General Approaches for Modeling Liquidity effects in Asset Markets and their Application to Risk Management Systems

Frank Milne
Economics Dept., Queen's University, Kingston, Ontario, Canada

June 29, 2010

Abstract

The paper begins by outlining the basic model to price and hedge securities. This model assumes perfectly liquid markets and is the basis for Risk Management (RM) models. The paper then explores three basic theoretical structures for modeling illiquidity in asset trading. We show that there are general results that can be deduced in these abstract models; but more specific models and results in the literature require much stronger assumptions and restrictions. Whether specific results in these more restrictive models will generalize is problematic.We suggest some other approaches that can supplement these RM models with illiquidity.

The current international crisis in credit markets has resulted in calls for improvements in (i) Risk Management (RM) systems in Financial Institutions (FI's), and (ii) the role of regulators and central banks in dealing with systemic crises. Many of these discussions gloss over the difficulties in operating RM systems, and assume that improvements can be made by better use of data, simple adjustments to incentives, and the use of regulatory capital and liquidity ratios. The problems in designing and operating RM systems, and regulatory responses are difficult, and in some cases almost intractable. But as I will argue below, there are some possible strategies we can follow that may improve RM performance by using recent models that include illiqudity in the pricing and hedging of financial assets. These types of models are in various stages of development and have not reached sufficient maturity to allow widespread implementation.The purpose of this paper is to lay out four general frameworks that can be used as a guide to various more specialized models in the literature. Using these framworks, it is possible to point to some general results that appear in more restricted settings. Also, we will discuss situations where highly restrictive models may be misleading: their frameworks rule out certain types behaviour or results that occur in a more general model.

RM systems have evolved over many decades. FI's that issue credit have long used credit ranking and other methods to manage their credit books. They have managed their books by adjusting rates, collateral, individual exposures and workouts in default. Because the lending book was largely illiquid, banks had limited ability to hedge their risks. Over time, these systems have become increasingly mechanised through credit scoring systems and other methods. But the big changes have occurred more recently, where securitisation has allowed RM's to hedge and trade credit risks. This requires different methods of pricing, hedging and managing credit exposures that must be integrated into more traditional systems. As we will argue below there were fundamental problems that occurrred in that integration.

The problems for the private sector RM systems can be summarised in two broad categories: (a) the underlying theoretical formulation of the RM systems; and (b) statistical calibration. The existing models are a synthesis of the traditional credit systems and the Arrow-Debreu model of trading, hedging and pricing assets and derivatives. This model, if taken seriously, implies there is a factor structure that can be used to price assets; that these factors (after diversification) can be traded in frictionless markets, and priced by arbitrage restrictions. As a system, it is general equlibrium plus linear algebra. Unfortunately it implies that the financial system does not add value and is welfare irrelevant. Modern Banking theory begins by acknowledging this deficiency and introduces various frictions to make sense of the existence of sophisticated financial systems. Banking theory has made very limited inroads into RM theory and practice, most of which has been dominated by the Arrow-Debreu model masquerading under the title of Financial Engineering. Recently the latter literature has been attempting to cope with the theoretical complexities introduced by frictions (eg transaction costs and "illiquidity") by reduced form methods, but the more general strategic and information problems that are of concern in the banking literature have been ignored. A further problem is that most banking theory models are relatively simple and of low dimension. They are exploratory, examining logical possibilities that could be consistent with stylized facts, but far from being implementable in any RM system.

At the regulatory level, a further layer of complexity is added in dealing with systemic risks. Whereas the RM systems take the environment as given, systemic risks require a full system to track interactions between FI's, and possible interactions with the real economy. An added requirement, if we are to justify intervention, is to have models with market failure(s). The final step is to explore possible regulatory frameworks that can rectify or ameliorate these failures. Building on recent Banking Theory there is a growing literature discussing bank failures and systemic risks. The literature on market failures has limited application for regulators and risk managers: it provides model exploration rather than directly implementable systems. Nevertheless, there has been some progress in practical modelling of systemic risks. These models have well-known deficiencies that became obvious in the Financial Crisis. Clearly there is much more work to be done in this area as the models should be used as prototypes for more sophisticated analysis.

1 Risk Management Systems

1.1 A Simple Model of a Financial Institution

We will begin by assuming a simple two date model for a Financial Institution. We assume that the institution has the same characteristics as an abstract investor portfolio problem (see for example Milne (2003)). The FI has an objective function that has standard utility function characteristics, that it maximizes over its stochastic net cash flow x_{1f} at time 1. The first constraint shows that the net cash flow at time 1 is generated by the FI's portfolio over J assets which have dividends or coupons R_{1j} and random prices p_{1j} at date 1 (we can assume a finite or an infinite state space and an associated probability distribution over asset returns), and asset j holdings of a_{j0} acquired at date 0. Note that the holdings a_{j0} can be positive or negative, indicating short selling or borrowing of riskless assets. Finally the FI is assumed to not take positions that generate negative cash flows in future states, after taking into account the amount of capital $C_1 > 0$ that the bank holds against losses from the portfolio. Later we will introduce other restrictions on the portfolio holdings that represent regulatory constraints.

$$\max_{a_f} U_f(x_{1f})$$

such that

$$\text{(a) } x_{1f} \leq \sum_j (R_{1j} + p_{1j}) a_{j0} + C_1$$

$$\begin{aligned} \text{(b) } W_0 &= \sum_j p_{0j} a_{j0} \\ \text{(c) } x_{1f} &\geq 0 \end{aligned}$$

The characterization of this problem is well-known and we will develop more sophisticated versions when we generalize the model later. Nevertheless it will suffice for our outline of Risk Management systems that are used by FI's.

The assumption that an FI has a utility function is highly problematic - we will concentrate on the construction of the constraint sets and asset returns. If the asset market return space spans the state space, then it is well-known that the utility function should satisfy $U_f(x_{1f}) = \Pi(x_{1f})$ where $\Pi(.)$ is a linear operator determined by the Arrow-Debreu prices or Martingale measure. If asset market returns are incomplete, then it is common to assume that the FI has a utility function. Whose preferences the utility function represents, is seldom made clear. As we will see below, the utility objective plays little role in operational RM systems: they concentrate on the constraint set and simulate the random payoff x_{1f} taking into account deviations from existing positions in an iterative process. Nevertheless, the objectives of the FI and its choice of asset portfolio and exposures is a serious issue that we will turn to later in this paper. Perverse incentives for management and/or traders could be captured - as a crude approximation - by preferences that place unduly low weights on losses, and/or have high tolerance for risks.

1.2 The Basic One Period Risk Management Model

The basic mathematical model that we sketched in the previous system can be made more operational introducing a factor structure on the asset return. The assumption that returns are generated by random factors has a long history in applied finance and underlies all RM systems. The idea is very simple: consider a fixed period of time, either a day or week, where a future asset price will vary randomly. For example consider a share whose price to day is p_{0j} and whose return is the random price p_{1j} and riskless coupon or dividend d_{1j} tomorrow: $R_{1j} = p_{1j} + d_{1j}$. Assume that the random price (plus dividends) can be described by a factor model:

$$R_{1j} = p_{0j}\ (1 + r) + \beta_{1j} F_{1j} + ...\beta_{Kj} F_K + \varepsilon_j$$

where r is the short term risk-free rate, the F's are K random, uncorrelated factor returns, the β's are called factor loadings, and ε_j is a random idiosyncratic term for the stock uncorrelated with the factors and the ε's for other stocks. This idea was first used in the 1960's to describe stock returns and simplify the calculation of covariances and variances for large equity portfolios.

It became apparent in the 1970's, that if the FI held a large diversified equity portfolio, so that the portfolio weighted ε's summed to approximately zero by the Law of Large Numbers, then the portfolio return could be described by a linear combination of the factor returns. The weights would be a combination of the portfolio weights and the β coefficients. Furthermore, in diversified portfolios, the prices of the assets would be restricted by possible arbitrage trades. To see this, ignore the diversifiable ε terms and assume that the number of factors is a small number – say two. Then we can deduce by elementary linear algebra that we can find a current price for each factor. Using these factor prices then every current stock price can be written as a linear combination of the factor prices using the β coefficients. If this linear pricing rule was not true, then any investor could take a portfolio of stocks and make unlimited profits. This factor pricing theory has various names, depending on the application: the Arbitrage Pricing Theory (APT); a one period version of financial derivative pricing; or the Generalized Modigliani-Miller Theorem (see Milne (2003) Chs.4 and 7.)

Financial economists observed that this one period method for pricing assets was simple and relatively easy to implement with standard econometric techniques. It had a number of limitations: the theory did not explain what the factors were, or whether the number of factors varied over time. Also, in identifying the factors, regression analysis would estimate the β coefficients. The problem was: were these coefficients stable over time; or would they be conditional on observable market variables? These issues have never been fully resolved, although, after strenuous empirical testing, there are some candidates for common factors (e.g. a stock market index, the short interest rate, industry factors derived from industry equity indices).

1.3 The Basic MultiPeriod Risk Management Theoretical Model

The next stage of development was when theorists understood that the same factor idea in a multiperiod setting. The underlying model is a simple extension of our model above where there are $T = \{0, 1, ..., T\}$ intermediate trading dates and (for simplicity) a finite tree structure for the evolution of information observable by all agents. At any time $t \in T$, there is an event ωt. Using this notation we can extend the FI problem to a multiperiod framework:

$$
\begin{aligned}
& Max_{a_f} U_f(x_f) \\
& \text{Such that} \\
\text{(a) } x_{\omega tf} \quad &\leq \quad \sum_j (R_{\omega tj} + p_{\omega tj}) a_{wt-1,j} - \sum_j p_{\omega tj} a_{\omega t,j} + C_{\omega t}, \\
& \text{for all feasible events and times } \omega t, t \neq 0 \text{ ;} \\
\text{(b) } W_0 \quad &= \quad \sum_j p_{0j} a_{0j} \text{ and } a_{\omega T,j} = 0 \text{ for all } j; \\
\text{(c) } x_{\omega t,f} \quad &\geq \quad 0, \text{for all } \omega t, t.
\end{aligned}
$$

This multiperiod model can be extended to a multifactor return structure so that we can derive a conditional factor structure for returns at each event.

Therefore the factor structure of returns can be reinterpreted as a conditional factor model, where the coefficients should be interpreted as conditional, and the number of factors could (in principle) vary over time or events.

Using this multiperiod factor model (for a derivation see Milne (2003) Chs 8-10) it can be used to price default-free bonds of different maturities. The trick is to observe that zero coupon bond prices for (simple transforms can be used to make the same argument for bond yields or forward rates) can be written as a factor structure model. This implies that the common factors will impact on bond prices depending on the β coefficients. Because bond prices converge to their face value at maturity, the β coefficients cannot be stationary. Other restrictions rule out dynamic arbitrage strategies.

These factor models have a further use: they provide a building block for derivative pricing and hedging. Consider the elementary equity binomial model of derivative pricing, that approximates the celebrated continuous time Black-Scholes (1973), Merton (1973) option pricing model. The idea is very simple: assume that the stock price evolves according to a one stochastic factor model plus a constant. Assume that the random factor is a binomial random variable. Then using the stock and the bond one can create a portfolio to replicate any derivative on the stock, one period ahead. Thus the option price must (otherwise arbitrage profits exist) equal the price of the replicating portfolio. Using this argument iteratively over time, assuming that the volatility parameter on the random factor is constant over time, and the risk free rate is constant over time, one can build up a dynamic portfolio strategy to replicate any European option payoff on the stock at time of maturity. Given a dynamic replication strategy, then the initial value of the strategy and the option price must equal to avoid an arbitrage opportunity.

This model is merely a simple prototype for more complex models that use more factors, or have more complex conditional volatility structures. Assuming a factor structure for bond prices, it is an easy step to create a bond option model where default-free bond prices follow a simple factor structure. By 1990 there were several bond option models in existence, and in short order they were implemented by major FI's on Wall St..

The next step made the bold assumption that one could apply the same factor idea to corporate bonds that could default. An early model by Merton (1973) had demonstrated the basic idea. Using an analogy between a stock option and a levered stock he was able to price the levered stock by the Black-Scholes-Merton model; and in turn by assuming the Modigliani-Miller theorem, deduce the value of the defaulting bond as a residual from the difference between the value of the firm and its equity value. This basic insight has spawned a whole battery of so-called "structural models" that extend this basic insight to more complex debt structures. A number of propriety models use this structure to price corporate debt.

A second group of models - the "reduced form" models (introduced by Jarrow and Turnbull (1995) and other theorists) - avoided describing the details of any firm's financial structure, but modeled default as another factor in the evolution of the bond price. Although simple in outline, the model can be extended to allow for additional information in bond ratings and other modifications to add realism to the bond pricing model. Given this structure, it is an easy step to use the replicating portfolio idea to create a perfect hedge for any credit derivative one wishes to dream up. Having created the replicating portfolio, then the price of the derivative must, by the familiar arbitrage-free argument, be the price of the replicating dynamic portfolio. Other variations of these models have been developed more recently to deal with complex derivatives on credit risks, and counterparty risks[1].

Both types of models, and their generalized versions, have been used extensively in the credit industry to model, price and hedge credit instruments. In turn the models have been modified to analyze Collateralized Debt Obligations (CDOs), Mortgage Backed Securities (MBS) and many variations that have allowed previously illiquid loans to be securitized and sold as part of larger packages or tranches via Conduits or Special Purpose Vehicles (SPV's). The underlying factor models used in this theory assume particular probability distributions over factors that explain default risk. Having created risks factors, specifying joint probability distributions, and making assumptions on the covariances between defaults of individual loans, one can create a portfolio of loans that reduces risks via standard diversification arguments. Having created this loan portfolio, then it can be sliced into tranches with increasing degrees of default risk. The safest tranche is modeled to be almost risk free; the second tranche (or mezzanine) has higher risk, and so on down, as the risks of the credit portfolio are sliced into packages that can be sold so as to mimic corporate bonds with different default risks or credit ratings.

Having dealt with trading, credit and derivative risks for an FI, there are other risks that can be incorporated into an FI's risk management system. For example, in recent years there have been attempts to model operational risks in the FI. The idea is that some FI losses have been due to errors in pricing, hedging or processing information, employee fraud, computer system failures, acts of terrorism, etc. Attempts have been made to model these risks as random shocks and calibrate the risks to actual data. The evidence suggests that high frequency small losses can be characterized with some degree of accuracy (e.g. small errors in entering data); but low frequency, large losses are far harder to estimate (e.g. large scale fraud, IT failure) and the FI must rely on internal audits, back-up systems etc. The operational risk models should be used in conjunction with standard auditing and other security practices to minimize the risks, given the costs of implementation. Other risks, that are harder to quantify, are legal risks and reputational risks that can arise when trading complex securities.

Returning to trading, credit and derivative risks, the factor structure inherent in the theoretical models allows the FI to aggregate risks into a net position for the FI. This can be illustrated by considering the simple, one period ahead returns for an FI and breaking the portfolio into different asset classes. For example, consider a simple model of an FI where there is an equity portfolio

[1]See Meissner (2005) for an accessible discussion of structural and reduced form models. See also Duffie and Singleton (2003) and Lando (2004).

(E), a government bond portfolio (B) and a derivative trading desk(D). Then we can partition the return equation into those porfolios to obtain:

$$x_{1f} = \sum_{j \in E} (R_{1j}^E + p_{1j}^E) a_j^E + \sum_{j \in B} (R_{1j}^B + p_{1j}^B) a_j^B + \sum_{j \in D} (R_{1j}^D + p_{1j}^D) a_j^D + C_1$$

Introducing factor structures for each type of security, we can aggregate the portfolio returns to obtain a factor exposure for the FI as a whole. Clearly the factor loadings will vary across types of securities, but the correlations between the factor exposures will be important in computing aggregate distributions of returns for the FI. For example, a common market factor will appear in equities and equity derivatives.

Given a short period ahead – one day or one week - the FI should be able to produce (in theory) an estimated probability distribution over their total returns. The last step is to calculate the Value at Risk (VaR). This measure of the loss exposure of the FI, is nothing more than the cut-off for lower tail threshold at 5 percent or 1 percent. FIs report their VaR as a measure of the potential losses given their holdings and return risks on their net exposure.

It is important to understand that, in both the stock, bond, credit and derivative models we have discussed, the factor structure allows one, in principle, to create a one period ahead net exposure where, if the model is to be believed, the firm issuing a derivative can perfectly hedge and have a net position of zero. This basic theoretical result, if taken literally, has surprising implications that many commentators overlook. One can show quite generally, that if derivatives can be replicated by static or dynamic portfolios of competitively traded securities, transactions costs are ignored, and the derivative industry is competitive, then trading derivatives adds not one dollar profit or jot of welfare advantage to any agent in the economy (see Milne (2003) Chs 9 and 10) This result is a direct implication of the multi-period, generalized version of the Modigliani and Miller Theorem from finance theory. To make the theory more realistic, one requires transaction costs, illiquidity, asymmetric information and other frictions to make sense of the FI industry (see Freixas and Rochet (2008) Ch.1). Statements about derivatives "completing markets" and leading to welfare benefits are problematic and should be treated with scepticism. Although it is possible to argue that improved diversification can act as a benefit to the economy, that assertion should be tempered by the risks introduced by the inappropriate use of financial instruments in markets that are inherently illiquid. As we will argue below, realistic features that make some financial markets imperfect, make the frictionless factor model an approximation at best; and at worst, a misleading paradigm that can lead to unexpected financial losses.

2 Modelling a Risk Management System with

Liquidity:Four Approaches in a Competitive Economy:

There have been various approaches to modelling different aspects of liquidity. We will provide a sketch of four basic (related) models and how they might, in principle, be incorporated into RM systems. We do not claim that this is an exhaustive set of approaches, but they do provide some standard ways for introducing liquidity effects into RM type models. The models abstract from asymmetric information between agents,and other subtle sources driving illiquidity, so that we can focus on reduced form methods that can be adapted to standard RM models. Also we will comment on some basic difficulties in actual implementation and suggestions for further research.

The first approach assumes a bid-ask spread for any traded asset in competitive security markets. Transaction costs can be interpreted narrowly as trading costs in using computers and paying traders, or more abstractly as a parsimonious reduced form that captures the impact of asymmetric information and or bargaining in asset markets. This type of transaction cost model has a long history in economic theory and has been explored in many variations. At one extreme it can allow a complete asset market structure (zero transaction costs on all assets); and at the other extreme the incomplete asset market model (zero transaction costs on one set of assets, and infinite transaction costs on the remainder). The intermediate case is more realistic in having moderate bid-ask spreads, so that contingent trades occur, or not occur, in equilibrium. This type of model does not seem to have been used in a systematic fashion in Risk Management models. In its general competitive equilibrium version, the model allows a structure that can mimic random illiquidity events when conditional bid-ask spreads widen dramatically. The transaction cost model's equilibrium allocation will not, in general, be efficient. Also by considering the incomplete market model version we know from the older literature that one can construct simple examples where multiple equilibria may exist and be Pareto ranked. Furthermore, the introduction of a new security can have the counterintuitive result that all agents are made worse off[2]. This class of model has inefficiency built in, so there may be role for government intervention. We will sketch a general model and some characterizations that will allow us to nest many simpler models and their results on allocations and prices in this class of model.

In the second approach we assume agents have constraints that may arise from regulation, margin requirements or other restrictions on their portfolio strategies. Some of these constraints come from regulations, and others are reduced forms that mimic restrictions imposed by private agents to deal with lack of transparency in counterparty risk. If we assume that there are incomplete asset markets (transaction costs on some asset markets will suffice to imply incompleteness), then we can model the implications of these constraints on asset prices and allocations. Furthermore, it should be possible in low dimension

[2]See Hart (1975) and Milne and Shefrin (1986).

versions of the model to deduce the implications of increasing or decreasing these constraints.

The previous two approaches dealt with market liquidity: that is, impediments to trading stocks, bonds and hedging and pricing derivatives. The third class of model considers funding liquidity. We will consider a class of single FI models that are related to the second class of model. They have (random) constraints on sources of funding to mimic withdrawal of funding sources or margin or collateral restrictions imposed by lenders on portfolios of the FI. These types of models can be embedded in a competitive economy to track the implications for asset allocations and prices when funding liquidity constraints tighten.

The fourth approach models illiquidity as a result of non-competitive behaviour: that is, traders can move prices by trading. This type of model is a more recent innovation and can be treated in small dimensional models that mimic dynamic oligopoly models. There have been four sub-branches in this literature: the first explores dynamic portfolio trading strategies in small dimension general equilibrium models ; the second considers dynamic derivative pricing and hedging; the third explores strategic behaviour by FI's in predating stressed competitors, and the fourth explores the liquidity impacts of trading and the role of liquidity provision by central banks.

3 Liquidity Models 1 and 2: Transaction Costs, Bid-Ask Spreads and Constraints on Asset Holdings

This economy is a straightforward extension of the standard, discrete time/event economy familiar from the literature on general equilibrium theory[3]. Consider an economy with a finite (or countable) time horizon $T = \{0, 1, ..., T\}$ and uncertainty characterized by a finite tree or filtration. The filtration is represented by $F = \{F_t, t \in T\}$ where each F_t is an increasingly finer partition of events. Including the initial node, there are $\Omega + 1$ nodes on the tree. For simplicity, we assume a single commodity. While it is not difficult to introduce multiple commodities and a sequence of spot markets, to do so adds nothing to our discussion of asset trading and pricing[4]. All agents have rational expectations of asset prices expectations at any event node and are not falsified by the evolution of the economy.

[3]See Milne (2003) and Magill and Quinzii (1996) for detailed text book discussions of the general equilibrium model without transaction costs.

[4]For a full description of the economy see Jin and Milne (1999). That paper provides conditions for the existence of an equlibrium in a finite economy with consumers, productive firms and intermediaries. See also Milne and Neave (2003) for a discussion of some properties of equilibria for this type of model. In much that follows, I draw on results in that paper.

3.1 Agents

Assume that agents are either consumers or financial firms - excluding financial brokers. We ignore productive firms, although the model can be extended easily to cover those types of agents. Consider the characteristics of a set of agents $I = \{1, ..., I\}$ or in a continuum $[0, 1]$. Assume agent $i \in I$ has a consumption set $X_i = \Re_+^{\Omega+1}$, and preferences described by a standard utility function U_i which is continuous; quasi-concave; strictly increasing and is differentiable on the $int(X_i)$.

The differentiability assumptions are convenient for using calculus. We will assume in our discussion below that optimal consumption vectors will be in the $int(X_i)$ thus allowing us to concentrate on financial asset trading restrictions and asset prices. Later will discuss the situation where the agent defaults. Otherwise the restrictions on the agent's consumption set and preferences are mild and allow a range of familiar special cases.

3.2 Security Returns and Prices

Agents trade in financial asset or security markets that transfer purchasing power across contingencies. We will restrict our discussion to securities with fixed payoffs (e.g. Bonds, Equities, Credit Securities, European options). Since trading and issuing securities involves the services of costly intermediation that consumes real resources, we introduce buying and selling prices for assets and a non-trivial role for financial intermediaries.

We describe buying and selling security prices by introducing an asset payoff matrix R and two asset price matrices[5]. First, assume that there is a finite but very large set of assets $K = \{1, ..., K\}$, including all the usual types, such as bonds, stocks, and European derivatives. Our definition of an asset k includes the event when it is traded. Thus an asset traded at one event, and the same asset traded at a subsequent event are considered different assets. This convention is introduced in our initial discussion to keep the notation as compact as possible. In later analysis we will use more detailed subscripts to show how more elaborate equations are merely reinterpretations.

Let the matrix of buying prices, B, be of dimension $(\Omega+1)\times K$. Similarly, let the matrix of selling prices, S, be of dimension $(\Omega+1)\times K$. (We will see below, that in equilibrium, $B \geq S$, i.e. buying prices will be at least as great as selling prices: the differential will be explained by transactions costs.) Finally, let R be an $(\Omega+1)\times K$ matrix of total returns or payoffs (dividends, coupons) for each asset. This structure will allow us to consider static and dynamic trading strategies where there are buying and selling prices for any multi-period asset, including bonds, stocks or European derivatives.

Although we do not introduce a factor structure for security prices and returns, it is an easy exercise to do so[6].

[5]This compact notation is used in Milne (2003) and Naik (1995).

[6]See Milne (2003) Ch. for the simple construction of orthonormal factors.

3.3 Brokers

The second group of agents are brokers or FI's who purchase and resell securities using a costly technology. This technology is a crude reduced form for the more complex system of asset trading, bargaining and search. There is an extensive literature that has modelled various types of liquidity problems in pricing and trading. What we are doing here is "black-boxing" that process to concentrate on the brokers and market clearing, using the well-known properties of a competitive economy to solve for prices and asset allocations. For broker $h \in H$ (assume that H is either a finite set or a continuum);assume there is a technology $T_h \subset \Re_+^K \times \Re_+^K \times \Re^{\Omega+1}$, with typical element $(\Delta_h^B, \Delta_h^S, y_h)$; where $\Delta_h^B$is the number of "bought" securities supplied by the broker; Δ_h^S is the number of "sold" assets purchased by the broker; and y_h is the vector of contingent commodities used up in the activity of intermediation. Our convention on "bought" and "sold" is from the viewpoint of the agent.

The technology incorporates pure dealing or broking activity, where the asset k passes through directly, i.e., $\Delta_{kh}^B = \Delta_{kh}^S$; or the more complex situation where the broker supplies transformation services as in the case of banks or other financial intermediaries. In this latter case intermediaries can buy assets, recombine their claims in a dynamic portfolio, and sell other assets backed by the portfolio. For simplicity in our discussion here we will assume that the intermediary h is a pure broker, and treat other intermediaries as a subset of the class of agents I. In summary, broker h's problem is:

$$\max V_h(x_h)$$

$$x_h \leq R[\Delta_h^S - \Delta_h^B] - S\Delta_h^S + B\Delta_h^B + y_h;$$

$$(\Delta_h^B, \Delta_h^S, y_h) \in T_h; x_h \in X_h.$$

We assume that the broker technology satisfies standard micro-economic conditions: T_h has the properties:(a) $0 \in T_h$ (b) T_h is closed (c) T_h is convex.

Notice that we have ruled out fixed costs or economies of scale in the transaction technology. Certain types of transaction technologies and intermediation have these non-convex properties. We will discuss this issue in below, explaining the modifications that can be made to our results to allow for this type of transaction technology. We assume that the broker is owned by a single entrepreneur to avoid the complications introduced by the failure of the Fisher Separation Theorem.

Assume that: (i) $X_h$is the non-negative orthant; and (ii) $V_h(\cdot)$ is differentiable, strictly increasing and quasi-concave.

We could assume that the broker has a stochastic cash flow endowment and an initial stock of securities, but for simplicity we will omit those simple generalizations here.

3.4 The Agent Problem

For each agent $i \in I$, there is a problem:

$$\max_{x_i \in X_i} U_i(x_i)$$

s.t.

$$x_i = R[\Delta_i^B - \Delta_i^S] + S\Delta_i^S - B\Delta_i^B + \bar{x}_i + y_i;$$

$$(\Delta_i^B, \Delta_i^S, y_i, S, B) \in T_i.$$

where $\Delta_i^B \in \Re_+^K, \Delta_i^S \in \Re_+^K$ are the agent's purchases (sales) of asset $k; \bar{x}_i$ is an exogenous random endowment; and T_i is the agent's own transaction technology, or constraints that arise from regulations, margin requirements etc.

T_i has the properties that for given S and B : (a) $0 \in T_i$;(b) T_i is closed;(c) T_i is convex. For an equilibrium to exist, we would require continuity conditions on T_i with respect to changes in the prices S and B. We will come back to this point later.

The agent and broker problems are virtually the same in the abstract, apart from a change in sign in the budget constraints, to represent the broker taking the other side of the market from agents, and the introduction of prices in the constraint set. This symmetry is deliberate because we wish to show that the difference is not one of substance, but of details in the structure of the transaction technology. This will become clear below when we characterize the agent's problem. Finally, the economy is closed by requiring market clearance for all contingent commodities and the buying and selling asset markets (We will assume here a finite number of agents, but with a continuum we would require integrals) .

$$\sum_h x_h + \sum_i x_i = \sum_h y_h + \sum_i y_i + \sum_i \bar{x}_i$$

$$\sum_i \Delta_i^B = \sum_h \Delta_h^B$$

$$\sum_i \Delta_i^S = \sum_h \Delta_h^S$$

Defn.1.1 A competitive equilibrium for this economy is the buying and selling price matrices B, S; bought asset demand and supplies (Δ_i^B, Δ_h^B); and sold asset demand and supplies (Δ_i^S, Δ_h^S) that are solutions to the agent, broker problems, and market clearance equations.

Jin and Milne (1999) provide sufficient conditions for the existence of an equilibrium in a version of this economy, with a finite number of agents and multiple commodities. That paper discusses the case with convex technologies for the brokers; and also explores existence results with non-convex technologies and large numbers (or a continuum) of agents.

4 Characterizations of Agent Optimality Conditions

4.1 The Agent's Problem: First Order Conditions and Personalized Martingale Pricing

In this section we characterize the agent's optimum using the Kuhn-Tucker first order conditions for a maximum. Given the quasi-concave nature of the problem, the first order conditions are necessary and sufficient for a maximum. This will allow us to derive a useful characterization of buying and selling asset prices and relate them to the personalized Martingale methodology. For simplicity, we assume that T_i is defined by a system of functional inequalities $F_{i\ell}(\Delta_i^B(\omega_t), \Delta_i^S(\omega_t), y_i(\omega_t); S(\omega_t); B(\omega_t); \omega_t) \geq 0$, for all ω_t and $\ell = 1, ..., L$. Later we will introduce a more general set of constraints that allow for more complex dynamic constraint sets. We begin by analyzing the consumer's problem. The Lagrangian is:

$$L = U_i(R[\Delta_i^B - \Delta_i^S] + S\Delta_i^S - B\Delta_i^B + y_i + \bar{x}_i) + \sum_k \delta_{ki}^B \Delta_{ki}^B + \sum_k \delta_{ki}^S \Delta_{ki}^S + \sum_{\omega_t} \sum_\ell \delta_{i\ell}(\omega_t) F_{i\ell}(\cdot)$$

where $\delta_{ki}^B, \delta_{ki}^S, \delta_{i\ell}$ are the appropriate Kuhn-Tucker multipliers. The first order conditions for a maximum with ($x_i^* \in intX_i, y_{i\ell}^*(\omega_t) > 0$) are:

$$\frac{\partial Li}{\partial \Delta_{ki}^B(\omega_t)} = \nabla U_i[R_k(\omega_t) - P_k^B(\omega_t)] + \delta_{ki}^B(\omega_t) + \sum_\ell \delta_{i\ell}(\omega_t) \frac{\partial F_{i\ell}}{\partial \Delta_{ki}^B(\omega_t)} = 0$$

$$\delta_{ki}^B(\omega_t)\Delta_{ki}^B(T_t) = 0; \delta_{ki}^B(\omega_t) \geq 0; \Delta_{ki}^B(\omega_t) \geq 0; \delta_{i\ell}(\omega_t) \geq 0;$$

$$F_{i\ell}() \geq 0; \delta_{i\ell}(\omega_t) F_{i\ell}(\omega_t) = 0$$

$$\frac{\partial L_i}{\partial \Delta_{ki}^S(\omega_t)} = \nabla U_i[-R_k(\omega_t) + P_k(\omega_t)] + \delta_{ki}^S(\omega_t) + \sum_\ell \delta_{i\ell}(\omega_t) \frac{\partial F_{i\ell}}{\partial \Delta_{ki}^S(\omega_t)} = 0$$

$$\delta_{ki}^S(\omega_t)\Delta_{ki}^S(\omega_t) = 0; \delta_{ki}^S(\omega_t) \geq 0; \Delta_{ki}^S(\omega_t) \geq 0$$

$$\frac{\partial L}{\partial y_i(\omega_t)} = \frac{\partial U_i}{\partial x_i(\omega_t)} + \sum_\ell \delta_{S\ell}(\omega) \frac{\partial F_{i\ell}}{\partial y_i(\omega_t)} = 0.$$

Note: $R_k(\omega_t)$ is the vector of future contingent payoffs for asset k conditional on ω_t.

Now we can rewrite the marginal utility vector, using a sequence of normalizations, so that it becomes a personalized, discounted, conditional martingale measure. Notice that these probability measures are merely personalized undiscounted Arrow-Debreu prices (or personalized martinglae measure) with the same properties as probabilities. Consider

$\omega_t \in F_t$, given $\omega_s \in F_s$, where $s > t$ and ω_s is a node reached by beginning at ω_t. Define $\gamma_i(s \mid \omega_t) = \sum_{\omega_s \in S(s|\omega_t)} (\lambda_i(\omega_s)/\lambda_i(\omega_t))$, where $\lambda_i(\omega_t) = \frac{\partial U_i}{\partial x_i(\omega_t)}$; and $S(s \mid \omega_t)$ is the subset of F_s which can be reached by starting from ω_t.

Finally, define: $\tilde{p}_i(\omega_s \mid \omega_t) = [\lambda_i(\omega_s)/\lambda_i(\omega_t)][\gamma_i(s \mid \omega_t)]^{-1}$. Given that $\nabla U_i >> 0$ by assumption, then by construction $(\tilde{p}_i(\omega_s \mid \omega_t))$ has the same properties as a conditional probability measure with full support.

We can rewrite the first order conditions as:

$$\sum_\ell \delta_{i\ell}(\omega_t) \frac{\partial F_{i\ell}}{\partial \Delta^B_{ik}(\omega_t)} + \frac{\delta^B_{ki}(\omega_t)}{\lambda_i(\omega_t)} - P^B_k(\omega_t) + \sum_{s>t} \gamma_i(s \mid \omega_t) \sum_{S(s|\omega_t)} R_k(\omega_s)\tilde{p}_i(\omega_s \mid \omega_t) = 0$$

$$\delta^B_{ki}(\omega_t)\Delta^B_{ki}(\omega_t) = 0; \delta^B_{ki}(\omega_t) \geq 0; \Delta^B_{ki}(\omega_t) \geq 0.$$

$$\sum_\ell \delta_{i\ell}(\omega_t) \frac{\partial F_{i\ell}}{\partial \Delta^S_{ik}(\omega_t)} + \frac{\delta^S_{ki}(\omega_t)}{\lambda_i(\omega_t)} + P^S_k(\omega_t) - \sum_{s>t} \gamma_i(s \mid \omega_t) \sum_{S(s|\omega_t)} R_k(\omega_s)\tilde{p}_i(\omega_s \mid \omega_t) = 0$$

$$\delta^S_{ki}(\omega_t)\Delta^S_{ki}(\omega_t) = 0; \delta^S_{ki}(\omega_t) \geq 0; \Delta^S_{ki}(\omega_t) \geq 0.$$

4.2 The Agent's Problem Without Personal Transaction Costs or Asset Constraints

Assume, for the moment that we can ignore personal transaction costs or constraints on asset trades. Consider first the simple case of a security k which lives only one period, i.e., $s = t + 1$. Assume that the security pays a dividend (or coupon) $R_k(\omega_{t+1})$. Then:

$$\frac{\delta^B_{ki}(\omega_t)}{\lambda_i(\omega_t)} - P^B_k(\omega_t) + \gamma_i(t+1 \mid \omega_t) \sum_{S(t+1|\omega_t)} R_k(\omega_{t+1})\tilde{p}_i(\omega_{t+1} \mid \omega_t) = 0.$$

$$\frac{\delta^S_{ki}(\omega_t)}{\lambda_i(\omega_t)} + P^S_k(\omega_t) - \gamma_i(t+1 \mid \omega_t) \sum_{S(t+1|\omega_t)} R_k(t+1)\tilde{p}_i(\omega_{t+1} \mid \omega_t) = 0$$

.

Immediately, we obtain the result:

$$P^B_k(\omega_t) - P^S_k(\omega_t) = \frac{\delta^B_{ki}(\omega_t) + \delta^S_{ki}(\omega_t)}{\lambda_i(\omega_t)} \geq 0.$$

By the monotonicity of utility $\lambda_i > 0$; if $P^B_k(\omega_t) - P^S_k(\omega_t) > 0$ (i.e., there is a bid-ask spread), then $\delta^B_{ki} + \delta^S_{ki} > 0$.

This implies that either the agent has:
(a) $\Delta^{\delta}_{ki} = \Delta^{B}_{ki} = 0$ – does not trade asset k;
(b) $\Delta^{\delta}_{ki} > 0; \Delta^{B}_{ki} = 0$ – sells , but does not buy asset k;
(c) $\Delta^{\delta}_{ki} = 0; \Delta^{B}_{ki} > 0$ – does not sell, but buys asset k.

Notice that the agent will not simultaneously buy and sell asset k; transactions costs make it inefficient to contemporaneously go long and short in an asset. This simple result is intuitive and fundamental showing that transaction costs induce asset trading stickiness around the contingent asset endowment point. In a one period model, this is obvious, but in a multiperiod context characterizations will be sensitive to the specification of the asset returns, contingent bid-ask spreads etc. We will discuss some general characteristics of dynamic asset strategies below.

Next, consider the general case where there are long-lived assets that have returns $\{R_k(\omega_s) \geq 0\}, \omega_s \in S(s \mid \omega_t), s > t$. Recalling the general conditions, we obtain:

$$P^B_k(\omega_t) - \frac{\delta^B_{ki}(\omega_t)}{\lambda_i(\omega_t)} = \sum_{s>t} \gamma_i(s \mid \omega_t) \sum_{S(s|\omega_t)} R_k(\omega_s)\tilde{p}_i(\omega_k s \mid \omega_t) = P^S_k(\omega_t) + \frac{\delta^S_{ki}(\omega_t)}{\lambda_i(\omega_t)};$$

or, defining $P^i_k(\omega_t) = \sum_{s>t} \gamma_i(s\omega_t) \sum_{S(s|\omega_t)} R_k(\omega_s)\tilde{p}_i(\omega_s \mid \omega_t)$ to be the personalized shadow price for asset k, $P^B_k(\omega_s) \geq P^i_k(\omega_s) \geq P^S_k(\omega_s)$.

Thus, for each asset k at event ω_t, there is a personalized discounted Martingale characterization of the asset pricing process when the current price lies between the buying and selling price [7].

Indeed, it is an easy extension to obtain for a long-lived asset:

$$\begin{aligned} P^B_k(\omega_t) &\geq \\ P^i_k(\omega_t) &= \gamma_i(t+1 \mid \omega_t) \sum_{S(t+1|\omega_t)} \{P^i_k(\omega_{t+1}) + R_k(\omega_{t+1})\}\tilde{p}_i(\omega_{t+1} \mid \omega_t) \\ &\geq P^S_k(\omega_t) \end{aligned}$$

which provides the explicit discounted Martingale structure. (Notice that there is an abuse of notation in that the asset k at ω_{t+1} should be treated as another asset k'.)

If the consumer is not trading asset k then $P^i_k(\omega_t), P^i_k(\omega_{t+1})$ will lie strictly between their buying and selling prices. But if the consumer is buying (or selling) at ω_t , then $P^B_k(\omega_t) = P^i_k(\omega_t)$(or $P^S_k(\omega_t) = P^i_k(\omega_t)$).

4.3 A Stochastic Linear Programming Version of the Model

So far we have characterized the agent's optimum in terms of shadow prices at the agent optimum. By exploiting these dual (Martingale) prices, we can simplify the agent's optimizing problem to a Linear Program. That is: at the

[7] This a well-known result. See for example Jouini and Kallal (1995).

optimum, x_i^* , consider the gradient vector $\nabla U_i(x_i^*)$ and the derived Martingale prices, $\{\gamma_i(), \tilde{p}_i()\}$, and solve the problem:

$$\max \sum_s \gamma_i(s \mid 0) \sum_{S(s|0)} x_i(\omega_s)\tilde{p}_i(\omega_s \mid 0)$$

s.t.

$$x_i \leq R[\Delta_i^B - \Delta_i^S] + S\Delta_i^S - B\Delta_i^B + \bar{x}_i$$

$$\Delta_i^S \geq 0; \Delta_i^B \geq 0.$$

In solving this problem notice that by the choice of $\{\gamma_i, \tilde{p}_i()\}, x_i^*$ must be a solution to the more general non-linear consumer problem. The LP formulation provides a basis for numerical computation of an optimal portfolio strategy. In general, it will be a large dimension problem and time consuming to solve. Nevertheless we can make some general observations from the structure of the problem. First, the solution will depend upon the preferences and optimal consumption plan of the consumer, as represented by shadow prices $\{\gamma_i, \tilde{p}_i()\}$. Second, depending upon the scale of the bid-ask spread of the different assets, some assets may be traded infrequently, or not at all (i.e. some assets will have corner solutions at the optimum).

With additional assumptions it is possible to simplify the characterization. For example, Edirisinghe, Naik and Uppal (1993) consider a binomial lattice, three securities (stock, bond and call option), and a linear objective, minimizing the initial cost of super-replicating a given payoff . Either one can think of their problem as a special case of our LP problem with the additional non-negativity constraint on final payoffs; or one can introduce a piece-wise-linear objective with severe penalties on negative cash-flows at the final replicating date.

Another rationalization of our linear framework would be to assume that the agent has a large endowment or portfolio so that, as a local approximation, asset trading has a small impact on the optimal x_i^*. In this case, it could be argued that the linear objective $\nabla U_i(x_i^*)$ is an appropriate objective for local hedging. Notice that this approximation can give a different solution to the Edirisinghe, Naik, and Uppal characterization. This formulation allows for the possibility of under or over replication of any portfolio that will trade-off cheaper trading strategies with the approximate contingent payoffs at the terminal date. Such a formulation approximates commercial practice in allowing the agent's portfolio payoffs to buffer approximate payoff hedges. Super-replication is too strong in assuming a very restrictive criterion for portfolio hedging.

Using this framework it is easy to add more realistic linear regulatory or legal constraints on asset allocations that are motivated by stress tests for an FI[8].

[8]The paper by Frauendorfer and Schurle (2007) oulines a stochastic linear programming model for an agent or bank that suffers random withdrawals from demand deposits or non-maturing accounts. The model allows one to solve a dynamic hedging problem with complex, multiple constraints.We will discuss this funding liquidity model in more detail below.

Non-linear constraints are complex to analyze, especially if they are non-convex. In this case there can be possible discrete jumps in optimal responses to price changes; and in an equilibrium settings, existence of a competitive equilibrium can be problematic. Also observe from the general first order conditions for the non-linear problem, that the constraints enter via contingent multipliers. These multipliers imply that the contingent returns must be modified by the marginal impact of future binding constraints. This provides a clue to the role of liquid securities in portfolios, in that they may earn a lower nominal return, but provide benefits in relaxing binding constraints that less liquid assets may not. This is the key to models that have liquidity premia for asset prices.[9]

Observe that the model collapses to the standard RM model when $B = S = P$.Thus the model nests the frictionless model. It is easy to rewrite the buying and selling prices in a factor structure so that for some asset classes the frictionless assumption is a reasonable approximation; but other asset classes have significant transaction costs, mimicking their less liquid markets.

Also notice that the model allows the transaction costs to be varied, conditionally to simulate illiquid periods. One could concentrate on the situation when $\bar{x}_i = R\overline{a}_i$ is the given portfolio of asset exposures at the current time and explore various scenarios choosing the Δ_i^B, Δ_i^S so that one can simulate the distribution of returns over one or several periods. This type of model can capture the slow speed of portfolio adjustment induced by illiquid markets. Also it provides a motivation for a central bank to intervene if it perceives that the bid-ask spreads are temporary and induced by asymmetric information between FI's. Of course that story requires the central bank, acting in cooperation with a regulator, to exploit its privileged information.

4.4 Further General Characterizations

Returning to the general problem, we can provide further general characterizations for the optimal portfolio program. First consider, $\beta(R, S, B, \bar{x}_i)$, or β for short in the opportunity set in our basic problem for the agent: it is a closed, convex set as the intersection of a set of linear inequalities, and it has an interior point. Furthermore, with a non-trivial bid-ask spread, there will be a kink at the endowment point $\bar{x}_i$. Consider the set $\rho(\bar{x}_i) = \{x_i \in \Re^{L+1} \mid U_i(x_i) \geq U_i(\bar{x}_i)\}$. By assumption, it is nonempty, closed and convex. Furthermore, $\bar{x}_i \in \beta 1 \bigcap \rho(\bar{x}_i)$. With sufficiently high transaction costs (i.e., a large $B - S$ spread) the only solution will be $\{x_i^* = \bar{x}_i; \Delta_i^{B*} = \Delta_i^{S*} = 0\}$, the no trade solution. The other extreme occurs where there are no transaction costs $(B = S = P)$ and the consumer will trade almost always. Of course, in general, with intermediate transaction costs, the optimal portfolio will be a sequence of trades and no trades across assets and events.

If we define τ^B, τ^S to be the matrix of proportional transaction cost rates around some transaction cost free price matrix P, then $B \equiv (1 + \tau^B)P$ and

[9] In addition, those models often require assumptions implying which marginal investors are trading at each event, determining prices. This approach simplifies the analysis greatly as it predetermines who are the marginal traders.

$S \equiv (1 - \tau^S)P$. If we consider all pairs $(\tau^B, \tau^S) \in \Re_+^{\Omega+1} \times \Re_+^{\Omega+1}$ and consider the partial ordering $\geq$, we can nest the opportunity sets $\beta(\tau^{B''}, \tau^{S''}, P, \bar{x}_i) \supseteq \beta(\tau^{B'}, \tau^{S'}, P, \bar{x}_i)$ for $(\tau^{B''}, \tau^{S''}) \geq (\tau^{B'}, \tau^{S'})$. By construction both opportunity sets contain $\bar{x}_i$ on their boundaries. Clearly, the consumer is worse off, or at least indifferent to higher transaction cost rates. The impact on asset trades is complex, and requires careful analysis, because the trades will be sensitive to assumptions on preferences, the asset price processes and returns, and the stochastic endowment stream.

4.5 Implications of Preference and Return Restrictions on Asset Trading

By placing additional restrictions on the preferences, the stochastic processes governing buying and selling prices, and the time horizon, it is possible to obtain tighter characterizations for portfolio strategies[10]. Most of these characterizations rely on expected, inter-temporally additively separable utility and simple Markov process assumptions on the prices and transaction costs to deduce Markov portfolio strategies. It is common to assume in addition that preferences are affine homothetic (i.e. HARA) and a limited menu of assets to introduce added simplicity to the characterization. In the absence of transaction costs, the HARA restriction on utility will allow portfolio demand functions to be affine linear in contingent wealth in the traded region. The introduction of transaction costs adds an additional layer of complexity.

4.6 Asset Pricing and Transaction Costs

A second literature focuses on the implications for asset prices when there are transaction costs[11]. This class of models introduces strong restrictions on agent preferences and their desire to trade. For example, an early class of models assumed risk neutrality, no borrowing and fixed (or random times) for agents exiting the market for a particular class of asset. The models showed that assets with long term cash flows earned a premium over short term assets as they were held by "patient" agents. It is not difficult to extend the model, allowing risk aversion, to show that there exist risk premia that are related to other factors. For example, the existence of market wide illiquidity will impact on the prices of all assets; the degree depending on the amount of correlation between the liquidity of the asset and market liquidity. Thus less liquid assets should earn higher expected returns.

This is an important literature because it can provide plausible reasons for various "anomalies" that appear in empirical tests of the standard, frictionless

[10] Constantinides (1979), Davis and Norman(1990) for an early sample of papers; and Karatzas and Shreve (1998) Ch.6 for a later survey. More recently see Liu (2004), Liu and Lowenstein (2007) and Dai, Jin and Liu (2008).

[11] For an excellent introduction and summary of these models see Amihud, Mendelson and Pedersen (2006) Chapter 2. In Chapter 3 they summarize the empirical literature using these types of theoretical models.

models This literature shows that evidence of high expected returns compared to standard factor pricing models, may simply signal lower liquidity, or lower contingent liquidity. Therefore standard RM models that do not take these liquidity factors into account will be misspecified.

This class of model is easy to incorporate into our general model. Consider a simple version of type class of model due to Amihud and Mendelson (1986)[12]. Assume an overlapping generations type of model with risk neutral agents. There constant riskless rate in each period r, a set of risky asset each period with iid returns, and all agents are risk neutral. Agents buy the illiquid assets at price P a time t and sell them next period $t+1$ when they die. When they sell the asset at $t+1$ they incur a proportional transaction cost of τ. This transaction cost can be introduced either by using the technology of a dealer, or by using a personal transaction cost for each agent, and assuming that the linear technology is identical for each agent. Recalling the agent's first order condition above:

$$\begin{aligned} P_k^B(\omega_t) &\geq \\ P_k^i(\omega_t) &= \gamma_i(t+1 \mid \omega_t) \sum_{S(t+1|\omega_t)} \{P_k^i(\omega_{t+1}) + R_k(\omega_{t+1})\} \tilde{p}_i(\omega_{t+1} \mid \omega_t) \\ &\geq P_k^S(\omega_t) \end{aligned}$$

But given the additional restrictive assumptions on preferences, returns and transaction costs:

$$\begin{aligned} P_k^B(\omega_t) &= P_k \\ &= \gamma(t+1 \mid \omega_t) \sum_{S(t+1|\omega_t)} \{P_k(\omega_{t+1}) + R_k(T_{t+1})\} p(\omega_{t+1} \mid \omega_t) \\ &= (1/1+r)) \sum_{S(t+1|\omega_t)} \{P_k^S(\omega_{t+1}) + R_k(\omega_{t+1})\} p(\omega_{t+1}) \\ &= (1/(1+r)) \sum_{S(t+1|\omega_t)} \{P_k - \tau + R_k(\omega_{t+1})\} p(\omega_{t+1}) \\ &= (1/r)\{E(R_k) - \tau\} \end{aligned}$$

It is easy to extend this model in a number of directions. For example, by modifying assumptions on the longevity of different agents, one can illustrate clientele effects so that agents with longer lives will hold less liquid assets, and shorter lived agents (or those who have short run liquidity demands) will pay more for shorter maturity assets.

Another variant on our general model is to consider an overlapping generations model where agents are born at t, die at $t+1$, and have identical exponential utility functions. Asset returns and transaction cost are stochastic and follow first order autoregressive processes. It is easy to show from our

[12]See Amihud, Mendelson and Pedersen (2006) Chapter 2 for details of this model and extensions.

first order conditions above, that asset expected returns can be described by a multi-factor Capital Asset Pricing Model (CAPM) formula that incorporates transaction cost risks as additional factors[13].

There has been an extensive empirical literature exploring these types of market microstructure models of illiquidity[14]. Empirical research shows that cross section and time series data on equity and bond market returns reflect liquidity effects. The greater the illiquidity, the higher the average return to compensate for illiquidity costs. Furthemore there is evidence of comovement in liquidity across some markets. Market efficiency can be effected by illiquidity. In an efficient market with minimal transaction costs, information effects should be very short-lived. But with significant trading costs there may be some empirical lags in price responses induced by trading costs. There appears to be evidence of such lags, correlated with measures of illiquidity. Related to these observations, are a few empirical investigations (e.g.index futures and cash markets) that show an apparent failure of arbitrage pricing. These investigations seem to imply that these examples are related to illiquidity.

Most of the empirical work on illiquidity explores normal times where there are cross section and time series variations in bid-ask spreads. Although this is valuable research in pricing and hedging assets and portfolios, it is of limited applicability in dealing with extreme events that occur in a liquidity crisis. Because a liquidity crisis can arise very quickly, time series analyses of prices, bid-ask spreads etc can induce slow reactions, so that trading and RM systems may be caught out. Clearly more forward looking quantitative and qualitative data, judgement, etc. should be applied in anticipating such disruptions.

4.7 Arbitrage Pricing Bounds with Transaction Costs

In this section we consider bounds induced on buying and selling prices by one-sided arbitrage restrictions. Given these results we are able to use them in more restricted settings to obtain new results on asset pricing with transaction costs. In particular we show that with equi-proportional transaction costs on static hedges, with non-negative positions, we can obtain standard arbitrage pricing implications on mid-prices or mid returns. In other words, standard frictionless pricing methods based on mid-price "approximations" used in the market can be rationalized theoretically.

4.8 General Results

In an economy without transaction costs, assets or portfolios that deliver the same stream of coupons, dividends, or payoffs will have the same market value. This result no longer holds in an economy with transaction costs. But there is an analogue with weak inequalities on Buying (Selling) prices of assets (portfolios)

[13]See Amihud, Mendelson and Pedersen (2006) for a simple exposition of the Acharya and Pedersen (2005) model.

[14]See Amihud, Mendelson and Pedersen (2006) Ch.3, and Subrhamanyam (2009) for surveys of this empirical literature.

with the same income stream. To see this, consider an economy with consumers and brokers. Assume that the consumers do not have a personal transaction technology and the buy-sell price differential is determined by the broker technology. Consider an asset k' and a portfolio of assets $\digamma$ with nonnegative holdings $\alpha_h(\omega_t) \geq 0$, such that at $\omega_t, R_{k'}(\omega_s) = \sum_k \alpha_k(\omega_t) R_k(\omega_s) \equiv R_{\digamma}(\omega_s)$ for all $\omega_s \in S(s \mid \omega_t)$ and $s > t$. This is a simple case where there is a set of assets with future returns that are linearly dependent. From the agent's first order conditions, we deduce immediately that

$$P^i_{k'}(\omega_t) = \sum_k \alpha_k(\omega_t) P^i_k(\omega_t) \equiv P^i_{\digamma}(\omega_t), \text{for every} i \in I;$$

and so,

$$P^B_{k'}(\omega_t) \geq P^i_{k'}(\omega_t) \geq P^S_{k'}(\omega_t).$$

Because $P^B_k(\omega_t) \geq P^i_k(\omega_t) \geq P^S_k(\omega_t)$, then if $\sum_k P^B_k(\omega_t)\alpha_k(\omega_t) < P^B_{k'}(\omega_t)$, we deduce $P^B_{k'}(\omega_t) > P^i_{k'}(\omega_t)$ and it follows that $\delta^B_{ki}(\omega_t) > 0$ and $\Delta^B_{ki} = 0$. That is, no consumer

will buy asset k' at ω_t.

A symmetric argument applies to the selling price. If $\sum_k P^S_k(\omega_t)\alpha_k(\omega_t) > P^S_{k'}(\omega_t)$, then no consumer will sell asset k' at ω_t.

If we reverse the inequalities, we cannot deduce anything about the activity of the asset markets, except to observe that trade in k' is more likely for the obvious reason that if

$$\sum_k P^B_k(\omega_t)\alpha_k(\omega_t) \geq P^B_{k'}(\omega_t) \geq P^S_{k'}(\omega_t) \geq \sum_k P^S_k(\omega_t)\alpha_k(\omega_t),$$

then asset k' provides a cheaper means (in terms of reducing transaction costs) of obtaining the same return.

Conversely, we can show the following inequaltities that will be important in our discussion in below.

For asset k' to have an active market then:

$$P^B_{k'}(\omega_t) \leq \sum_{k \in \digamma} P^B_k(\omega_t)\alpha_k(\omega_t)$$

$$P^S_{k'} \geq \sum_{k \in \digamma} P^S_k(\omega_t)\alpha_k(\omega_t)$$

Assume the converse for the buying inequality, then $P^B_{k'} > \sum_{k \in \digamma} P^B_k(\omega_t)\alpha_k(\omega_t)$. For any i, then $\sum_{k \in \digamma} P^B_k(\omega_t)\alpha_k(\omega_t) \geq \sum_{k \in \digamma} \alpha_k(\omega_t) P^i_k(\omega_t) = P^i_{k'}(\omega_t)$. But this implies

$P^B_{k'}(\omega_t) > P^i_{k'}(\omega_t)$; which in turn implies $\delta^B_{k'i}(\omega_t) > 0$; or $\Delta^B_{ki} = 0$ for any i. Thus no consumer buys k' and the market is inactive, a contradiction.

A symmetric argument applies for the selling inequality.

Note: The inequalities show that an active market for k' requires the two conditions to be necessary conditions, but they are not sufficient for activity.

4.9 Transaction Costs, Arbitrage Pricing and Static Hedging

So far we have assumed that transaction costs may vary widely across different types of securities. Although this can be plausible for certain types of securities, there are other groups of securities where it is reasonable to assume equi-proportional transaction costs. For this section, assume that there are brokers with constant returns to scale technology at each event. With equi-proportional transaction costs $B \leqq (1+\theta)S$ for some scalar $\theta \geq 0$. Note that equality occurs if the market is active, i.e. there is trade in the security. When all securities are active then they have the same spread between buying and selling prices.

The second assumption we make is that our set of securities is sufficiently rich to include spanning sets for each set of successor events. In the zero transaction cost case, the spanning argument merely requires a set of securities that have returns that span the successor events. With transaction costs, we assume that there exists a set of securities with returns such that a nonnegative combination spans the successor events. As a first pass, consider the set of securities to include the full set of Arrow-Debreu securities (we will weaken that strong assumption subsequently). Then we can prove the following result:

Proposition 1 *Given that security k' has an active market, has nonnegative returns, and there exists a full complement of Arrow-Debreu securities with active markets, then:*

$$P_{k'}^{B}(\omega_t) = \sum_{k\imath AD} \alpha_k(\omega_t) P_k^B(\omega_t)$$

$$P_{k'}^{S}(\omega_t) = \sum_{k\in AD} \alpha_k(\omega_t) P_k^S(\omega_t),$$

where AD is the set of Arrow-Debreu securities for all events in ω_{t+1}.

Proof. ■

By the assumption of equi-proportional transaction costs

$$\sum_{k\imath AD} \alpha_k(\omega_t) P_k^B(\omega_t) = \sum_{k\in AD} \alpha_k(\omega_t)(1+\theta) P_k^S(\omega_t)$$

¿From the inequalities in the previous section:

$$P_{k'}^{B}(\omega_t) \leq (1+\theta) \sum_{k\imath AD} \alpha_k(\omega_t) P_k^B(\omega_t) = \sum_{k\in AD} \alpha_k(\omega_t) P_k^B(\omega_t);$$

and

$$P_{k'}^{B} = (1-\theta) P_{k'}^{S} \geq (1+\theta) \sum_{k\imath AD} \alpha_k(\omega_t) P_k^B(\omega_t) = \sum_{k\in AD} \alpha_k(\omega_t) P_k^B(\omega_t).$$

But this implies:

$$P_{k'}^{B}(\omega_t) = \sum_{k \in AD} \alpha_k(\omega_t) P_k^{B}(\omega_t),$$

which proves the first equality.

A similar proof implies the second equality. ‖

What is reasonable about this result is that it mimics the usual frictionless arbitrage argument by providing two spanning price results, one for each buying and selling price. It might be thought that the assumption of the existence of active Arrow-Debreu securities is the key to the result. But it is easy to show the following more general result:

Given that k' has an active market, and this return can be spanned by a nonnegative portfolio of active securities $\mathcal{F}$, then

$$P_{k'}^{B}(\omega_t) = \sum_{k \in \mathcal{F}} \alpha_k(\omega_t) P_k^{B}(\omega_t)$$

$$P_{k'}^{S}(\omega_t) = \sum_{k \in \mathcal{F}} \alpha_k(\omega_t) P_k^{S}(\omega_t)$$

Formally identical to the previous proposition, replacing the set AD with $\mathcal{F}$.

This result provides a key to a chain of results on arbitrage pricing with transaction costs. We make two observations. First, the assumption that asset k' has an active market is important: if asset k' has an inactive market, then it is possible that its buy-sell spread lies outside the replicating bounds and the security is not traded and economically irrelevant.

Second, if we consider the midpoints of the buy-sell prices, then the proposition implies that

$$P_{k'}^{am}(\omega_t) = \sum_{\zeta} \alpha_k(\omega_t) P_k^{am}(\omega_t),$$

where $P_k^{am}(\omega_t) \equiv 1/2(P_k^{B}(\omega_t) + P_k^{S}(\omega_t)), k \in V, k'$. In short, we obtain an analogue of the frictionless arbitrage pricing result for the arithmetic average of the buy-sell prices.

This last observation suggests the possibility that we can extend some standard arbitrage pricing results to allow for proportional transaction costs. The trick is to understand that one must find securities that will span returns so long as the spanning portfolio has nonnegative weights. Note that this result is does not apply to the case where the spanning weights can be positive and negative. In this case (when buying the security) the prices on the right hand side will be buying prices for positive weights and sellling prices for negative weights (i.e.shorting). Another limitation is that the underlying transaction costs are equiproportional. If this is not true, then we can have a situation where the replicating portfolio is less expensive than the asset and there is no trade in the asset. That is:

$$P^B_{k'}(\omega_t) \geq \sum_k P^B_k(\omega_t)\alpha_k(\omega_t) \geq \sum_k P^S_k(\omega_t)\alpha_k(\omega_t) \geq P^S_{k'}(\omega_t).$$

To elaborate: consider a two date example where agents hold no position in either the derivative or the replicating assets. Will the agent trade the derivative? Examination of the inequalities reveal that the agent will not trade the derivative, because its transaction cost is dominated by the replicating portfolio. If we extend the example to three dates where at the intermediate date the agent has inherited a position in the derivative, then the agent will not trade the derivative, perhaps preferring to change the position in the replicating portfolio. In this case the agent is locked into the derivative, but has a partial hedge (because of the spread) in the replicating portfolio. Clearly, this reasoning underlies much of the trading and hedging behavior of traders when faced with different spreads.

4.10 Dynamic Trading and Arbtitrage Bounds with Transaction Costs

Cetin, Jarrow, Protter and Waracha (2006)[15] consider dynamic hedging and pricing for derivatives when there are transaction costs on stock and no trasnaction costs on a bond. The basic idea is quite simple. Assume that an agent wishes to hedge a European call option. To fix ideas, we will assume the finite date/state version of the model. In the standard frictionless model, so long as there are sufficient assets at each event to hedge the subsequent nodes, the asset markets can be dynamically spanned. Standard arbitrage pricing arguments imply the existence of a unique martingale measure (Arrow-Debreu prices) and any future, stochastic payoff can be spanned.

Now assume that the agent faces asset trading costs for a stock that varies with the scale of trading at any event. A second interpretation, is that the trading cost arises from temporary price pressure effects when the agent buys or sells. Assume that there is a riskless bond that trades with no transaction costs. In a variation on the two date version of our discussion above one can see that the trading costs on the stock will induce a bid-ask spread for the replicated derivative. The model must assume some objective for the agent: sometimes it is assumed to be a linear objective using a martingale measure, and at other times super-replication is required. As we argued above in discussing the stochastic linear programing model with hedging, there are problems with these types of objectives as they finesse the larger problem of the agent's portfolio problem.

To fix ideas assume that objective is to minimize a final date, linear function of deviations from the derivative payoff after transaction costs. For simplicity of exposition, assume the riskless rate is zero. Assume the transaction cost function does not vary over time and is not stochastic. Now recalling the first

[15] This paper is one of a series using the same technology - see Jarrow (2007) for an introduction to this literature; and Jarrow and Protter (2008) for a more technical discussion.

order conditions for the agent's problem with personalized transaction costs, and assume that the final random payoff is the deviation of the final portfolio from the derivative payoff. That payoff will depend on the accumulated returns ex transaction costs $Dev_k(\Delta_{ik}^B, \Delta_{ik}^S, \omega_T)$. Assuming that the transaction cost function will be assumed to be zero at zero trades, smooth, increasing with the volume of buying and selling, and the other simplifications, the conditions can be interpreted as saying that contingent prices for the hedge will depend on the value of the expected deviations from the final payoff, and the costs of trading at each event. Consider a simple case, where there is a neighbourhood around zero trading where the transaction costs are zero. Then for a sufficiently "small" derivative position, the transaction costs and final deviation payoffs are zero and the standard costless arbitrage pricing result follows[16]. Conversely, if the optimum has non-zero trades, the transaction cost functions on the buy and sell sides are symmetric, then the first order conditions, after iterating over time imply:

$$\sum_{S(T|0_t)} Dev_k(\Delta_{ik}^B, \Delta_{ik}^S, \omega_T)\tilde{p}_i(\omega_T \mid 0) = \sum_t \sum_{S(\omega_t|0)} MC(\Delta_{ik}(\omega_t))\tilde{p}_i(\omega_t \mid 0)$$

where $MC(\Delta_{ik}(\omega_t))$ is the marginal cost of trading $\Delta_{ik}(\omega_t)$, where a positive number is buying and a negative number is selling. This condition is straightforward to interpret. The current cost of the deviation portfolio has two components: the first is the expected cost of the deviations from a perfect hedge, plus the present, expected value of the transaction costs of the dynamic hedge. This trades off the costs of precision with the transaction costs of producing precision.

Embedding this analysis into a market structure, to obtain bounds on prices, requires additional assumptions. For example, if there were many dealers writing the derivative for other agents, the price will depend upon any transaction cost differentials across the writing agents and their ability to tolerate payoff deviations. In the extreme case where there was a continuum of identical dealers, pricing would be determined competitively. But if there are few dealers, the competitive assumption may not be appropriate and the dealer can extract additional rents. Clearly these issues loom large for dealers writing derivatives with illiquid underlying asset markets.

This model assumes that only a dynamic hedge is feasible. But we know that many traders do not dynamically hedge on illiquid markets, but may take an initial static hedge using derivatives with different strike prices on the underlying. By including this class of hedging, the problem is enlarged to include static and dynamic hedging strategies for agents writing derivatives. The optimal strategy may be a combination of static and dynamic hedging strategies.

Finally observe that in the interpretation of this model where the transaction costs is a temporary price pressure effect, the model is silent on the impact of

[16] In continuous time, this result is mathematically delicate, requiring restrictions on trading processes - see Jarrow and Protter (2008).

other traders trading the underlying security at the same time. This situation requires an analysis of possible strategic trading that we will take up below.

5 Liquidity Model 2: The Agent's Problem With Personal Transaction Costs or Asset Constraints

Let us reintroduce the constraint functions, $F_{i\ell}()$ that can reflect personal transaction costs or constraints on asset trading. For simplicity we will assume that market transaction costs are zero, i.e., $P^B = P^S$. The first order conditions become:

$$\frac{\delta^B_{k^{\cdot}i}(\omega_t)}{\lambda_i(\omega_t)}+\sum_\ell \frac{\delta_{i\ell}(\omega_t)}{\lambda_i(\omega_t)}\frac{\partial F_{i\ell}}{\partial \Delta^B_{ik}(\omega_t)}-P_k(\omega_t)+\sum_{s>t}\gamma_i(s \mid \omega_t)\sum_{S(s|\omega_t)} R_k(\omega_s)\tilde{p}_i(\omega_s \mid \omega_t)=0.$$

$$\frac{\delta^S_{ki}(\omega_t)}{\lambda_i(\omega_t)}+\sum_\ell \frac{\delta_{i\ell}(\omega_t)}{\lambda_i(\omega_t)}\frac{\partial F_{i\ell}}{\partial \Delta^S_{ik}(\omega_t)}+P_k(\omega_t)-\sum_{s>t}\gamma_i(s \mid \omega_t)\sum_{S(s|\omega_t)} R_k(\omega_s)\tilde{p}_i(\omega_s \mid \omega_t)=0.$$

Clearly these conditions imply that either $\Delta^B_{ki}(\omega_t) \geq 0$ or $\Delta^S_{ki}(\omega_t) \geq 0$, but we cannot have $\Delta^B_{ki}(\omega_t) > 0$ and $\Delta^S_{ki}(\omega_t) > 0$. Furthermore, we can define personalized buying and selling prices for asset k at event ω_t by:

$$\sum_\ell \frac{\delta_{iR}(\omega_t)}{\lambda_i(\omega_t)}\frac{\partial F_{i\ell}}{\partial \Delta^B_{ik}(\omega_t)} - P_k(\omega_t) \circeq -P^{Bi}_k(\omega_t).$$

$$\sum_\ell \frac{\delta_{i\ell}(\omega_t)}{\lambda_i(\omega_t)}\frac{\partial F_{i\ell}}{\partial \Delta^S_{ik}(\omega_t)} + P_k(\omega_t) \circeq -P^{Si}_k(\omega_t).$$

Given these definitions, then we can analyze the agent's problem as we did above, recalling that the personalized buying and selling prices are evaluated at the optimum. We could reintroduce market-based transaction costs, $P^B_k(\omega_t) > P^S_k(\omega_t)$,so that the personalized transaction costs would include the personalized plus the market transaction cost. As the reader can check, this is an easy extension of our analysis. Later we will provide examples of this class of model in the literature.

Example 1: Restricted Access

Basak and Cuoco (1998) have explored a exchange economy with just two investors and two assets: a riskless bond, and a risky asset. One investor is restricted from trading in the risky asset, but the other investor trades both securities. It follows immediately from market clearance that the unrestricted investor must bear all the risk; and given that her consumption is a fraction of total consumption, one can justify a higher risk premium than the unrestricted economy.

Using our framework, we can consider Basak and Cuoco's model as a discrete event/date special case of our framework. For the unrestricted investor, the first order conditions collapse to the standard martingale condition. But for the restricted investor, the bond equation is the unrestricted condition, and the risky asset has personalized transaction costs set sufficiently high that no participation is warranted.

Example 2: Short Selling Constraints

A second interpretation of our conditions assumes that the constraint functions represent short selling constraints, VaR constraints or other regulatory constraints. For example, we can introduce a variation on the personalized transaction costs idea to incorporate short selling constraints or more complicated constraints associated with dynamic strategies. (Notice that the set T_i can allow for more general constraints than our current constraints representation, $F_{i\ell}() \geq 0$.) To illustrate the idea, consider a short-selling constraint on long-lived asset K, but no constraints on any other asset. Furthermore, assume $P_k^B(\omega_t) = P_k^S(\omega_t) \forall \omega_t, \forall k$. Given that the consumer begins at ω_s with no endowments of the asset (non-negative asset endowments are a trivial extension), then a short-sale constraint at ω_t is represented by

$$\sum_{\mathbf{P}(\omega_t)} (\Delta_{Ki}^B(\omega_t) - \Delta_{Ki}^S(\omega_t)) \geq 0$$

where $\mathbf{P}(\omega_t)$ is the chain of predecessor events (including ω_t) stretching from ω_s to ω_t. Clearly these constraints have more complex interactions than our previous event constraints. The consumer's Lagrangian for this problem is:

$$L_i = U_i(R[\Delta_i^B - \Delta_i^S] + P[\Delta_i^S - \Delta_i^B] + \bar{x}_i) + \sum_k \delta_{ki}^B \Delta_{ki}^B + \sum_k \delta_{ki}^S \Delta_{ki}^S + \sum_{\omega_t} \delta_K(\omega_t) \sum_{\mathbf{P}(\omega_t)} (\Delta_{Ki}^B(\omega_t) - \Delta_{Ki}^S(\omega_t)).$$

The first order condition for asset K for this problem are:

$$\frac{\delta_{Ki}^B(\omega_t)}{\lambda_i(\omega_t)} + \sum_{s>t} \sum_{S(s|`\omega_t)} \frac{\delta_{Ki}(\omega_s)}{\lambda_i(\omega_t)} - P_K(\omega_t) + \sum_{s>t} \gamma_i(s \mid \omega_t) \sum_{S(s|\omega_t)} R_K(\tau_s) \tilde{p}_i(\omega_s \mid \omega_t) = 0$$

$$\frac{\delta_{Ki}^S(\omega_t)}{\lambda_i(\omega_t)} + \sum_{s>t} \sum_{S(s|\omega_t)} \frac{\delta_{Ki}(\omega_s)}{\lambda_i(\omega_t)} + P_K(\omega_t) - \sum_{s>t} \gamma_i(s \mid \omega_t) \sum_{S(s|\omega_t)} R_K(\omega_s) \tilde{p}_i(\omega_s \mid \omega_t) = 0$$

or regrouping them we obtain the interpretation:

$$\frac{\delta_{Ki}^B(\omega_t)}{\lambda_i(\omega_t)} - P_K(\omega_t) + \sum_{s>t} \gamma_i(s \mid \omega_t) \sum_{S(s|\omega_t)} [R_K(\omega_s) \tilde{p}_i(\omega_s \mid \omega_t) + \frac{\delta_{Ki}(\omega_s)}{\lambda_i(\omega_t)}] = 0$$

$$\frac{\delta_{ki}^S(\omega_t)}{\lambda_i(\omega_t)} + P_K(\omega_t) - \sum_{s>t} \gamma_i(s \mid \omega_t) \sum_{S(s|\omega_t)} [R_K(\omega_t)\tilde{p}_i(\omega_s \mid \omega_t) + \frac{\delta_{Ki}(\omega_s)}{\lambda_i(\omega_s)}] = 0.$$

¿From these two conditions it follows easily that the consumer will not buy and sell asset K simultaneously at ω_t .But if the consumer buys (or sells) at ω_t, the marginal condition is:

$$P_K(\omega_t) = \sum_{s>t} \gamma_i(s \mid \omega_t) \sum_{S(s|\omega_t)} [R_K(\omega_s)\tilde{p}_i(\omega_s \mid \omega_t) + \frac{\delta_{Ki}(\omega_s)}{\lambda_i(\omega_t)}].$$

That is, the price of asset K at ω_t must equal the personalized discounted expected return, where the return is adjusted for the Kuhn-Tucker multipliers associated with possible future short-sale binding constraints. It is possible to rewrite this condition in a more recognizable form, by defining

$$\tilde{\delta}_{Ki}(\omega_s) \equiv \frac{\delta_{Ki}(\omega_s)}{\lambda_i(\omega_t)\tilde{p}_i(\omega_s \mid \omega_t)},$$

so that we obtain:

$$P_K(\omega_t) = \sum_{s>t} \gamma_i(s \mid \omega_t) \sum_{S(s|\omega_t)} [R_K(\omega_s) + \tilde{\delta}_{Ki}(\omega_s)]\tilde{p}_i(\omega_s \mid \omega_t).$$

This equation gives a standard personalized martingale interpretation for pricing asset K, where future returns include a "yield" from a marginal relaxation of future short-sale constraints, $\{\tilde{\delta}_{Ki}(\omega_s)\}$. From standard non-linear programming theory we know that the Kuhn-Tucker multipliers can be interpreted as the marginal utility of a small relaxation of their associated constraint at the optimum (i.e. the envelope theorem). Thus $\tilde{\delta}_{Ki}(\omega_s)$ is the personalized undiscounted "price" of relaxing the short-sale constraint for the event. Thus any increase in the holding of asset K will have a future benefit in relaxing every contingent binding constraint.

The General Technology: VaR, Margin and other Constraints on asset Trading

The argument on short selling constraints can be generalized to more complicated constraints on asset trades. For example there could be dynamic constraints on asset trades that arise from legal or risk management considerations, or margin and collateral constraints. Because there are many different types of constraints (we will not do the formal modelling here) we will argue that the general principles are straightforward. Any dynamic constraint that involves accumulated trading positions (our short-selling example is a simple case) will have dynamic multipliers that will signal convenience yields for slackening future constraints. A good example of this type of constraint would be VaR constraints which are a sequence of prespecified Baysian updated constraints on asset returns and trading strategies. Other examples are various margin requirements where in certain market situations, margin or collateral constraints are tightened. In our model, they could be modelled by having constraints that tightened on asset holdings in certain states of the world, certain constellations of asset prices etc. Intuitively, asset that are not constrained will sell at a price premium compared to assets that are constrained. But, as we observe from the first order conditions, those situations should be incorporated into the optimal solution and appear through the pricing equations in the multiplier terms. Thus the way to think of these contingent constraints, is to assume that they are anticipated in the optimal strategy and incorporated in asset prices.

It is clear that by using the stochastic linear programming approach discussed above in a previous section, that one can construct large dimensional LP problems that consider portfolio strategies for any class or classes of securities with linear legal or risk management constraints[17]. Any realistic program will be of considerable dimension and require sensitivity analysis in specifying asset return processes and the constraint sets. In addition one can add transaction costs to obtain a program that would combine trading and risk management constraints as well as bid-ask spreads that mirror periods of contingent "low" (wide bid-ask spreads) or "high" (narrow bid-ask spreads) liquidity.

5.1 Equilibrium Examples

Given the flexibility of the model and the requirement that agents differ in characteristics to induce trade, the equilibrium problem and the implications for asset prices is in general difficult to solve and characterize, except for special, highly parameterized cases. We will give a brief discussion of some models that use this approach.

5.2 Equilibrium Asset Pricing with VaR Constraints and Margin Requirements

In a recent paper, Danielsson and Zigrand (2008) have a simple variation on the three date incomplete market economy with VaR constraints on the FI's problem. Given a highly parameterized model, they are able to characterize the equilibrium and prices. Because the VaR constraint set is a function of asset prices it is possible that no equilibria exists[18]. They explore the situation where the constraint sets are sufficiently regular to allow an equilibrium, then they characterize the allocation and prices. Because the economy is a second best allocation it is not surprising that they find apparently counter-intuitive results:

[17] It is well-known that non-linearities can be accomodated in linear progams by piece-wise linear approximations.

[18] This problem can occur in any abstract general equilbrium model where constraint sets - other than the linear budget constraint - are functions of prices.

weakening these constraints may reduce welfare by increasing risk premia, illiquidity and volatility in periods of systemic risks (modelled as correlated shocks to agents asset demands).

Similar effects can be obtained by imposing margin requirements on asset holdings. In this case, general falls in risky asset prices will trigger increased margin requirements. In turn this can reduce agent holdings of risky assets, leading to further falls in prices. Because the constraint includes asset prices, then the usual stabilizing effect of falling prices increasing demand, can be overturned by binding asset constraints that imply falling prices will lead to increased asset sales, and reinforce a decline in asset prices[19].

This idea is easy to illustrate in our more general structure. To keep the argument as simple as possible consider a two date example of the constrained agent model. It is well known[20] that for the unconstrained case, the asset demand functions for an agent with standard neoclassical preferences (the standard risk averse von Neumann Morgenstern utility functions satisfies these conditions) will have downward sloping demand curves for assets so long as the assets are normal goods in the conventional demand theory sense. But with binding constraints which are a function of asset prices, this result can be overturned. With a binding constraint, a reduction in the price of an asset will have two effects: first, there will be the standard demand theory response showing an increase in the demand for that asset; and the second effect is the impact via the constraint where a decline in the prices can force a decrease in the demand for the asset. The net effect can lead to a decline in the demand for the asset and a increase in holdings for other assets. This apparently paradoxical result can lead to large falls in an asset price as margin and other constraints force selling. One can construct extreme theoretical examples with constraints where no equlibrium price exists.

5.3 Summary of the Asset Constraint Approach

This type of model has some interesting features in that they mimic FI asset strategies, financial market architecture and provide a role for FI's. They can mimic infrequent trading in certain markets and the demand for liquid assets when the FI's portfolio has illiquid assets and random exposures or calls for cash. It can mimic the difficulties in hedging derivatives and the reason why derivative pricing may not provide a unique valuation, but be limited to a bid-ask spread.

This second best economy can generate seemingly perverse welfare results. We know from the incomplete markets literature that weakening constraints may make agents worse off. Therefore, it should not be difficult to construct examples, in this constrained economy, that generate seemingly perverse welfare results. This example should be a lesson for those who recommend regulatory constraints or activity without thinking carefully about agent and market reactions.

[19]See Brunnermeier and Pedersen (2008) for a model along these lines; and Adrian and Shin (2008) for an informal discussion.

[20]See Milne (2003), Chapter 3.

One of the weaknesses of the transaction cost model is that it does not address the fundamental causes of transaction costs or non-regulatory asset constraints and their evolution. It is a reduced form model that mimics more fundamental problems in search and asymmetric information. For example margin requirements arise because the positions of the borrower or sort seller are obscure to the lender. Therefore they impose margin or collateral to safeguard their investments. A more complete model would include these deicisions as component of the model. Reflection reveals that this will introduce game theoretic components to the model, where lenders can induce changes in behavior in lenders. We will discuss this type of model below.

The model assumes a given bid-ask spread, where trading does not impact on prices i.e. asset markets are competitive. Furthermore if asset markets are complete, and there is a rich supply of derivative assets, then asset constraints can be circumvented, and the resulting economy is effectively complete and Pareto Optimal. A realistic model must observe transaction costs and constraints as reduced forms, representing fundamental reasons why markets are not complete and/or distorted.

This approach can be a useful approximation in some asset markets; but in "thin" non-competitive markets it is a poor approximation to bilateral or non-competitive trading where price making is important. That is the topic of the next section.

5.4 Liquidity Model 3: Funding Liquidity as Constraints on Sources of Funding

5.5 Liquidity Model 4: Liquidity as a Pecuniary Externality

5.6 A Basic Risk Management Model with Liquidity

The third strategy for modelling liquidity is to assume that the FI can move prices as they trade. In other words, the FI has market power. It is well-known that in markets for small stocks, trading can move prices. But this phenomenon occurs in more active markets when a large trader buys or sells large quantities of the security. Many FI's are reluctant to discuss this price-making power as it would bring them under the scrutiny of security regulators looking for market manipulation of prices. One suspects that the more neutral term "liquidity" is a disguise for this price-making power.

To begin, assume that we have the basic model of an FI risk management system with one modification - the FI can move prices when it trades. (We will assume for simplicity that there are no bid-ask spreads or other transaction costs.) Assume that the asset price matrix in our model is a function of the FI's asset trades:

$$\max V_h(x_h)$$

$$x_h \leq R(\overline{a}_h + \Delta_h) - P(\Delta_h)\Delta_h + C.$$

We assume that the stochastic price matrix $P(\Delta_h)$ is a function of the trading strategy. (For the moment we will ignore the situation where prices are a function of other traders' strategies.) The price function will be assumed to have certain properties. First it will be assumed that over the set of feasible trades, the price function will not allow arbitrage profits (i.e. There does not exist a trading strategy Δ_h^* such that $R\Delta_h^* - P(\Delta_h^*)\Delta_h^* \geq 0$ and one component of the vector is strictly positive). Secondly, it is possible to model the price process such that trades trigger a transitory price movement, or a permanent price movement. (The implications for optimal trades for these two different specifications are quite different.) For simplicity, we could assume that trading in asset k at event ω_t only moves that asset price at ω_t in response to a contemporaneous demand curve. This assumption could be relaxed by assuming that there is an impact on the prices of related assets. (This is plausible if the second asset is a derivative, normally priced by arbitrage.) Also there can be an underlying random price process for prices even though the FI does not trade[21]. As we observe in a previous section there are results on the trading strategies for such a problem. Nevertheless, we can intuit what type of results can be obtained in simple examples. There will be a tendency to spread out trades to minimize the impact of price pressure from trading. FI's that can be subject to large exogenous, random exposures will be less likely to hold illiquid assets; or if they do, will hold a cushion of assets that are relatively liquid (i.e. can be traded with little price impact). One would expect that for agents to hold an illiquid asset, all other payoff characteristics the same, the agent will require a lower price, or a return premium.

It is not difficult, in principle, to add other constraints to the FI's problem to incorporate VaR constraints etc.. Indeed, it is possible to allow for bid-ask spreads on asset prices and price-making to make the liquidity model more complete. As far as I am aware there has been little work in this area of detailed modelling of liquidity in the FI's Risk Management problem.

5.7 Pecuniary Externalities and Risk Management

The problem above ignores the possibility that other agents' trading strategies can impact on the prices. To accommodate that situation, the problem is modified so that there are a set of traders, such that each trader explicitly observes the trading strategies of the other traders. This game will be sensitive to the timing of trades and the assumption of symmetric or asymmetric information. To make the points as simple as possible, assume simultaneous trading at ω_t and

[21]Both cases have been analysed in a simple model of arbitrage option pricing in a series of papers by Jarrow, Protter and their co-authors. For a survey of this literature see Jarrow and Protter (2008). See also more recently, Cetin, Jarrow, Protter and Warachka (2006) and Jarrow (2007).

symmetric information. A number of papers make this assumption and close the model by assuming a contingent residual excess demand curve resulting from a competitive fringe of price taking agents[22].

$$\max V_h(x_h)$$

$$x_h \leq R(\overline{a}_h + \Delta_h) - P(\Delta_h; \Delta_{-h})\Delta_h + C.$$

where $P(\Delta_h; \Delta_{-h})$ is the stochastic price process which is a function of agent $h's$ contingent trades and the trading strategies of the other players. In general this game can be analyzed as a subgame perfect equilibrium. Of course the price functions will require restrictions to ensure existence, added restrictions will ensure uniqueness, and additional assumptions to aid tractability in characterizing the equilibrium.

It has been shown by Brunnermeier and Pedersen (2005), in a continuous time set-up, that a distressed trader can be predated by other traders. It is clear that this type of structure will allow a range of Industrial Organisation type modelling results to be obtained[23]. In an infinite horizon, repeated game version of the model one can exploit Folk Theorem type results to obtain periods of cooperation in providing liquidity between FI's; followed by defections and punishment phases[24]. This type of analysis captures long periods of cooperation, punctuated by much shorter periods of predation and liquidity crises.

One extension of this type of model is to explore the role of regulation and central banks in providing liquidity support via the interbank market for a distressed FI, or to FI's in general[25]. It is clear that there are interesting variations on this type of analysis, which raise serious issues of predation, restrictive practices and anti-trust issues where the predation may be motivated in a time of crisis to knock out a competitor in an oligopolistic financial sub-market, or to enhance a takeover or a "consolidation" of a subset of the financial industry.

As far as I know, there has been no serious attempt to incorporate this type of analysis into Risk Management models or FI strategic planning in any systematic fashion. Given the highly sensitive nature of this type of analysis, one would expect that any such analysis would not appear in the public domain.

6 Systemic Risks, Liquidity and Regulation: The Role of Simulations

There is a recent literature that simulates banking crises through interbank exposures. These models have been developed to describe contagion in a banking

[22]Some papers assume continuous time set-ups, but there will be discrete time/state analogue results. See Pedersen (2010) for a survey of this class of model.

[23]Brunnermeier and Pedersen (2005) provide a quick sketch of several possible extensions.

[24]See Carlin, Lobo and Viswanathan (2007).

[25]See Acharya, Gromb and Yorulmazer (2008).

system where a shock to the asset position in one bank spreads via exposures to other banks in the system. This theory explores the degree and pattern of exposures between banks to show that contagion can be contained or will spread through the system. The work of Elsinger, Lehar and Summer (2006), is a prototype of this type of analysis. They introduced a model where a shock in one bank would have an impact on the exposures of other banks. Using data on aggregate bank exposures, and reserves, they used an iterative algorithm to track possible default chains. Having noted the initial impact on an FI, the next round checks via an iterative procedure to see if there is another bank failure. If not, the failure is contained. If another bank or banks fail, then the process iterates again, and so on until there are no further bankruptcies. The procedure has been run as an empirical model and some central banks have been exploring this methodology to test for possible contagion.

The Bank of England[26] and other central banks have been running variations on this type model. Clearly there are serious attempts to improve and extend these models, but they have some serious limitations that require modifications and extensions[27]. One of the limitations of this type of simulation is the restrictive behavioral assumptions on bank asset strategies. In reality, any FI that suffers a serious loss will respond by altering their trading, lending and capital strategies. Furthermore, FI's that have exposure to the stressed FI will alter their behavior. In turn this can impact adversely on asset valuations so that contagion can spread to an FI that has no exposures to the distressed FI or set of FI's. Another limitation is the use of aggregate data that blurs important characteristics of asset default probabilities and loss given default. This is important for illiquid securities that may take some time to sell or be held to maturity. There will be progress in these types of macro simulations, but I would like to propose a more micro simulation as a complementary stress test.

7 War Games and Simulations using FI Risk Management Models and Strategies

As a complement to the macro style models, one possibility is to exploit the detail that is embedded in the RM systems of large FI's. Large FI's are important strategically in many financial markets, often playing crucial roles in trading. They can contain, or propagate, serious counterparty risks with other large players. Thus the detail of their trading activities, especially when there is financial innovation, can play an important role in helping to analyze their role in propagating systemic risks in the financial system.

One possible strategy in analyzing these types of situations, would be to exploit the FI's RM systems in simulating a shock. This could be regarded as a system wide shock that impacted on all the main players. In a real time

[26]For example see Allessandri, Gai, Kapadia, Mora and Puhr (2008)

[27]See Upper (2007) for a summary of the methods and a critique.

game, the FI's could determine their portfolio rebalancing and other decisions. Then the RM systems would be run again after making assumptions on asset price adjustments given the order flows. This game would require that regulators and central banks play a key role in the game. FI's will be reluctant to reveal their strategies to competitors, but the regulator is central in seeing all trading strategies for all the FI's. Such a game would be a non-trivial exercise and would require careful and repeated playing. These games could be played regularly and treated as part of an educational process, testing the RM systems of national FIs and sharpening the regulator's responses to possible scenarios. The games would not be able avoid all systemic risks (some scenarios may be difficult to anticipate), but they should be carefully designed to think through the implications of possible serious events that do not occur often, new but possible risky scenarios, stress test the implications of new financial products, consider the impact of financial distress from unregulated vehicles and FI's, and think through new regulations and industry responses. The reports of such games should be collected and made available by appropriate national and international regulators. They should be available for research and RM teaching.

Given the international nature of financial markets, there is a clear requirement that national regulators be linked so that cooperation can act smoothly in an international financial crisis. Such organizations and cooperation already exist, but the current crisis will be testing their efficiency. National financial games could be linked between major financial centers to track international FI's. This structure would rehearse international workouts to test regulatory inconsistencies and legal complexities that arise from international transactions and workouts[28].

8 Conclusion

This paper has summarized the basic model underlying modern Risk Management systems. It outlined theoretical deficincies with the model[29]. Using the current credit crisis as a stress test of the standard RM approach, we suggested that some theoretical modifications may help in dealing with liquidity issues that are omitted from the basic model. Also the role of regulators and intervention are related to these more elaborate models, attempting to bring the RM model more in line with the insights of market microstructure theory, banking theory, liquidity and financial crisis models.

[28]See Kubarych (2001) for a discussion of a large strategic and financial game played in the US.

[29]For a dicussion of the empirical problems that are encountered in implementing RM systems, see Milne (2009).

References

[1] Acharya, V., D.Gromb and T. Yorulmazer (2008), "Imperfect Competition in the Interbank Market for Liquidity as a Rationale for Central Banking", Working Paper.

[2] Acharya, V. and L.H. Pedersen (2005), "Asset pricing with liquidity risk", Journal of Financial Economics, Elsevier, 77(2), 375-410.

[3] Adrian, T. and H.Shin,(2008), "Financial Intermediaries, Financial Stability and Monetary Policy", August 2008, Jackson Hole Meetings, Federal Reserve Bank of Kansas City.

[4] Alessandri, P., P. Gai, S. Kapadia, N. Mora and C. Puhr (2008) "A Framework for Quantifying Systemic Stability", Working Paper.

[5] Allen, F. and D.Gale (2007), Understanding Financial Crises, Oxford University Press.

[6] Amihud, y., H.Mendelson and L.Pedersen (2006) Liquidity and Asset Prices, Now Publishers Incorporated, Boston.

[7] Basak, S. and Cuoco, D. (1998). "An equilibrium Model with Restricted stock Market Participation", The Review of Financial Studies, 11, vol 2, 309-402.

[8] Black, F.and M.Scholes (1973), "The Pricing of Options and Corporate Liabilities", Journal of Political Economy, 3, pp.637-654.

[9] Brunnermeier, M. and L. Pedersen (2005), "Predatory Trading", Journal of Finance, 60(4), pp. 1825-1863.

[10] Brunnermeier, M. and L.Pedersen (2008), "Market Liquidity and Funding Liquidity", forthcoming *Review of Financial Studies.*

[11] Carlin, B., M.Lobo and S. Viswanathan (2007), "Episodic Liquidity Crises: Cooperative and Predatory Trading", Journal of Finance, Vol LXII, No. 5, 2235-2274.

[12] Cetin, U, R. Jarrow, P.Protter and M.Warachka (2006), "Pricing Options in an Extended Black Scholes Economy with Iliquidity: Theory and Empirical Evidence",Review of Financial Studies, 19, n.2, 493-529.

[13] Constantinides, G. (1979). "Multiperiod Consumption And Investment Behavior with Convex Transaction Costs," Management Science, 25,1127-1137.

[14] Dai, M. H. Jin, H. Lui (2008) "Illiquidity, Portfolio Constraints and Diversification" Working Paper.

[15] Danielsson, J. (2002), "The Emperor has no clothes: Limits to risk modelling", Journal of Banking and Finance, 26, 1273-1296.

[16] Danielsson, J., B. Jorgensen, C de Vries and X. Yang (2008) "Optimal Portfolio allocation under the probabilistic VaR constraint and incentives for financial innovatio", Annals of Finance, 4; 345-367.

[17] Danielsson, J. and J-P Zigrand (2008) "Equilbrium Asset Pricing with Systemic Risk", Economic Theory, 35; 293-319.

[18] Davis, M.H. and Norman, A.R. (1990). "Portfolio Selection with Transaction Costs," Mathematics of Operations Research , 15, 676-713.

[19] Duffie, D. and K.Singleton (2003), Credit Risk: Pricing, Measurement and Management, Princeton University Press.

[20] Edirisinghe, C., Naik, V. and Uppal R. (1993). "Optimal Replication of Options with Transaction Costs," Journal of Quantative and Financial Analysis,28, 117-138.

[21] Elsinger, H., A. Lehar and M.Summer (2006), "Risk Assessment for Banking Systems", Management Science, 52, pp.1301-1314.

[22] Frauendorfer, K. and M. Schurle (2007), Dynamic modelling and optimization of non-maturing accounts, Wiley.

[23] Freixas, X. and J-C Rochet (2008), The Microeconomics of Banking, MIT Press. Second Edition.

[24] Hart, O. (1975) "On the Optimality of Equilibrium when the Market Structure is Incomplete," Journal of Economic Theory 11 (December 1975) 418-443.

[25] Jarrow, R. and S. Turnbull (1995), "Pricing Options on Financial Securities Subject to Credit Risk", Journal of Finance, 59, pp.53-85.

[26] Jarrow, R. and P. Protter (2005), "Liquidity Risk and Option Pricing Theory", Cornell Working Paper.

[27] Jarrow, R. and P. Protter (2008), "Forward and Futures Prices with Bubbles", Cornell Working Paper.

[28] Jin, X. and Milne, F. (1999). "The Existence of Equilibrium in a Financial Market with Transaction Costs," in M. Avellanda, (ed.) Quantitative Analysis in Financial Markets, World Scientific Publishers.

[29] Jouini, E. and Kallal, H. (1995). "Martingales and Arbitrage in Securities Markets with Transaction Costs," Journal of Economic Theory, 66, 178-197.

[30] Karatzas, I. and Shreve,S. (1998). Methods of Mathematical Finance, Springer-Verlag.

[31] Kubarych, R.(2001), Stress Testing the System: Simulating the Global Consequences of the Next Financial Crisis, Council on Foreign Relations Press.

[32] Lando, D. (2004), Credit Risk Modeling, Princeton University Press.

[33] Liu, H. (2004) “Optimal Consumption and Investment with Transaction Costs and Multiple Risky Assets”, Journal of Finance, Vol LIX, No.1 289-338.

[34] Liu, H, and M.Lowenstein (2007) “Optimal Portfolio Selection with Transaction Costs and ”Event Risk”” Working Paper.

[35] Magill, M. and Quinzii, M. (1996). Theory of Incomplete Markets, Cambridge University Press.

[36] Meissner, G. (2005), Credit Derivatives: Application, Pricing and Risk Management, Blackwell publishing.

[37] Merton, R.(1973), “A Theory of Rational Option Pricing”, Bell Journal of Economics and Management Science, 4(1), pp.141-183.

[38] Milne, F.(2003), Finance Theory and Asset Pricing, Oxford University Press.Second Edition.

[39] Milne, F.(2009), “The Complexities of Financial Risk Management and Systemic Risks”, Bank of Canada Review, pp.15-29.

[40] Milne, F. and E. Neave (2003) “A General Equilibrium FinancialAsset Economy with Transaction Costs and Trading Constraints”, Working Paper No.1082, Economics Dept.Queen’s University.

[41] Milne, F. and H.Shefrin (1986), “Information and Securities: Pareto Dominance and the Second Best”, Journal of Economic Theory, 43:2, pp.314-328.

[42] Naik, V. (1995). ”Finite State Securities Market Models and Arbitrage”, Ch.2 in R. Jarrow et al, (eds.) Finance, Handbook on OR and MS, Vol 9,Elsevier Science.

[43] Subrahmanyam, A. (2009), ”The Implications of Liquidity and Order Flows for Neoclassical Finance”, Pacific Basin Finance Journal, 17, 527-532.

[44] Upper, C. (2007), “Using Counterfactual Simulations to Assess the Danger of Contagion in Interbank Markets.” BIS Working Paper no.234.