Chapter 31

ARBITRAGE AND MARKET FRICTIONS

SHASHIDHAR MURTHY, Rutgers University, USA

Abstract

Arbitrage is central to finance. The classical implications of the absence of arbitrage are derived in economies with no market frictions. A recent literature addresses the implications of no-arbitrage in settings with various market frictions. Examples of the latter include restrictions on short sales, different types of impediments to borrowing, and transactions costs. Much of this literature employs assumptions of continuous time and a continuous state space. This selected review of the literature on arbitrage and market frictions adopts a framework with discrete states. It illustrates and discusses a sample of the principal results previously obtained in continuous frameworks, clarifying the underlying intuition and enabling their accessibility to a wider audience.

Keywords: arbitrage; frictions; asset pricing; review; short sales constraints; sublinear pricing functional; super martingales; discrete state space; transactions costs; borrowing constraints

31.1. Introduction

The concept of arbitrage and the requirement that there be no arbitrage opportunities is central to finance. Essentially, an arbitrage opportunity is an investment where one can get something for nothing: a trading strategy with zero or negative current cost that is likely to yield a positive return and sure to not entail a future liability. Thus, the requirement that there be no arbitrage is a minimal desired attribute of a properly functioning securities market.

The implications of the absence of arbitrage are central to much of finance, simultaneously illuminating many areas and giving rise to new fields of inquiry. From early developments of the spot-forward parity relationships to the fundamental irrelevance propositions of Modigliani and Miller (1958), many arguments have at least implicitly used the main intuition of no-arbitrage that close substitutes must obey the law of one price, viz. two securities with the same payoffs must have the same price. Modern day application of this intuition came to the fore with the Black and Scholes (1973) model of option pricing. A first systematic analysis of the implications of no arbitrage was then carried out by Ross (1976, 1978). The principal question in such analysis is: given a set of some primitive assets, how much can one infer about the valuation of other assets if there are to be no arbitrage opportunities? Both the analysis of Ross (1976, 1978) and its generalization by Harrison and Kreps (1979) assume that investors are able to trade in frictionless markets.

A recent, burgeoning literature addresses the implications of no-arbitrage in settings with various market frictions. Examples of the latter include restrictions on short sales, different types of impediments to borrowing, and transactions costs. This paper reviews a selected portion of this literature and surveys the principal results obtained. Much of this literature employs the assumption of continuous time or an infinite dimensional
state space. Here, a discrete framework is adopted in the interest of clarifying the intuition behind previously obtained results and rendering them accessible to a wider audience.

The principal implication of no-arbitrage in a frictionless setting may be summarized by what is sometimes known as the Fundamental Theorem of Asset Pricing (Dybvig and Ross, 1987). This theorem states that the absence of arbitrage is equivalent to the existence of both a strictly positive linear pricing rule and a solution to the choice problem of some investor who prefers more to less. Apart from implying that the law of one price holds, this result has several alternative representations and implications. One of the best known is that the no-arbitrage value of a claim is the cost of a portfolio that exactly replicates or hedges the claim’s payoff. A second is that relative prices of assets must be martingales under a “risk-neutral” probability measure.

Rather than purport to be an exhaustive survey, this paper reviews a sample of the main results from the literature on arbitrage and market frictions. One striking result is that the cheapest way to hedge a given liability may be to hedge a larger liability. This was first shown by Bensaid et al. (1992) in a transactions costs setting. An implication of this is that pricing may fail to be linear and instead be sublinear: the value of the sum of payoffs may be less than the sum of the values of the individual payoffs. Thus, there may be room for financial innovation, or departures from Modigliani–Miller (1958) type irrelevance, where an intermediary pools securities, and then strips them; see Chen (1995) for a discussion. When there are no frictions, the price paid when buying a claim is also the amount received in going short or writing the claim. Market frictions which result in sublinearity of the valuation or pricing rule can lead to bid–ask spreads on derivative securities even when there are no transactions costs (i.e. bid–ask spreads) in trading the primary securities, as shown by Luttmer (1996).

Furthermore, departures from the law of one price and the martingale property may occur under frictions. In the presence of a short sales constraint that changes elastically depending on the collateral posted, Hindy (1995) showed that an asset’s value depends not only on its dividends but also on the collateral services it provides. When investors face short sales or borrowing constraints, Jouini and Kallal (1995a,b) show that asset prices may be super martingales.

The rest of this paper is organized as follows. A basic framework is set out in the next section, following which the benchmark case of no frictions is discussed in Section 31.3. Due to limitations of space, we formally illustrate the above results considering primarily the case of no short sales in Sections 31.4 and 31.5. However, we also briefly outline the impact of other types of frictions such as constraints on portfolio weights that permit some short sales (such as that under a leverage constraint or margin restriction), and transactions costs in Section 31.6. We conclude with some remarks relating to the consistency with equilibrium of results obtained from the no-arbitrage approach under frictions.

### 31.2. A Basic Framework

Consider an economy over dates $t = 0$ and $T$. Uncertainty is described by a discrete state space $\Omega$ with typical member $\omega \in \{1, \ldots, N\}$ denoting the final state of nature realized at date $T$ where $N < \infty$. The probabilities of these states are $\{p(\omega)\}$ corresponding to an underlying probability measure $P$.

Investors trade a set of primitive assets which are in positive net supply, and whose prices are taken as given. Asset $j = 1, \ldots, J$ has price $S_j(0)$ at date 0 and the future price $S_j(\omega) \equiv D_j(\omega)$ in state $\omega$ at date $T$, where $D_j$ is a given random dividend or payoff. Asset $j = 1$ is taken to be a risk-free bond with current price of unity; (one plus) its constant interest rate is denoted $R$. A portfolio choice is $z = (z_1, \ldots, z_J)$, comprising holdings of shares of the various assets at date 0. Investors choose portfolios to maximize their preferences that are strictly increasing in consumption at dates 0 and $T$. 
Trading in assets is subject to market frictions that take the form of a constraint on short sales and/or borrowing. The formulation we will consider for most of this paper restricts holdings of shares of some or all assets to be at least as large as exogenously given lower bounds: \( z_j \geq -\xi_j \), where \( \xi_j \geq 0 \). In the case of no short sales of asset \( j \), \( \xi_j = 0 \); if instead some limited but fixed amount of short sales is permitted, \( \xi_j > 0 \). Similarly, note that the no borrowing case corresponds to \( \xi_j = 0 \), since asset \( j = 1 \) is the risk-free bond. A portfolio that satisfies the short sales constraint is termed admissible.

Investors can use the primitive assets to create, i.e., exactly replicate, various payoffs using admissible portfolios. Every such payoff \( x \equiv \{x(\omega)\} \) where \( x(\omega) = \sum_j z_j S_j(\omega) \) is hence said to be marketed, i.e., available for purchase and/or sale. In the presence of market frictions, the set of marketed payoffs is not limited to those payoffs that can be explicitly replicated. For instance, consider a payoff \( x \) of 1 in some state \( \omega' \) and 0 in other states whose replication requires a portfolio that involves a short position in asset \( j \) (and positions in other securities). Suppose, the latter short position is equal to the maximum amount permitted of \( \xi_j \). Then, the payoff \( 2x \) cannot be exactly replicated because it would require a short position of \( 2\xi_j \) shares. However, the payoff \( 2x \) may still be marketed if there exists a portfolio that produces at least 2 in state \( \omega' \) and 0 elsewhere; i.e., if the payoff can be super-replicated.

Thus, it is natural to define a price for an arbitrary payoff \( x \) as the minimum cost

\[
\phi(x) \equiv \left\{ \sum_j z_j S_j(0): x(\omega) \leq \sum_j z_j S_j(\omega), \forall \omega \right\}
\]  

(31.1)

at which it can be exactly replicated or super-replicated by an admissible portfolio, where the associated functional \( \phi(.) \) is termed a pricing or valuation rule.

An arbitrage opportunity is an admissible portfolio \( z \) that either has (i) a nonpositive cost \( \sum_j z_j S_j(0) \) when initiated and a date \( T \) payoff \( x \equiv \{x(\omega)\} \) where \( x(\omega) = \sum_j z_j S_j(\omega) \), which is positive in some states and nonnegative in others, or (ii) a negative current cost and a nonnegative future payoff in all states.

31.3. Exact Replication and Prices under no Frictions

At this stage, it is useful to present the principal result on the implications of the absence of arbitrage for the benchmark case where there are no market frictions. This result, known as the Fundamental Theorem of Asset Pricing, is due to Ross (1976, 1978). Given the definition of the pricing or valuation operator \( \phi(.) \), it is clear that there are no arbitrage opportunities if and only if the pricing rule in Equation (31.1), denoted \( \phi^*(.) \) here, is positive and linear.

Apart from implying that the law of one price must hold, the linearity property means that \( \phi^*(\lambda x) = \lambda \phi^*(x) \) for all \( \lambda \), i.e., the price functional is homogeneous. It is useful to further interpret the above result in terms of an implicit state price vector \( \xi \equiv \{\xi(\omega)\} \) where \( \xi(\omega) \) is the price of a state security that pays 1 unit in state \( \omega \), and 0 elsewhere. The linearity and positivity of \( \phi^*(.) \) are equivalent to \( \phi^*(x) = \sum_\omega \xi(\omega) x(\omega) \) and \( \xi(\omega) > 0 \), respectively. The pricing rule \( \phi^*(.) \) values every marketed payoff precisely because the latter can be exactly replicated, or hedged, using a portfolio of existing assets: it assigns a value equal to the cost of the replicating portfolio.

Another useful interpretation of the linearity of \( \phi^*(.) \) is that there exists a (“risk-neutral”) probability measure \( Q^* \) that is equivalent to the underlying measure \( P \) under which relative or normalized asset prices are martingales. Thus,
every primitive asset’s current price relative to, say, the price of the bond (which is 1), is equal to the expectation under $Q'$ of its future payoffs relative to that of the bond: $S_j(0) = E^{Q'} [D_j R^{-1}]$. Equivalently, the value of every payoff satisfies $\phi(x) = \sum q'(\omega)x(\omega)R^{-1}$, where $q'(\omega)$ denotes the risk-neutral probability of state $\omega$ under $Q'$. These well-known implications of no-arbitrage in frictionless markets provide the basis of most option pricing models, following Black and Scholes (1973), Merton (1973), and Cox and Ross (1976).

31.4. No Short Sales

We now return to the economy with frictions of Section 31.2, and consider the case of no short sales. As in the frictionless case, it is clear that there are no arbitrage opportunities in this setting only if every nonnegative marketed payoff $x$ (which is positive in some state) has price $\phi(x) > 0$. We proceed by recording a result below that is the counterpart to Proposition 1.

**Proposition 2.** Suppose the only friction is that the short sales of some assets is prohibited, i.e. $\underline{x}_j = 0$ for some $j$, and $\overline{x}_j = \infty$ for the rest. Then there are no arbitrage opportunities if and only if the pricing rule in Equation (31.1), denoted $\phi^{NS}(\cdot)$ here, is positive and sublinear. Furthermore, there exist underlying positive hypothetical linear pricing rules $\phi^{\ast}(\cdot)$ such that $\phi^{NS}(x) \geq \phi^{\ast}(x)$, for all marketed payoffs $x$. Also, there exists a new probability measure associated with $\phi^{NS}(\cdot)$ under which the (normalized) price process of an asset is a supermartingale if the asset cannot be sold short, and a martingale if the asset can be sold short.

The proof follows from Garman and Ohlson (1981), Chen (1995), Jouini and Kallal (1995a,b), and Luttmer (1996), and rather than reproduce it here, we will shortly present a simple binomial example where the result is explicitly illustrated. (Also note that while some of these papers consider transactions costs, their results apply here). But first, a few implications of the sublinearity property and the supermartingale property are discussed.

Observe that, in contrast to Proposition 1, the pricing rule $\phi^{NS}(\cdot)$ is not linear but sublinear. The sublinearity implies that the value of a portfolio of two payoffs $x$ and $y$ may be less than the sum of the values of the payoffs, i.e. $\phi^{NS}(x + y) \leq \phi^{NS}(x) + \phi^{NS}(y)$.

It also implies that $\phi^{NS}(\lambda x) = \lambda \phi^{NS}(x)$ for all $\lambda \geq 0$, i.e. the price functional is positively homogeneous.

Chen (1995) discusses the role of financial innovation in such a context. He shows that an innovator (who is assumed to not face any short sales constraint, unlike other investors) can earn profits by purchasing a “pooled” payoff $x + y$ at a cost $\phi^{NS}(x + y)$, stripping it into individual components $x$ and $y$, and selling (i.e. issuing) the latter at prices $\phi^{NS}(x)$ and $\phi^{NS}(y)$, respectively. Other investors cannot earn the same profits because they cannot short-sell (i.e. issue) the individual component securities $x$ and $y$. In a frictionless economy, in contrast, the linearity of the pricing rule $\phi^{\ast}(\cdot)$ leaves no role for such financial innovation; i.e. the Modigliani–Miller (1958) invariance proposition holds.

Next, consider the relationship between the value of a security with payoff $x$ and another security with payoff $-x$. In a frictionless world, the values of these two securities (the second security is essentially equivalent to going short the first) are the negative of each other, i.e. their values sum to 0. This follows from the linearity (homogeneity) of the valuation rule $\phi^{\ast}(\cdot)$. Under no short sales, the valuation rule $\phi^{NS}(\cdot)$ is only positively homogeneous, and thus $\phi^{NS}(-x)$ may differ from $-\phi^{NS}(x)$. The intuition is just that the cost of super-replicating a payoff $x$ will in general differ from that for the payoff $-x$. Also note that since the value of a zero payoff must be zero, $\phi^{NS}(x) + \phi^{NS}(-x)$ is at least as large as $\phi^{NS}(x + (-x)) = \phi^{NS}(0) = 0$; i.e. the sum of the values of both securities may be positive. Consequently, the ask price $\phi^{NS}(x)$ of the payoff $x$ may exceed the bid price $-\phi^{NS}(-x)$. Thus, as Jouini and Kallal (1995a,b) and Luttmer (1996) show, a derivative security’s price may exhibit a bid–ask spread even where there are no transactions costs (i.e. bid–ask spreads) in trading the primitive assets.
As we noted in Section 31.3, asset prices (normalized by, say, the bond) in frictionless economies are martingales under the risk-neutral probability measure. In other words, one cannot expect to earn more than the risk-free rate after correcting for risk. In sharp contrast, Proposition 2 shows that there exists a risk-neutral probability measure, say \( Q^{NS} \), under which (normalized) prices of assets subject to short sales constraints are super martingales. In other words, \( S(t)/R^{-1} = E^{Q^{NS}}[D_t] \) for such assets: their prices after correcting for risk and the risk-free return are expected to be nonincreasing. This is compatible with the absence of arbitrage opportunities from the perspective of a risk-neutral investor because an asset whose price is expected to decrease relative to the bond cannot be sold short. This super martingale property was proved by Jouini and Kallal (1995a,b) in a model with short sales constraints (and transactions costs).

### 31.5. A Simple Binomial Model

As an example of a simple model that explicitly illustrates the results of Proposition 2 and their significance, we now consider a one-period binomial model. A stock and bond are traded with the constraint that no short sales of the stock is permitted, but borrowing (short sales of the bond) is allowed. The stock’s current price is \( S \) and its end-of-period price is \( uS \) in state \( u \), and \( dS \) in state \( d \). The bond has current price of unity and one plus a risk-less return of \( R \) in state \( d \) and \( u \).

Consider a payoff \( x \equiv (x_d, x_u) \) comprised of \( x_d \) in state \( d \) and \( x_u \) in state \( u \). Hedging any such payoff requires a portfolio of \( z_s \) shares of the stock and \( z_b \) units of the bond that satisfies

\[
  z_s\omega S + z_bR \geq x_0; \quad z_s \geq 0; \quad \omega \in \{d, u\}
\]

where \( \omega \in \{d, u\} \) denotes both the future state and the return of the stock. Note from Equation (31.2) we allow for the possible super-replication of the payoff; also observe that \( z_s \) must satisfy the no-short-sales constraint. Since the cost of the hedge portfolio is \( z_sS + z_b \) it follows, using Equations (31.1) and (31.2), that the value of the payoff is

\[
  \phi^{NS}(x) \equiv \text{Min} \{ z_sS + z_b; \; z_s\omega S + z_bR \geq x_0; \; z_s \geq 0; \; \omega \in \{d, u\} \}
\]

i.e. it equals the cost of the cheapest hedge portfolio.

Denote the risk-neutral probability of state \( u \) in the frictionless counterpart to the above example by \( q^* \equiv (R - d)/(u - d) \). It is then easy to verify that the solution to (31.3) is:

\[
  \phi^{NS}(x) = [q^*x_u + (1 - q^*)x_d]R^{-1} \quad \text{if} \quad x_u \geq x_d \quad \text{(31.4)}
\]

and

\[
  \phi^{NS}(x) = x_dR^{-1} \quad \text{if} \quad x_u < x_d. \quad \text{(31.5)}
\]

In other words, for a payoff such as that of a call option, where \( x_u > x_d \), the value is given by (31.4) and is no different from what it would be in a frictionless world. This is because exact replication, or an exact hedge, of the call entails a long position in the stock and borrowing. In contrast, for a security such as a put option, where \( x_u < x_d \), the value in Equation (31.5) is just the discounted value of the payoff in the “down” state discounted at the risk-free return. The reason is that an exact hedge or replication of the put would require short sales of the stock and is hence infeasible due to the no-short-sales constraint. Instead, the cheapest super-replication of the put involves a long bond position with face value \( x_d \).

To see that the valuation functional \( \phi^{NS} \) in Equations (31.4) and (31.5) is sublinear, compare the value of the payoff \( (dS, uS) \) from the stock with the sum of the values \( \phi^{NS}(dS, 0) \) and \( \phi^{NS}(0, uS) \). The former is obviously \( \phi^{NS}(dS, uS) = [q^*uS + (1 - q^*)dS]R^{-1} = S. \) However, the latter sum, \( \phi^{NS}(dS, 0) + \phi^{NS}(0, uS) = dSR^{-1} + [q^*uS + (1 - q^*)]0 \) \( R^{-1} = S + dSR^{-1}q^* \), exceeds the current stock price, and this proves the sub-linearity. The intuition is that the cost of hedging the combined payoff \((dS, uS)\) is less than the sum of the costs of hedging \((dS, 0)\) and \((0, uS)\) because hedging \((dS, 0)\) entails super-replication.
Finally, we show how the super martingale property of Proposition 2 comes about. Recall that with no frictions, \( q' = (R - d)/(u - d) \) is the risk-neutral probability of state \( u \) under which the stock, bond, and all other payoffs (i.e. options) are martingales. Now define the probability \( q \in [0, q'] \) and the associated hypothetical linear valuation rule \( \phi^q(x) = [qx_u + (1 - q)x_d]R^{-1} \). It is easy to verify that the actual sublinear valuation rule \( \phi^{NS}(x) \) of the economy with short sales constraints in Equations (31.4) and (31.5) is related to the sets \( \{ q \} \) and \( \{ \phi^q(\cdot) \} \) by:

\[
\phi^{NS}(x) = \max \{ \phi^q(x); q \in [0, q'] \} .
\]

Compared to the probability \( q' \), every other probability \( q \in [0, q'] \) places less weight on the “up” state and more weight on the “down” state. Hence, under each of these probabilities \( q \in [0, q'] \), the stock’s (normalized) current value exceeds its expected future value, i.e. \( S/R^{-1} > [quS + (1 - q)dS] \). In other words, the stock has a price process which is a super martingale because it cannot be sold short.

### 31.6. Other Types of Frictions

Due to limitations of space, we have so far considered primarily the case of no short sales. In this section, we briefly outline the impact of other types of frictions.

Consider an alternative formulation of a short sales constraint where the admissible extent of short sales of an individual asset varies with the value of the investor’s portfolio and with any collateral pledged. Such a constraint recognizes that some assets (such as a very liquid, short-term Treasury bill) are judged to have “high” value as collateral, and thus better afford the ability to maintain a short position than is the case with other assets (such as an illiquid, off-the-run Treasury bond) deemed to have “low” collateral value. In such a setting, Hindy (1995) proved that the absence of arbitrage implies that every asset’s price admits a decomposition into a dividend-based value and a residual that depends on the asset’s “collateralizability.” Thus, the law of one price may not hold: asset \( k \) may sell at a higher price than asset \( l \) even if their payoffs are the same if a one dollar worth of asset \( k \) allows investors the ability to short more of a third asset \( j \) than does a dollar worth of asset \( l \).

Transactions costs in trading some or all assets constitute yet another type of market friction. In a binomial stock price model with proportional transactions costs, Bensaid, et al. (1992), showed that even when an option’s payoff can be exactly replicated, it can be cheaper to hedge an option with a strategy that results in a payoff that dominates that of the option when there are transactions costs. This result is foreshadowed in Boyle and Vorst (1992) who derive the cost of exactly hedging an option in an identical framework, and show that their hedge portfolio’s cost is increasing in the number of trading periods for a high enough transaction cost parameter, and for options close to at-the-money—i.e. those which have a lot of convexity and whose exact replication requires a lot of rebalancing. Thus, the intuition from these papers is essentially that the benefits of exact replication can be traded off against savings on transactions costs. It should also be intuitively clear that in such settings that the cost of super-replicating a pool of payoffs may be cheaper than the sum of the costs of super-replicating the individual payoffs. In other words, the sublinearity result of Proposition 2 will continue to hold.

### 31.7. Conclusion

We have provided a review of the principal results which obtain when there are no arbitrage opportunities in a world where investors have to contend with market frictions. We conclude with some remarks about the consistency of these results with equilibrium.

One of the advantages of the no-arbitrage approach to valuation is that it allows one to make predictions about prices that are independent of particular investor attributes such as risk aversion, endowments etc. The reason is that the prices of
the existing primitive assets effectively subsume the risk preferences of the marginal investor. Furthermore, in the absence of frictions, all investors’ marginal utility-based valuations of all traded assets coincide: i.e. any investor may be taken to be the marginal agent supporting prices.

When there are frictions, investors’ valuations may be heterogeneous, and hence differ from that predicted by the no-arbitrage approach. For instance, when there are short sales constraints, Chen (1995) showed that the price of a security derived from the no-arbitrage condition may be lower than the price that the seller of the security can actually receive by selling it to the investor who values it most. Furthermore, as Detemple and Murthy (1997) showed, the introduction of what may otherwise be considered redundant securities can upset a given equilibrium in the presence of constraints on portfolio weights. More recently, Hara (2000) shows that even when introduction of a new security does not change utility-maximizing consumption choices it may give rise to a multiplicity of each investor’s security demands which in turn raises subtle equilibrium issues.

Thus, while routine application of the no-arbitrage approach in the presence of market frictions is not necessarily as useful as in a frictionless world, it nevertheless presents exciting new challenges for future research in asset pricing.

NOTES

1. Some other papers relevant to arbitrage and market frictions, which we do not discuss are Dybvig and Ross (1986), Jarrow and O’Hara (1989), and Prisman (1986).

2. Other important types of market frictions include (i) a constraint on portfolio weights (such as that under a leverage constraint or margin restriction) where the permitted amount of short sales or borrowing varies with the value of the portfolio, (ii) unlimited short sales at a cost that increases with the extent of short sales, and (iii) transactions costs that have either or both a fixed component and a variable component.

3. Given the availability of a risk-free bond, every payoff has such a minimum cost. Also note that each primitive asset must satisfy \( \phi(D_j) = S_j(0), j = 1, \ldots, J \), for if this were not true, they would not be held by any investor (which is incompatible with the fact that they are in positive net supply).

4. In this finite dimensional setting, the new probability measure associated with \( \phi^{NS}(\cdot) \) need not be equivalent to \( P \); i.e. the new measure need not assign positive probabilities to the same states that \( P \) does. However, limiting arguments can be used in an infinite state space to establish equivalency.

5. Note that such a violation of the law of one price does not occur in Sections 31.4 and 31.5 where we considered a simpler type of short sales constraint.

REFERENCES


