



Queen's Economics Department Working Paper No. 1082

A General Equilibrium Financial Asset Economy with Transaction Costs and Trading Constraints

Frank Milne
Queen's University

Edwin Neave
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

9-2003

A General Equilibrium Financial Asset Economy with Transaction Costs and Trading Constraints*

Frank Milne[†] and Edwin H. Neave
School of Business and Department of Economics
Queen's University Kingston
K7L 3N6 Ontario, Canada

September 5, 2003

Abstract

This paper presents a unified framework for examining the general equilibrium effects of transactions costs and trading constraints on security market trades and prices. The model uses a discrete time/state framework and Kuhn-Tucker theory to characterize the optimal decisions of consumers and financial intermediaries. Transaction costs and constraints give rise to regions of no trade and to bid-ask spreads: their existence frustrate the derivation of standard results in arbitrage-based pricing. Nevertheless, we are able to obtain as dual characterisations of our primal problems, one-sided arbitrage pricing results and a personalised martingale representation of asset pricing. These pricing results are identical to those derived by Jouini and Kallal (1995) using arbitrage arguments. The paper's framework incorporates a number of specialised existing models and results, proves new results and discusses new directions for research. In particular, we include characterisations of intermediaries who hold optimal portfolios; brokers who do not hold portfolios, and consumer-specific

*We have benefited from comments from seminar participants at the University of Chicago, Carnegie-Mellon, London Business School, Queen's University, and the University of British Columbia. This research has been supported by the Canadian Social Sciences and Humanities Research Council

[†]Correspondence: Frank Milne Economics Dept , Queen's University Fax: (613) 533-6668; email: milnef@qed.econ.queensu.ca

transactions costs and trading constraints. Furthermore we show that in the special case of equiproportional transaction costs and a sufficient number of assets, there is an analogue of the arbitrage pricing result for European derivatives where prices are interpreted as mid-prices between the bid-ask spread. We discuss the effects of non-convex transaction technologies on prices and trades.

Introduction

Financial theorists have recently shown renewed interest in incorporating transaction costs and trading constraints into the analysis of financial market trading and asset pricing. In particular Jouini and Kallal (1995), Edirisinghe, Naik and Uppal (1993), and Luttmer (1996), with other researchers, explore the implications of transaction costs for asset prices and hedging possibilities.

Since much of the literature employs restrictive approaches in analyzing applied problems, we see a demand for a unified framework to encompass these more specialized models. This paper constructs a multiperiod general equilibrium framework where consumers and financial intermediaries trade financial assets with frictions. Our discrete-time-state approach uses General Equilibrium theory and finite dimensional non-linear programming to analyse the effects of transaction costs and trading constraints on asset pricing and hedging possibilities in a competitive economy. We demonstrate that a number of recent models are special cases of our general structure.

The simplest version is the degenerate case where there are no asset transaction costs or trading constraints on any conceivable asset. Then the model collapses to the familiar frictionless asset model - see Milne (1995), Naik (1995), Magill and Quinzii (1996) and Pliska (1997). If there are complete markets, then lack of arbitrage opportunities implies a unique martingale, or Arrow Debreu pricing condition for derivative securities. Incomplete markets can be seen as a step toward reality, by assuming that transaction costs or trading constraints restrict trades to a frictionless subset of asset markets. Both of these models are subsets of our general model with non-trivial transaction costs.

One version of this model allows intermediaries to trade assets between buyers and sellers using a costly technology. Given that these brokers do not hold portfolios, their behaviour implies the existence of a technology-induced, bid-ask spread in asset prices that can evolve stochastically over time. This technology can be thought of as a simplified reduced form approximation of a market microstructure that incorporates strategic choices in ordering and executing trades.

A second version allows the consumer to operate a wholly-owned transaction technology, which implies consumer-specific transaction costs. The bid-ask spread between the consumer and its transaction technology are now unobservable shadow prices deduced from the consumer's optimizing problem. A special case of this model is where the consumer or intermediary has constraints on trading. For example we could have legal restrictions on portfolio holdings of foreign assets, short-selling constraints or Value at Risk constraints on portfolio positions. A variant of this model examines properties of the induced preferences that result from optimising over the technology. We use this interpretation in considering some recent literature. For example, one version allows for personal costly trading.

A third version allows an intermediary, or dealer, to construct portfolios and securities while operating between the buying and selling prices. This construction allows us to consider banks or other financial intermediaries that operate to transform securities through repackaging of returns. A good example would be firms writing exotic options, who (partially) hedge using portfolios of other securities.

We postulate two classes of agents: consumers and financial intermediaries. Given consumers have smooth quasi-concave preferences and a convex technology then standard Kuhn-Tucker theory provides necessary and sufficient first order conditions for an optimum. As a special case of this model we obtain a dual linear programming model of dynamic portfolio choice with transaction costs (for a detailed analysis of a similar case see Edirisinghe, Naik and Uppal (1993)).

Intermediaries are treated similarly to consumers. It is well-known that the Fisher separation theorem can fail in incomplete asset markets - a special case of our model. Therefore we assume that each intermediary has a utility function which represents the preferences of the single entrepreneur/owner. Given the assumption of differentiable, quasi-concave utility and a convex technology, we have a standard production problem that allows us to characterise the intermediary's optimal production and dynamic portfolio strategy.

Transactions costs restrict arbitrage possibilities, and in their presence standard arbitrage arguments fail. However, with positive transactions costs, we can still derive upper and lower bounds on asset prices and trades which reduce to standard arbitrage conditions when transactions costs converge to zero. From the first order conditions for a maximum we deduce a personalised martingale characterisation of asset pricing such that any agent's personal discounted valuation lies between the bid-ask spread. This result is a discrete version of an infinite dimensional result obtained by Jouini and Kallal (1995) who used arbitrage arguments. We show how this result on personalised martingale pricing extends to personal production technologies, including the special

case of asset trading constraints. This extension of the model provides a unified treatment of optimal portfolio choice with transaction costs, a variety of constraints, and the dual pricing characterisation at the optimum.

Combining the agent models in a competitive market clearing model, we address a number of issues. It is easy to see that one can construct rationalisations of asset market architecture and intermediary clusters by appropriately choosing transaction costs and constraints in asset trades. A more difficult problem is characterising asset prices in such an economy. There are a small number of highly restrictive models that attempt such characterisations. We discuss these models and provide some cautionary comments on deducing general or robust results in more general versions of the model.

We are able to demonstrate a new and useful result on asset pricing with transaction costs. Using one sided arbitrage arguments, and assuming equi-proportional transaction costs over a sufficient number of securities in the market, we deduce an exact arbitrage pricing result that is applied to buying and selling prices. This result is the keystone for providing tight restrictions on buying and selling prices for derivative securities. In particular we show that we can deduce an exact pricing result: by taking the mid-prices of bid-ask spreads we can deduce any discrete, frictionless arbitrage pricing result in the literature. For example, the Binomial Option stock option pricing result of Cox, Ross and Rubinstein (1979) will hold when prices or returns are interpreted as mid-points between the bid and ask. Other examples are the family of arbitrage pricing interest rate models of the Heath, Jarrow and Morton (1992) class, or their discrete counterparts - see Carr and Jarrow (1995), Jarrow (1995), Jarrow (1996). (Note: we restrict our discussion to European derivatives because the optimal exercise strategy will depend upon the agent holding the security.)

The paper begins in Section 1 by outlining the basic model. Section 2 uses Kuhn-Tucker theory to characterize the consumer and intermediary problems. Section 3 considers derivative assets and deduces one-sided arbitrage pricing results. Section 4 provides a brief discussion of the general equilibrium implications of trading restrictions and the difficulty in obtaining robust results. Section 5 discusses the modifications necessitated by non-convex technologies. Section 6 discusses market innovation and its connection with the recent literature that explores welfare and pricing implications of opening markets. Section 7 links the discussion in Section 6 with the earlier discussion in Section 4, by making some cautionary observations on the robustness of comparative static results. Section 8 concludes.

1 Basic Model with Transaction Costs

Our economy is a straightforward extension of the standard, discrete time/event economy familiar from the literature on general equilibrium theory. (See Milne (1995) and Magill and Quinzii (1996) for detailed text book discussions of the general equilibrium model without transaction costs. Naik (1995), Carr and Jarrow (1995) Jarrow (1995), Pliska (1997) provide detailed accounts of discrete time/event models of derivative pricing without transaction costs, and some early work on transaction costs.) Consider an economy with a finite time horizon $T = \{0, 1, \dots, T\}$ 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. All agents have rational expectations in the sense of Radner (1972): asset price expectations at any node are not falsified by the evolution of the economy.

1.1 Consumers

First, consider the characteristics of a set of consumers $I = \{1, \dots, I\}$. Assume consumer $i \in I$ has a consumption set $X_i \subset \mathfrak{R}^{\Omega+1}$, and preferences described by a utility function U_i .

A.1 $X_i \subset \mathfrak{R}^{\Omega+1}$ is closed, convex, bounded below, and $0 \in X_i$.

A.2 $U_i : X_i \rightarrow \mathfrak{R}$ is (a) continuous; (b) quasi-concave; (c) strictly increasing.

A.3 $U_i : X_i \rightarrow \mathfrak{R}$ is differentiable on the $\text{int}(X_i)$.

The differentiability assumptions are convenient for using calculus. We will assume in our discussion below that optimal consumption vectors will be in the $\text{int}(X_i)$ thus allowing us to concentrate on financial asset trading restrictions and asset prices. Otherwise the restrictions on the consumer's consumption set and preferences are mild and allow a range of familiar special cases. In particular, the consumption set can be considered to be net consumption, $x_i - w_i$, (we could ignore a separate endowment, $w_i \in \mathfrak{R}_+^{\Omega+1}$, for the consumer), and preferences can be defined over net consumption. Preference restrictions are relatively mild, and allow von Neumann-Morgenstern preferences as well as recursive and other non-standard preferences under uncertainty. These are all special cases of the assumed consumer utility function. (For examples of non-standard utility functions used in asset pricing models see Dow and

Werlang (1992), Epstein and Zin (1989), Kelsey and Milne (1996)).

1.2 Security Returns and Prices

Consumers trade in financial asset or security markets that transfer purchasing power across contingencies. We will restrict our discussion to securities with fixed payoffs (eg. Bonds, Equities, European options). Since trading in 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 pay-off matrix R and two asset price matrices. First, assume that there is a finite but very large set of assets $K = \{1, \dots, K\}$, 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 pay-offs (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.

An example illustrates the flexibility of our framework. (For more examples in the model without transaction costs, see Milne (1995), Naik (1995), and Magill and Quinzii (1996) Chapter 4.) Consider the return and price matrices as set out in Table 1. The underlying tree is described in Figure 1. There are three dates, and binomial uncertainty at each non-terminal node.

Table 1 and Figure 1 about here

At node ω_t , we denote $P_k^B(\omega_t)(P_k^S(\omega_t))$ as the buying (selling) price of a unit of asset k traded at ω_t . Asset k is a unit claim on a contingent return stream $R_k(\omega_s), \forall \omega_s \in S(s | \omega_t), \forall s > t$. (We denote $S(s | \omega_t)$ as the successor nodes to ω_t at time $s > t$.) For example, asset $k = 1$ has a buying (selling) price $P_1^B(\omega_0)(P_1^S(\omega_0))$

at ω_0 , which entitles the holder to one unit of the commodity at ω_{11} and ω_{12} . Notice that by holding one unit of $k = 2$ and 3 , an agent can replicate the payoff of asset $k = 1$. Later we will deduce implications on the relative magnitudes of the buying and selling prices of assets $k = 1, 2, 3$.

Observe that we can incorporate dynamic portfolio strategies of trading in long-lived assets by specifying the payoffs of an asset. The central idea is to expand the set of asset descriptors to incorporate not only their trading date/event, but also if they are bought or sold. Consider the example of assets $k = 4$ and 5 . Asset 4 has payoffs at $t = 1, 2$, and asset 5 has the same payoff at $t = 2$. We can think of asset 5 as the same asset as asset 4, but traded ex-dividend/coupon at ω_{11} .

Therefore, an agent could buy one unit of asset 4 at ω_0 , at the buying price, and liquidate the position at ω_{11} by selling one unit of asset 5. Notice that the agent will have received the intermediate return of 2 at ω_{11} , but have a zero net position on the return at ω_{21} and ω_{22} . Furthermore, the agent will incur transaction costs on the trade by selling asset 5 at the selling price, and closing out the position.

1.3 Intermediaries

The second group of agents are intermediaries who purchase, transform and resell securities using a costly technology. For intermediary $h \in \{1, \dots, H\} = H$ assume there is a technology $T_h \subset \mathfrak{R}_+^K \times \mathfrak{R}_+^K \times \mathfrak{R}^{S+1}$, with typical element $(\Delta_h^B, \Delta_h^S, y_h)$; where Δ_h^B is the number of "bought" securities supplied by the intermediary; Δ_h^S is the number of "sold" assets purchased by the intermediary; 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 consumer.

The technology can incorporate 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 intermediary 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. Our model can also incorporate constraints that are specific to an intermediary and do not appear directly in the bid-ask spread. A special case of this model is when the technology T_h incorporates dynamic constraints on trade; for example, the technology can constrain short sales. We will discuss this type of constraint in more detail below in section 2. In summary, intermediary h 's problem is:

$$\max V_h(x_h) \tag{1.1}$$

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

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

We assume that the intermediary technology satisfies standard micro-economic conditions:

A.4 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 section 5 below, explaining the modifications that can be made to our results to allow for this type of transaction technology. We assume that the intermediary is owned by a single entrepreneur to avoid the complications introduced by the failure of the Fisher Separation Theorem.

A.5 (i) X_h is closed, convex, bounded below and $0 \in X_h$;

(ii) $V_h(\cdot)$ is differentiable, strictly increasing and quasi-concave.

We could assume that the intermediary has a stochastic endowment, but for simplicity we will subsume it into the definition of the consumption set.

1.4 The Formal Consumer Problem

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

$$\max_{x_i \in X_i} U_i(x_i) \tag{1.2}$$

s.t.

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

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

where $\Delta_i^B \in \mathfrak{R}_+^K, \Delta_i^S \in \mathfrak{R}_+^K$ are the consumer's purchases (sales) of asset k ; \bar{x}_i is an exogeneous endowment; and T_i is the consumer's own transaction technology.

A.5

T_i has the properties:

- (a) $0 \in T_i$
- (b) T_i is closed
- (c) T_i is convex.

The consumer and intermediary problems are identical in the abstract, apart from a change in sign in the budget constraints, to represent the intermediary taking the other side of the market from consumers. 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.

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

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

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

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 problems (1.1) - (1.2), and market clearance equations (1.3). \square

Jin and Milne (1999) provide sufficient conditions for the existence of an equilibrium in a generalized version of this economy, with multiple commodities. The paper discusses the case with convex technologies for the intermediaries; and also explores existence results with non-convex technologies.

2 Characterizations of Agent Optimality Conditions

2.1 The Consumer's Problem: First Order Conditions and Personalized Martingale Pricing

In this section we characterize the consumer's optima using the Kuhn-Tucker first order conditions for a maximum. Given A.1-A.3 then the first order conditions are necessary and sufficient for a maximum - for example, see Mangasarian (1969), or Takayama (1985). In our discussions using Kuhn-Tucker conditions we will assume that the constraint set has an interior point, i.e., the constraint qualification is satisfied. 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 $F_{i\ell}(\Delta_i^B(\omega_t), \Delta_i^S(\omega_t), y_i(\omega_t); \omega_t) \geq 0$, for all ω_t and $\ell = 1, \dots, L$. 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 $(x_i^* \in \text{int}X_i, y_{i\ell}^*(\omega_t) > 0)$ are:

(2.A.1)

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

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

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

(2.A.2)

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

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

(2.A.3)

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

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 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 $\gamma_i(s | \omega_t) = \sum_{\omega_s \in S(s|\omega_t)} (\lambda_i(\omega_s)/\lambda_i(\omega_t))$, where $\lambda_i(\omega_t) = \frac{\partial U_i}{\partial x_i(\omega_t)}$; and $S(s | \omega_t)$ is the subset of F_s which can be reached by starting from ω_t .

Finally, define: $\tilde{p}_i(\omega_s | \omega_t) = [\lambda_i(\omega_s)/\lambda_i(\omega_t)][\gamma_i(s | \omega_t)]^{-1}$. Given that $\nabla U_i \gg 0$ by Assumption (A.2), then by construction $(\tilde{p}_i(\omega_s | \omega_t))$ has the same properties as a conditional probability measure with full support.

Recalling the conditions (2.A.1) and (2.A.2), we can rewrite them as:

(2.A.4)

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

$$\delta_{ki}^B(\omega_t)\Delta_{ki}^B(\omega_t) = 0; \quad \delta_{ki}^B(\omega_t) \geq 0; \quad \Delta_{ki}^B(\omega_t) \geq 0.$$

(2.A.5)

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

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

The Consumer's Problem Without Personal Transaction Costs: Assume, for the moment that we can ignore personal transaction costs. To analyze (2.A.4) and (2.A.5) 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:

(2.A.6)

$$\frac{\delta_{ki}^B(\omega_t)}{\lambda_i(\omega_t)} - P_k^B(\omega_t) + \gamma_i(t+1 | \omega_t) \sum_{S(t+1|\omega_t)} R_k(\omega_{t+1}) \tilde{p}_i(\omega_{t+1} | \omega_t) = 0.$$

(2.A.7)

$$\frac{\delta_{ki}^S(\omega_t)}{\lambda_i(\omega_t)} + P_k^S(\omega_t) - \gamma_i(t+1 | \omega_t) \sum_{S(t+1|\omega_t)} R_k(t+1) \tilde{p}_i(\omega_{t+1} | \omega_t) = 0$$

Immediately, we obtain the result:

(2.A.8)

$$P_k^B(\omega_t) - P_k^S(\omega_t) = \frac{\delta_{ki}^B(\omega_t) + \delta_{ki}^S(\omega_t)}{\lambda_i(\omega_t)} \geq 0.$$

By the monotonicity of utility $\lambda_i > 0$; if $P_k^B(\omega_t) - P_k^S(\omega_t) > 0$ (i.e., there is a bid-ask spread), then $\delta_{ki}^B + \delta_{ki}^S > 0$.

This implies that either the consumer has:

- (a) $\Delta_{ki}^\delta = \Delta_{ki}^B = 0$ – does not trade asset k ;
- (b) $\Delta_{ki}^\delta > 0$; $\Delta_{ki}^B = 0$ – sells, but does not buy asset k ;
- (c) $\Delta_{ki}^\delta = 0$; $\Delta_{ki}^B > 0$ – does not sell, but buys asset k .

Notice that the consumer will not simultaneously buy and sell asset k ; transactions costs make it inefficient to contemporaneously go long and short in an asset.

Next, consider the general case where there are long-lived assets that have returns $\{R_k(\omega_s) \geq 0\}$, $\omega_s \in S(s | \omega_t)$, $s > t$. Recalling the general conditions (2.A.6), (2.A.7), we obtain:

(2.A.9)

$$P_k^B(\omega_t) - \frac{\delta_{ki}^B(\omega_t)}{\lambda_i(\omega_t)} = \sum_{s>t} \gamma_i(s | \omega_t) \sum_{S(s|\omega_t)} R_k(\omega_s) \tilde{p}_i(\omega_s | \omega_t) = P_k^S(\omega_t) + \frac{\delta_{ki}^S(\omega_t)}{\lambda_i(\omega_t)};$$

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

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 (cf. Dermody and Prisman (1988), Dermody and Rockafellar (1991), (1995), Jouini and Kallal (1995), Ortu (1996)).

Indeed, it is an easy extension of (2.A.9) to obtain for a long-lived asset:

(2.A.10)

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

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 $P_k^i(\omega_t), P_k^i(\omega_{t+1})$ will lie strictly between their buying and selling prices. But if the consumer is buying (selling) at ω_t , then $P_k^B(\omega_t) = P_k^i(\omega_t) (P_k^S(\omega_t) = P_k^i(\omega_t))$.

2.2 A Linear Programming version of the Model

So far we have characterized the consumer's optimum in terms of shadow prices at the consumer optimum. By exploiting these dual (Martingale) prices, we can simplify the consumer's optimizing problem to a Linear Program (for example, see Mangasarian (1969) or Takayama (1985)). That is: at the 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:

(2.A.11)

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

s.t.

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

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

In solving this problem notice that by the choice of $\{\gamma_i, \tilde{p}_i(\cdot)\}$, 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 $\{\gamma_i, \tilde{p}_i(\cdot)\}$. 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 characterisation. 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 (2.A.11) 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 rationalisation of our linear framework would be to assume that the consumer or 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.

2.3 Further General Characterisations

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 problem (2.A.11): 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 $\rho(\bar{x}_i) = \{x_i \in \mathfrak{R}^{L+1} \mid U_i(x_i) \geq U_i(\bar{x}_i)\}$. By assumption A.2, it is nonempty, closed and convex. Furthermore, $\bar{x}_i \in \beta \cap \rho(\bar{x}_i)$. With sufficiently high transaction costs (i.e., a

large $B - S$ spread) the only solution to (2.A.11) 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 $S \equiv (1 - \tau^S)P$. If we consider all pairs $(\tau^B, \tau^S) \in \mathfrak{R}_+^{\Omega+1} \times \mathfrak{R}_+^{\Omega+1}$ 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 price process and the stochastic endowment stream.

2.4 Implications of Preference and Return Restrictions

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 characterisations of the portfolio strategies (cf. Neave (1970), Constantinides (1979), Davis and Norman(1990) for an early sample of papers; and Karatzas and Shreve (1998) Ch.6 for a recent survey and references.) Most of these characterisations 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 (ie. HARA) and a limited menu of assets to introduce added simplicity to the characterisation. 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.

To illustrate this type of analysis, consider the consumer problem (1.2) without a personal technology. Consider the consumer's optimal portfolio strategy. If the consumer's utility function is homothetic, then it follows easily from the consumer optimum conditions that a scalar multiple of the contingent endowment vector implies that the optimal asset portfolio strategy will be the same scalar multiple of the original portfolio strategy. This result is quite general and applies to any stochastic price process and buy-sell structure of prices. If in addition, we assume that the rate of return vector for the buy-sell prices at each contingency is identical, there

are no intermediate contingent endowments, the homothetic utility function is von Neumann-Morgenstern with a constant discount factor (this is equivalent to being a member of the HARA class of utility functions), and the conditional probabilities over rates of returns are identical at each non-terminal contingency, then it is easy to show that the optimal portfolio strategy in each contingency will be a scalar transform of a normalised optimal strategy. The proof follows easily from observing that under the stated conditions, the first order inequality conditions (2.A.10) will be invariant to scale. Thus we can define a proportional optimal asset portfolio strategy that is invariant to scale. Notice that in the case of no transaction costs this result reduces to the old Samuelson (1969) result. But as we have argued, this intuitive result generalizes with our transaction cost conditions to include not only trading regions for asset trades, but to the no-trade region as well.

With a finite horizon, the optimal proportional strategy will not be stationary; but if we allow an infinite horizon, then the optimal proportional strategy will be stationary for the obvious reason that the problem is identical at each event/time except for the inherited portfolio proportions. Although this result is attractive and intuitive, it is easy to destroy it by minor modifications of the stringent conditions. For example, variations in the investment opportunity set by allowing contingent volatility changes in the price processes, or contingent changes in the buying and selling prices through variations in the spread, will alter the contingent portfolio strategies in non-trivial ways. (For simulations of simple examples with this type of model see Lynch and Balduzzi (2000).)

2.5 The Consumer's Problem With Personal Transaction Costs:

Let us reintroduce the personal transaction cost functions, $F_{i\ell}(\cdot)$. For simplicity we will assume that market transaction costs are zero, i.e., $P^B = P^S$. The first order conditions (2.A.4) and (2.A.5) become:

(2.A.12)

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

(2.A.13)

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

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

(2.A.14)

$$\sum_{\ell} \frac{\delta_{iR}(\omega_t)}{\lambda_i(\omega_t)} \frac{\partial F_{i\ell}}{\partial \Delta_{ik}^B(\omega_t)} - P_k(\omega_t) \equiv -P_k^{Bi}(\omega_t).$$

(2.A.15)

$$\sum_{\ell} \frac{\delta_{i\ell}(\omega_t)}{\lambda_i(\omega_t)} \frac{\partial F_{i\ell}}{\partial \Delta_{ik}^S(\omega_t)} + P_k(\omega_t) \equiv -P_k^{Si}(\omega_t).$$

Given these definitions, then we can analyse the consumer'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, $P_k^B(\omega_t) > P_k^S(\omega_t)$, 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.

2.6 An Example: Restricted Access:

In a recent paper, Basak and Cuoco (1998) have explored a two consumer exchange economy with just two assets: a riskless bond, and a risky asset. One consumer is restricted from trading in the risky asset, but the other consumer trades both securities. It follows immediately from market clearance that the unrestricted consumer 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 consumer, (2.A.12-13) collapse to the standard martingale condition. But for the restricted consumer the bond equation is the unrestricted condition, and the risky asset has personalised transaction costs set sufficiently high that no participation is warranted.

2.7 The General Technology: The Special Case of Short Selling Constraints:

We 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:

(2.A.16)

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

(2.A.17)

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

or regrouping them we obtain the interpretation:

(2.A.18)

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

(2.A.19)

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

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:

$$(2.A.20) \quad P_K(\omega_t) = \sum_{s>t} \gamma_i(s | \omega_t) \sum_{S(s|\omega_t)} [R_K(\omega_s) \tilde{p}_i(\omega_s | \omega_t) + \frac{\delta_{Ki}(\omega_s)}{\lambda_i(\omega_t)}] = 0.$$

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 (2.A.20) in a more recognizable form, by defining

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

so that we obtain:

(2.A.21)

$$P_K(\omega_t) = \sum_{s>t} \gamma_i(s | \omega_t) \sum_{S(s|\omega_t)} [R_K(\omega_s) + \tilde{\delta}_{Ki}(\omega_s)] \tilde{p}_i(\omega_s | \omega_t) = 0.$$

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 $\tilde{\delta}_{K_i}(\omega_s)$ is the personalised 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. For further discussion of this type of constraint in a general equilibrium model see Detemple and Murthy (1997). They solve an explicit two consumer example to illustrate possible solutions to general equilibrium prices and allocations. Our result shows that the interpretation of personalised future yields on the short sale constrained security is quite general.

2.8 The General Technology: VAR and other Constraints

The argument on short selling constraints can be generalised to more general constraints on asset trades. For example there could be complicated dynamic constraints on asset trades that arise from legal or risk management considerations. Because there are many different types of constraints, we will not do the formal modelling here, but we 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 Variance at Risk (VAR) constraints which are a sequence of prespecified Bayesian updated constraints on asset returns and trading strategies.

It is clear that by using the linear programming approach discussed above, that one can construct large dimensional LP problems that consider portfolio strategies for any class or classes of securities with legal or risk management constraints. Any realistic program will be of considerable dimension and require serious sensitivity analysis in specifying asset return processes and the constraint sets. In addition one can add transaction costs in the sense of (2.A.11) 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.

2.9 The General Technology: Induced Preferences

As our last example of the consumer with a general technology, consider the induced preferences for a reduced form of the consumer problem. The basic idea here is an old one in economic theory (see Milne (1981) for an analysis and references to the earlier literature). In our example, consider any vector of asset trades and define the maximised utility over the consumer's constraint set and given net trades. It is

easy to show that with our assumptions above, one can derive a continuous, quasi-concave utility function over the closed, convex set of feasible net security trades (for a proof see Milne (1981)). It is not difficult to show that if the primitive utility function is differentiable (assumption A.3), then the induced utility function may not be differentiable if the functions defining the constraint set are not differentiable or there are corner solutions implicit in the solution (see Machina (2000) for details). Furthermore, if the primitive utility function is additively separable over contingent consumption (for example, the standard intertemporally additively separable, expected utility function), then the induced utility may not be additively separable. Examples of this type analysis are given in Kelsey and Milne (1999). This research is suggestive that so-called "non-standard preferences" may mimic standard preferences with constraint sets on trading and consumption.

2.10 The Intermediary's Problem: Characterization

In this section we will characterize the general intermediary's problem and explore special cases of the model. Recall the intermediary problem 1.1:

(2.B.1)

$$\max V_h(x_h)$$

(i)

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

(ii)

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

Given the assumptions on the utility function and the technology T_h (A.4, A.5), then (2.B.1) is a quasi-concave programming problem. The technology is sufficiently general to cover a number of interesting cases. As we observed above, the consumer and intermediary's problems are identical. But we will concentrate here on the intermediary interpretation and this will depend upon the formulation and interpretation of the technology for transactions.

2.11 Case 1: The Pure Broker

First consider the simple case of a pure broker where the technology is restricted, the agent h holds no inventory and trades directly with the market i.e. $\Delta_h^B = \Delta_h^S = \Delta$. Furthermore, to keep the argument as simple as possible, assume that there is an elementary transaction technology described as a sequence of contingent production functions $F_h(\Delta_h^B(\omega_t), \Delta_h^S(\omega_t), y_h(\omega_t); \omega_t) \geq 0$, where the contemporaneous commodity is used to facilitate trades through the market.

Setting up the Lagrangian, we find:

(2.B.2)

$$\begin{aligned} L_h = & V_h(R[\Delta_h^S - \Delta_h^B] - S\Delta_h^S + B\Delta_h^B + y_h) \\ & + \sum_{\omega_t} \delta_h(\omega_t) F_h(\Delta_h^B(\omega_t), \Delta_h^S(\omega_t), y_h(\omega_t); \omega_t) + \delta_h^B \Delta_h^B + \delta_h^S \Delta_h^S \\ & + \sum_{\omega_t} \theta_h(\omega_t) (\Delta_h^B(\omega_t) - \Delta_h^S(\omega_t)) \end{aligned}$$

The first order necessary and sufficient conditions for a maximum imply:

(2.B.3)

$$\frac{\partial V_h}{\partial x_h(\omega_t)} [P_k^B(\omega_t) - P_k^S(\omega_t)] + \delta_{kh}(\omega_t) = -\delta_h(\omega_t) \frac{\partial F_h(\omega_t)}{\partial \Delta_{kh}(\omega_t)}$$

where $\Delta_h^B = \Delta_h^S = \Delta_h$; $(\frac{\partial F_h}{\partial \Delta_{k^B}^B} + \frac{\partial F_h}{\partial \Delta_{k^S}^S}) \equiv \frac{\partial F}{\partial \Delta_{k^h}}$; and $\delta_{kh}(\omega_t) \equiv \delta_{kh}^B(\omega_t) + \delta_{kh}^S(\omega_t)$. It follows that if the broker trades, $\Delta_{kh}(\omega_t) > 0$, then $\delta_{kh}(\omega_t) = 0$, and the price spread is proportional to $\frac{\partial F}{\partial \Delta_{kh}}$.

The price spread will depend upon the general forces of demand and supply, except in the special case where all intermediaries have identical production technologies and there is constant returns to scale in production. To analyse that case, assume that at ω_t when asset k is traded it absorbs a constant proportion of the contemporaneous commodity:

$$\Delta_{kh}(\omega_t) - g_k(\omega_t) \cdot y_{kh}(\omega_t) = 0 \text{ for all } h = 1, \dots, H$$

and $g_k(\omega_t) > 0$. Substituting into (2.B.3), we obtain:

(2.B.4)

$$(P_k^B(\omega_t) - P_k^S(\omega_t)) - \frac{\delta_{kh}(\omega_t)}{\lambda_h(\omega_t)g_h(\omega_t)} = g_k(\omega_t)^{-1}.$$

If the asset is not traded $\Delta_{kh}^* = 0$ for all h then $P_k^B(\omega_t) - P_k^S(\omega_t) \leq g_k(\omega_t)^{-1}$. A good example of this phenomenon would be a large contingent increase in transaction costs at ω_t (compared to the previous event) which eliminates trade in asset k at ω_t . Alternatively, if the asset is traded, then we have

(2.B.5)

$$P_k^B(\omega_t) - P_k^S(\omega_t) = (g_k(\omega_t))^{-1},$$

and the asset pricing spread is determined by the transaction technology. By specifying the matrix $G = [g_k(\omega_t)]$ of transaction coefficients and assuming that trades take place at all events ω_t , (i.e. all markets are active) one can determine the stochastic process of the spread from the transaction technology, i.e., $B - S = G$. If some asset markets are inactive we simply get the weaker statement $B - S \leq G$.

The model above assumed competitive markets, but we can make the model more realistic by assuming that the brokers are specialists who trade the security by posting bid and ask prices. Given this Bertrand type pricing behaviour it follows from elementary microeconomics that the brokers will act competitively in the sense of equating price to marginal cost, so long as there are two or more brokers acting non-collusively in any security market.

2.12 Case 2: The Dealer

To analyse the case of a dealer who trades on their own account, we return to the more general formulation (2.B.1). For simplicity assume that the transaction technology for the dealer is described by the functional inequality $F_h(\cdot)$ as in (2.B.2), but the dealer is not constrained to a pure pass-through function for asset traders. The modified Lagrangean becomes:

(2.B.6)

$$L_h = V_h(R[\Delta_h^S - \Delta_h^B] - S\Delta_h^S + B\Delta_h^B + Y_h)$$

$$\begin{aligned}
& + \sum_{\omega_t} \delta_h(\omega_t) F_h(\Delta_h^B(\omega_t), \Delta_h^S(\omega_t), Y_h(\omega_t); \omega_t) \\
& + \delta_h^B \Delta_h^B + \delta_h^S \Delta_h^S.
\end{aligned}$$

which is identical to the Section A consumer's problem apart from the sign change on buying and selling assets. We can modify (2.A.4), (2.A.5) to obtain.

(2.B.7)

$$\sum_{\ell} \delta_{h\ell}(\omega_t) \frac{\partial F_{h\ell}}{\partial \Delta_{hk}^B(\omega_t)} + \frac{\delta_{kh}^B(\omega_t)}{\lambda_h(\omega_t)} + P_h^B(\omega_t) - \sum_{s>t} \gamma)_{h(s|\omega_t)} \sum_{S(s|\omega_t)} R_k(\omega_s) \tilde{p}_h(\omega_s | \omega_t) = 0$$

$$\delta_{kh}^B(\omega_t) \Delta_{kh}^B(\omega_t) = 0; \quad \delta_{kh}^B(\omega_t) \geq 0; \quad \Delta_{kh}^B(\omega_t) \geq 0.$$

(2.B.8)

$$\sum_{\ell} \delta_{h\ell}(\omega_t) \frac{\partial F_{h\ell}}{\partial \Delta_{hk}^S(\omega_t)} + \frac{\delta_{kh}^S(\omega_t)}{\lambda_h(\omega_t)} + P_k^S(\omega_t) + \sum_{s>t} \gamma)_{h(s|\omega_t)} \sum_{S(s|\omega_t)} R_k(\omega_s) \tilde{p}_h(\omega_s | \omega_t) = 0$$

$$\delta_{kh}^S(\omega_t) \Delta_{kh}^S(\omega_t) = 0; \quad \delta_{kh}^S(\omega_t) \geq 0; \quad \Delta_{kh}^S(\omega_t) \geq 0.$$

Generalising our discussion of personalized buying and selling prices in section A, define

$$P_k^{Bh} \equiv \sum_{\ell} \delta_{h\ell}(\omega_t) \frac{\partial F_{h\ell}}{\partial \Delta_{hk}^B(\omega_t)} + P_k^B(\omega_t),$$

$$P_k^{Sh} \equiv - \sum_{\ell} \delta_{h\ell}(\omega_t) \frac{\partial F_{h\ell}}{\partial \Delta_{hk}^S(\omega_t)} + P_k^S(\omega_t).$$

Clearly when dealers buy, they should take into account not only the buying price, but the additional cost incurred in their inhouse transactions. Similarly, when they sell, they should deduct the inhouse costs from the sale proceeds.

Furthermore, because the dealer does not have to pass the asset trade through like a broker, they can take one side of the market and absorb this trade from their existing inventory. Any general characterization of trades will be a complicated amalgam of preferences, asset returns and the dealer's transaction costs. There is a simple Martingale representation that summarizes these features:

(2.B.9)

$$P_k^{Bh}(\omega_t) = \sum_{s>t} \gamma_h(s | \omega_t) \sum_{S(s|\omega_t)} R_k(\omega_t) \tilde{p}_h(\omega_s | \omega_t) - \frac{\delta_{kh}^B(\omega_t)}{\lambda_h(\omega_t)}$$

(2.B.10)

$$P_k^{Sh}(\omega_t) = \sum_{s>t} \gamma(s | \omega_t) \sum_{S(s|\omega_t)} R_k(\omega_t) \tilde{p}_h(\omega_s | \omega_t) + \frac{\delta_{kh}^S(\omega_t)}{\lambda_h(\omega_t)}$$

To make further progress in characterizing asset prices, we would require more detailed assumptions on utility, the transaction technology and the structure of asset returns. The following is a simple example of the type of analysis we have in mind.

2.13 Example: Inter Dealer Trading:

It is well known that in many markets trading is confined almost exclusively to specialist dealers, mutual funds, banks etc. One can justify this by choosing a pattern of transaction costs that exclude ordinary consumers, but allow dealers to trade securities between themselves at sufficiently low transaction costs (Merton (1991) uses this idea in his discussion). For example consider the extreme simplification where dealers face zero transaction costs, but consumers face infinite transaction costs except for a small subset of assets which have zero or small transaction costs. The latter could be mutual fund products or riskless deposits at banks. From our first order conditions it is easy to adapt the standard martingale pricing results on securities where the trades are undertaken by dealers, but consumers are excluded. This would approximate the pricing behavior we see in the market. Basak and Cuoco (1998) is an extreme version of this type of model, where their unrestricted consumer plays the role of the dealer, and their restricted consumer is limited to buying a riskless mutual fund.

2.14 Intermediary Portfolio Constraints:

We can introduce additional constraints on trading securities in exactly the same way as we did with the consumer. In the absence of trading costs we can adapt those arguments to the case of an intermediary. For example, risk management security trading constraints (e.g.VAR) can be incorporated in exactly the same way as we indicated in the consumer's problem above.

2.15 Intermediaries with Constant returns to Scale Technology

It is not difficult to adapt the model to the situation where the dealer has a constant marginal cost technology and can hold an inventory. It follows that competitive dealers will price at marginal cost just as our earlier broker example, but they will be able to hold inventory as well. The marginal cost pricing conditions will be identical to the broker model.

3 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 general 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 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 rationalised theoretically.

3.1 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) 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 F with nonnegative holdings $\alpha_k(\omega_t)$, such that at ω_t , $R_{k'}(\omega_s) = \sum_k \alpha_k(\omega_t) R_k(\omega_s) \equiv R_F(\omega_s)$ for all $\omega_s \in S(s | \omega_t)$ and $s > t$. This is a simple case where there is a set of assets with future returns that are linearly dependent. Recalling (2.A.10), we deduce immediately that

(3.1.1)

$$P_{k'}^i(\omega_t) = \sum_k \alpha_k(\omega_t) P_k^i(\omega_t) \equiv P_F^i(\omega_t), \text{ for every } i \in I;$$

and so,

(3.1.2)

$$P_{k'}^B(\omega_t) \geq P_{k'}^i(\omega_t) \geq P_{k'}^S(\omega_t).$$

Because $P_k^B(\omega_t) \geq P_k^i(\omega_t) \geq P_k^S(\omega_t)$, then if $\sum_k P_k^B(\omega_t) \alpha_k(\omega_t) < P_{k'}^B(\omega_t)$, we deduce $P_{k'}^B(\omega_t) > P_{k'}^i(\omega_t)$ and it follows from (2.A.9) that $\delta_{k'i}^B(\omega_t) > 0$ and $\Delta_{k'i}^B = 0$. That is, no consumer will buy asset k' at ω_t .

A symmetric argument applies to the selling price. If $\sum_k P_k^S(\omega_t) \alpha_k(\omega_t) > P_{k'}^S(\omega_t)$, 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

(3.1.3)

$$\sum_k P_k^B(\omega_t) \alpha_k(\omega_t) \geq P_{k'}^B(\omega_t) \geq P_{k'}^S(\omega_t) \geq \sum_k P_k^S(\omega_t) \alpha_k(\omega_t),$$

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

Conversely, we can prove the following proposition that will be important in our discussion in Section 3.2.

Proposition 3.1.1

For asset k' to have an active market (or for asset k' to be effective in Ortu's (1996) terminology), then:

(3.1.4)

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

(3.1.5)

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

Proof: Assume the converse for (3.1.4), then $P_{k'}^B > \sum_{k \in F} P_k^B(\omega_t) \alpha_k(\omega_t)$. For any i , given (3.1.1), $\sum_{k \in F} P_k^B(\omega_t) \alpha_k(\omega_t) \geq \sum_{k \in F} \alpha_k(\omega_t) P_k^i(\omega_t) = P_{k'}^i(\omega_t)$. But this implies $P_{k'}^B(\omega_t) > P_{k'}^i(\omega_t)$; which in turn implies $\delta_{k'i}^B(\omega_t) > 0$; or $\Delta_{ki}^B = 0$ for any i . Thus no consumer buys k' and the market is inactive, a contradiction.

A symmetric argument applies for (3.1.5). ||

Note: The proposition shows that an active market for k' requires (3.1.4) and (3.1.5) as necessary conditions, but they are not sufficient for activity.

Corollary 3.1

If $A_{k'} \equiv \{(\alpha_k(\omega_t)) \in \mathfrak{R}_t^K \mid R_F(\omega_t) = R_{k'}(\omega_t)\}$ is non-trivial, then (3.1.4) (3.1.5) apply for all $(\alpha_k(\omega_t)) \in A_{k'}$.

Now let us turn to the broker's problem. If the market for k' is active (i.e., there is at least one buyer and one seller for k') then we know that (3.1.3) must hold. But given the assumption on intermediary technology we know that from (2.B.4) and Proposition 3.1.1 that:

$$g_{k'}(\omega_t)^{-1} \leq \sum_k \alpha_k(\omega_t) [1 + \frac{\delta_k(\omega_t)}{\lambda_h}] g_k(\omega_t)^{-1}.$$

(3.1.6)

In words, the intermediary technology provides a cheaper means of trading asset k' than the replicating portfolio. In the case where all the replicating assets are traded, (3.1.6) reduces to

$$g_{k'}(\omega_t)^{-1} \leq \sum_k \alpha_k(\omega_t) g_k(\omega_t)^{-1},$$

(3.1.7)

so that activity and the relative bid-ask spread is determined solely by the technology.

Consider a long-lived European derivative asset and its replicating portfolio during the life of the asset. There is no reason that (3.1.3-7) should be true for all $\omega_s \in S(s | \omega_t), s > t$. That is, if we consider the stochastic process representing the δ 's and g ()'s we would have periods where an agent is locked into the derivative and finds it unprofitable to trade in this illiquid market because of its large spreads compared to the lower spread in the replicating portfolio. This situation is represented by:

(3.1.8)

$$P_{k'}^B(\omega_t) \geq \sum_k P_k^B(\omega_t) \alpha_k(\omega_t) \geq \sum_k P_k^S(\omega_t) \alpha_k(\omega_t) \geq P_{k'}^S(\omega_t).$$

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 (3.1.3) and (3.1.8) reveals that the consumer 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, 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 behaviour of traders when faced with different spreads.

3.2 Equi-Proportional Transaction Costs, Effective Substitutes and Arbitrage Pricing

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. Recalling our results in section 2.B, we conclude that with equi-proportional transaction costs $B \leq (1 + \theta)S$ 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 proportion 2 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. (Nonnegativity is not a serious constraint if we allow negative payoffs as mirror images of positively held securities.) 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 (3.2.1) 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:

(3.2.1)

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

(3.2.2)

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

where AD is the set of Arrow-Debreu securities for all events subsequent to T_t .

Proof: By the assumption of equi-proportional transaction costs

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

From Proposition: (3.1.1)

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

and

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

But this implies:

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

which proves (3.2.1).

A similar proof implies (3.2.2). \parallel

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:

Proposition (3.2.2)

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

(3.2.3)

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

(3.2.4)

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

Proof: Formally identical to Proposition 3.2.1, replacing the set AD with F . \parallel

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 Proposition 3.2.2 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 the standard arbitrage pricing results for European derivatives 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. Clearly this requires a sufficiently rich set of securities to achieve our result, a significantly richer set than the relatively sparse set required without transaction costs.

3.3 Example 1: The Binomial Option Pricing Model:

Consider the familiar binomial tree where the state space is defined by the arithmetic prices of the stock. There is a riskless bond, and all possible European call and put options are available (assuming that exercise prices are on a finite grid, there will be a finite but very large number of securities.) All assets have the same proportional transaction costs. Beginning at the end of the tree, we use Proposition 3.2.2 and the midpoint pricing result to deduce the one period binomial pricing proposition on mid-prices. Retreating to the previous node we can use exactly the same argument to imply that all active derivative prices will satisfy the midpoint pricing argument. Repeated recursion implies the standard Binomial Option pricing result where prices are arithmetic averages of the buy-sell spread.

Notice that the same logic implies the Put-Call parity result on mid-prices.

3.4 Example 2: Rates of Return and Interest Rate Models:

Our discussion above used arithmetic averaging, but we could have used geometric averages and converted the argument to rates of return. Now the spread will be between the appropriate buying, selling and mid, rates of return, or interest rates in the case of riskless bonds. By using the same backward induction argument as above, we can translate the discrete models of rates of returns in stock options (for example Cox, Ross and Rubinstein (1979), or the models of interest rate derivative pricing (see Jarrow (1996) for a survey) to obtain standard option pricing results. Indeed we can go further and adapt any discrete model which uses spanning and the martingale technology to price European options. For example, we can apply our result to models with stochastic volatility or regime shifts.

These examples demonstrate that the frictionless asset pricing models can be rationalised as a good (indeed exact!) approximation in a more realistic world with transaction costs, where the prices used for modelling are the mid-prices or mid-

returns. Thus, the frictionless asset pricing models may be a better approximation than many researchers believe in providing unbiased estimates of asset prices.

3.5 Redundant Securities and Activity:

In the previous section we demonstrated a new result: that arbitrage pricing on mid-prices could replicate frictionless pricing results. It is important to explore the limits of this argument to see if it is robust to variations in our assumptions. We assumed that transaction costs were equally proportional on all assets. But for practical purposes we could have transaction costs being much higher on some assets than others. Thus our pricing result will still hold if we concentrate on a subset of spanning active securities. The inactive securities are not traded and are irrelevant for economic agents. Observe that in the standard frictionless model spanned assets are redundant and play no effective role in improving the welfare of agents. Our modification allows an extension of that idea to an economy with transaction costs, where there is a subset of redundant non-traded assets which are too expensive to trade because they have large transaction costs. Nevertheless, assume there are sufficient assets with equi-proportional transaction costs that provide a spanning set of traded assets. This subset will suffice for the conclusion of our Proposition (3.2.2).

In the standard frictionless argument, any agent can replicate a European derivative by trading dynamically in a few underlying securities. (For example in the Binomial model we require only the stock and bond.) But in our model we merely require activity of the markets, or even more weakly, that the market is on the margin of activity. Our argument requires the role of both buyers and sellers. Given equi-proportional transaction costs, then mis-pricing on one side of the market will trigger a trade on the other side of the market (by market clearance) introducing activity and the elimination of the original mis-pricing. In other words, standard frictionless arbitrage arguments do not exploit market clearance as we do in an essential fashion.

4 Constraints on Arbitrage Trades and Implications for Pricing

In this section we will outline some results on trading and pricing when some, or all, agents have constraints on trading securities. We will assume no transaction costs, but introduce constraints on agent trades, exploring cases where arbitrage

pricing continues to hold. First, consider an economy where there exists at least one competitive agent who has no constraints on security trades. It follows easily from our results in section 3, that standard arbitrage pricing results hold. Notice that one unconstrained competitive agent is sufficient for this argument to eliminate arbitrage opportunities. We can weaken this argument to require merely one or more unconstrained agents at every event where arbitrage possibilities could arise. Thus all agents could be partially constrained, but there is sufficient slack in the constraints in equilibrium that there are always one or more agents who can eliminate arbitrage pricing possibilities. This argument has the weakness that we cannot prespecify which agents will be crucial to eliminating arbitrage pricing differentials. Instead we require that in equilibrium there is at least one agent at each event who is unconstrained in eliminating any potential arbitrage.

In the more general situation, where constraints can be binding in essential ways, potential arbitrage opportunities can be frustrated by trading restrictions. This is obvious. But the implications of this observation are subtle: even in simple models with restricted numbers of agents and securities the analysis becomes complex to characterise. (See Detemple and Murthy (1997)).

5 Non-Convex Transaction Technologies

Our discussion so far has assumed that the transaction technology is convex. But in practice there are fixed costs to executing many transactions. This observation is consistent with non-convex transaction technology sets for consumers and intermediaries. As discussed in Jin and Milne (1999), non-convexities can frustrate the existence of an equilibrium for our model. Nevertheless, by assuming many agents and restrictions on the degree of non-convexity, it is possible to prove in the limit that an equilibrium exists. This result allows us to adapt most of our analysis to the non-convex transaction technology case with minor modifications. Assuming continuously differentiable utilities, the consumer and intermediary first order conditions will be necessary, but not sufficient for a maximum. In addition, the demand/supply correspondence for the agents need not be convex valued, as the non-convexity can imply the existence of "gaps" in the optimal responses.

If the non-convexities are "large" in the economy then the approximation argument fails. This is the familiar territory of monopolistic competition models, or strategic interaction where scale economies lead to a delicate trade-off between scale economies and product/asset differentiation. Such a model requires a different for-

mulation to the competitive construction in this paper (see Allen and Gale (1994)).

In our model, we have assumed that consumers and intermediaries are identical in the abstract formulation. But it is obvious that consumers with small wealth will be faced with a disadvantageous technology because of small scale trades, whereas large intermediaries will have overcome the fixed costs of setting up trading desks, information networks, etc. for large trading volumes. Therefore, it is obvious that trading will be much more frequent with the large intermediaries than with the small consumer. The existence of initial set up costs, can explain the size and scope of large scale intermediaries. Given our dynamic structure it is easy to model the evolution of intermediaries where a new security market opens, new high cost firms enter, and as time passes, the technology evolves so that costs fall and competition eliminates all but a few large survivors. Strictly speaking we would require increasing costs to eventually dominate the initial economies of scale to bound the firm sizes throughout this evolutionary process.

A final observation: our pricing results in section 3 remain unchanged as long as we satisfy the assumption of activity, and that constant marginal costs operate after overcoming some initial fixed costs.

6 Opening New Asset Markets

In the literature we can distinguish between different interpretations of "opening" an asset market. We relate these interpretations to our model and illustrate the strengths and weaknesses of each concept. In our model of market equilibrium all contingent asset markets exist, but contingent asset market activity is defined by positive trade in an asset at an event. All agents observe conditional buying and selling prices at each event, and thus do not have to estimate prices in inactive markets. It is easy to construct examples where a particular asset (e.g., an exotic European option) is not traded for the first few periods, but because of demand and supply conditions and/or transaction costs, is traded after a threshold event. For example, an extreme case is where the transaction costs are wholly prohibitive until some time t' , and thereafter are zero for the security. Clearly trading will only take place for events after t' , and the plans of agents prior to t' will take into account the eventual opening of the security market. In a situation of this type, markets "opening" are rationally anticipated, and the optimal decisions and their implications for security pricing continue to apply. As far as we know this interpretation of rationally anticipated security market opening is new to the finance literature, although it is implicit in the original economic theory

literature.

Security market opening need not be anticipated by agents. In these circumstances Hart (1975) showed via an example that this kind of security market "opening" (a surprise to agents) can lead to any welfare result. Subsequently Milne and Shefrin (1987), Cornes and Milne (1989), and more generally Elul (1995) have found conditions under which one can obtain any desired utility change across agents by appropriate opening of security markets. These results are closely related to Second-Best General Equilibrium comparative static exercises. For example in the economics literature in International Trade theory we can illustrate such a result. Starting from an arbitrary tariff-distorted equilibrium, there are paths of gradual tariff reform that progress via strict Pareto improvements in welfare (cf. Turunen-Red and Woodland (1991)). The connections between the international trade literature and the asset market literature are made explicit in Cornes and Milne (1989).

Also related to these results, is the observation (see Hart (1975)) that there can be multiple equilibria for an asset economy with incomplete markets, and that unlike the standard frictionless case, the equilibria can be Pareto ranked. In our discussion of asset economies we can include incomplete markets as a limiting case where some markets are traded with zero transaction costs and the remainder have infinite transaction costs. Thus our characterisation is of a particular equilibrium and ignores the possibility of multiple rational expectation equilibria. An exception to this general possibility is an economy with a representative agent (see Luttmer (1996)). In that economy there will be no trade with or without transaction costs, and the allocation is a trivial Pareto Optimum. But there can be non-unique supporting prices if the transaction cost technology has a kink at zero trade.

Another branch of the literature abandons the General Competitive Equilibrium framework and accepts a less ambitious Partial Equilibrium analysis (cf. Madan and Soubra (1991)). The model allows choice over the asset payoffs to be introduced, trading off the demand responses of buyers with the cost of innovation. This approach, which is at variance with our price-taking assumption, is related to General Equilibrium Monopolistic Competition models as formulated, for example, by Allen and Gale (1994).

A third branch of the literature attempts to price new securities that are about to be traded. In our framework one can consider a security not traded at an event, and deduce its shadow evaluations by different agents. Using our personalized Martingale construction, each agent has a shadow evaluation for the securities return stream. By introducing restrictions on the consumer's optimal portfolio and preferences it is possible to obtain bounds on the shadow valuation of any security (cf. Constantinides

(1994); Constantinides and Zariphopoulou (1995), including references to an extensive literature on this methodology). However, there is no reason to believe that shadow valuations prior to trade, will remain unchanged after trade: the introduction of the new asset can affect all shadow prices and asset prices. (This observation is merely the dual to the primal models of Hart (1975) and his successors discussed above.)

For example, consider a two-date, binomial event economy with a riskless bond traded on a frictionless market, and a stock traded with positive transaction costs. Consider the introduction of a European call option on the stock, traded at zero transaction cost. By the shadow price method one can estimate the value of the option, prior to trade. But with the introduction of the option, the new equilibrium has the bond and option trading, but not the stock. Furthermore the new contingent allocation is Pareto Optimal as the bond and option span the binomial state-space. In this example the shadow price method provides a poor estimate of the option price after the option is introduced.

7 Comparative Static Analysis, Security Prices and Robustness: A Cautionary Note:

There are growing number of papers that solve particular versions of General Equilibrium models with restricted numbers of securities, simple transaction costs and/or constraints on security trades and two agents. The aim of this literature is to provide results on the impact of these frictions on equilibrium security allocations and prices. These models are highly stylised with very restrictive assumptions on preferences, return processes, the number of agents, the transaction technology, constraints etc. and aim to demonstrate intuitive relationships or generate counter-examples to conjectured results. As a source of counter examples, these models play a very useful function demonstrating that general equilibrium models can generate complex feedback effects that can overturn first order impacts. These type of models should be handled with caution. In this paper we have tried to provide general characterisations that do not rely on particular specifications of preferences, endowments etc. Particular cases can be fitted into our framework and we have given numerous examples that can be accommodated in this way. As we observed in the previous section, one should be very careful in extending the analysis to consider comparative static exercises that attempt to compare equilibria before and after the opening of a security market. Without strong restrictions, perverse welfare results can be obtained from opening or closing new security markets. As a dual result to that observation, Boyle

and Wang (2000) show that generically any variation in security trading possibilities will have an impact on all prices. This result will apply to opening or closing markets, or merely tightening or loosening binding constraints on security trades. This type of argument should make one cautious in making strong predictions on relative price changes, resulting from the impact of regulatory or other changes that alter trading opportunities, when the results are derived from comparative static analyses of highly restrictive models.

8 Conclusion

We have constructed a general equilibrium of an asset economy with transaction costs and trading constraints that is sufficiently flexible to allow for a wide variety of situations discussed in the literature. Using basic calculus tools we have provided a series of characterizations of the equilibria illustrating a number of interesting properties of prices and asset allocations. Given the generality of the structure, it should be clear that we have in no way exhausted the results that can be obtained from the model - there are deeper implications that will be the topic of further research.

9 Appendix

Table 1

$\omega_t \setminus k$	R					$R_t \setminus k$	$b = B, S$				
	1	2	3	4	5		1	2	3	4	5
ω_0/k	*	*	*	*	*		$P_1^b(\omega_0)$	$P_2^b(\omega_0)$	$P_3^b(\omega_0)$	$P_4^b(\omega_0)$	*
ω_{11}/k	1	1	*	2	*		*	*	*	*	$P_5^b(\omega_{11})$
ω_{12}/k	1	*	1	*	*		*	*	*	*	*
ω_{21}/k	*	*	*	3	3		*	*	*	*	*
ω_{22}/k	*	*	*	4	4		*	*	*	*	*
ω_{23}/k	*	*	*	*	*		*	*	*	*	*
ω_{24}/k	*	*	*	*	*		*	*	*	*	*

The asterisk simply means that the vectors are mapped into subsets of the contingent event space. An alternative formulation would be to have them all mapped into the full event space and replace the asterisk by a zero.

BIBLIOGRAPHY

- Allen, F. and D. Gale (1994). *Financial Innovation and Risk Sharing*. MIT Press.
- Basak, S. and Cuoco, D. (1998). "An equilibrium Model with Restricted stock Market Participation", *The Review of Financial Studies*, 11, vol 2, 309-402.
- Boyle, P. and Wang, T. (2001). "Valuation of New Securities in an Incomplete Market: The Catch 22 of derivative Pricing", *Mathematical Finance*, Vol.11/3, July, 267-284.
- Carr, P. and Jarrow, R. (1995). "A Discrete Time Synthesis of Derivative Security Valuation Using a Term Structure of future Prices", Chapter 7 in R. Jarrow et al, (eds.) *Finance, Handbook on OR and MS*, Vol 9, Elsevier Science.
- Constantinides, G. (1979). "Multiperiod Consumption And Investment Behavior with Convex Transaction Costs," *Management Science*, 25, 1127-1137.
- Constantinides, G. (1986). "Capital Market Equilibrium with Transaction Costs," *Journal of Political Economy*, Vol. 94, No. 4, 842-862.
- Constantinides, G. (1994). "Option Bounds Based on Stochastic Dominance," Graduate School of Business, University of Chicago.
- Constantinides, G. and Zariphopoulou, T. (1995). "Universal Bounds on Option Prices with Proportional Transaction Costs," mimeo, University of Chicago.
- Cornes, R. and Milne, F. (1989). "A Simple Analysis of Mutually Disadvantageous Trade," *Economic Studies Quarterly*, 40, 122-134.
- Cox, J., Ross, S. and Rubinstein, M. (1979). "Option Pricing: A Simplified Approach", *Journal of Financial Economics*, 7, 229-63.
- Davis, M.H. and Norman, A.R. (1990). "Portfolio Selection with Transaction Costs," *Mathematics of Operations Research*, 15, 676-713.
- Debreu, G. (1959). *Theory of Value*, Yale University Press.
- Dermody, J. and Prisman, E. (1988). "Term Structure Multiplicity and Clientele in Markets with Transaction Costs and Taxes," *Journal of Finance*, 43, 893-911.
- Detemple, J. and Murthy, S. (1997). "Equilibrium Asset Prices and No-Arbitrage

- with Portfolio Constraints," *Review of Financial Studies*, 10, No.4 ,
1133-1174.
- Dermody, J. and Rockefellar, R. (1991). "Cash Stream Evaluation in the Face of Transaction Costs and Taxes," *Mathematical Finance*, 1, 31-54.
- Dermody, J. and Rockefellar, R. (1995). "Tax Basis and Nonlinearity in Cash Stream Evaluation," *Mathematical Finance*, 5, 97-119.
- Dow, J. and Werlang, S. (1992). "Uncertainty Aversion and the Optimal Choice of Portfolio," *Econometrica*, 60, 197-204.
- Duffie, D. (1990). "Money in General Equilibrium Theory," chapter 3 in B. Friedman and F. Hahn (eds.), *Handbook of Monetary Economics*, Vol. 1, North-Holland.
- Edirisinghe, C., Naik, V. and Uppal R. (1993). "Optimal Replication of Options with Transaction Costs," *Journal of Quantative and Financial Analysis* 28, 117-138.
- Ellickson, B. (1993). *Competitive Equilibrium*, Cambridge University Press.
- Elul, R. (1995). "Welfare Effects of Financial Innovation in Incomplete Market Economies with Several Consumption Goods," *Journal of Economic Theory*, 65, 43-78.
- Epstein, L. and Zin, S. (1989). "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 837-909.
- Epstein, L. and Wang, T. (1994). "Intertemporal Asset Pricing Under Knightian Uncertainty," *Econometrica*, Vol. 62, No. 2, 283-322.
- Hart, O. (1975). "On the Optimality of Equilibrium When the Market Structure is Incomplete," *Journal of Economic Theory*, 11, 418-443.
- Heath, D., Jarrow, R. and Morton, A. (1992). "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims," *Econometrica*, 60, 77-106.
- Heller, W. and Starr, R. (1976). "Equilibrium with Non-Convex Transaction Costs: Monetary and Non-Monetary Economics," *Review of Economic Studies*, 43, 195-215.
- Jarrow, R. (1995). *Pricing Interest Rate Options*, Ch.8 in R. Jarrow et al, (eds.) *Finance, Handbook on OR and MS*, Vol 9, Elsevier Science.

Jarrow, R. (1996). *Modelling Fixed Income Securities and Interest Rate Options*, McGraw-Hill.

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.

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

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

Kelsey, D. and Milne, F. (1999). "Induced Preferences. NonAdditive Beliefs and Multiple Priors", *International Economic Review*, 40, 2, pp.455-477.

Luttmer, E. (1996). "Asset Pricing in Economies with Frictions," *Econometrica*, 64, 6.

Lynch, A. and Balduzzi, P. (2000). "Predictability and Transaction Costs: the Impact on Rebalancing Rules and Behavior," *Journal of Finance*, 55, 5, pp.2285-2310.

Machina, M. (2000). "Payoff Kinks in Preferences over Lotteries" mimeo, UNSD.

Madan, D. and Soubra, B. (1991). "Design and Marketing of Financial Products," *Review of Financial Studies*, 4, 361-384.

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

Mangasarian, O. (1969). *Non-Linear Programming*, McGraw-Hill.

Merton, R. (1991). "The Financial System and Economic Performance," *Journal of Financial Services Research*, 4, 263-301.

Milne, F. (1981). "Induced Preference and the Theory of the Consumer," *Journal of Economic Theory*, 24, 205-213.

Milne, F. (1995). *Finance Theory and Asset Pricing*, Oxford University Press.

Milne, F. and Shefrin, H. (1987). "Information and Securities: Pareto Dominance and the Second Best," *Journal of Economic Theory*, 43, 418-443.

Milne, F. and Smith C. (1980). "Capital Asset Pricing with Proportional Transaction Costs," *Journal of Financial and Quantitative Analysis*, Vol. XV, No. 2, 253-265.

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.
- Neave, E. (1970). "The Stochastic cash Balance Problem with Fixed Costs For Increases and Decreases," *Management Science*, 16, 472-490.
- O'Hara, M. (1995). *Market Microstructure Theory*. Blackwell.
- Ortu, F. (1996). "Arbitrage and Martingales in Finite Dimensional Securities Markets with Transaction Costs: A Linear Programming Approach," mimeo, Università di Trieste.
- Ostroy, J. and Starr, R. (1990). "The Transactions Role of Money," chapter 1 in B. Friedman and F. Hahn (eds.), *Handbook of Monetary Economies*, Vol. 1, North-Holland.
- Pliska, S. (1997). *Introduction to Mathematical Finance: Discrete Time Models*, Blackwell.
- Radner, R. (1972). "Existence of Equilibrium of Plans, Prices and Price Expectations in a Sequence of Markets," *Econometrica*, 40, 289-303.
- Samuelson, P. (1969). "Lifetime Portfolio Selection by Stochastic Programming," *Review of Economics and Statistics*. Reprinted in R.
- Merton (ed.) *The Collected Scientific Papers of Paul A. Samuelson*, Vol. 3, MIT Press 1972.
- Takayama, A. (1985). *Mathematical Economics*, Cambridge University Press.
- Turunen-Red, A. and Woodland, A. (1991). "Strict Pareto Improving Multilateral Reforms of Tariffs," *Econometrica*, 59, 1127-1152.

