Cash Settlement, Price Manipulation, and the Modigliani-Miller Theorem

Albert S. Kyle

Robert H. Smith School of Business
University of Maryland

Very Preliminary First Draft

May 1, 2007

Please do not quote without permission

CASH SETTLEMENT, PRICE MANIPULATION, AND THE MODIGLIANI-MILLER THEOREM

Albert S. Kyle

May 1, 2007

[Preliminary Version: Please do not quote without the permission of the author.]

Some futures contracts call for physical delivery of the underlying asset. Examples are the bond contracts originally traded on the CBOT. Other futures contracts call for liquidation of outstanding positions at expiration by cash settlement. An example is the S&P 500 futures contract traded on the CME. When a cash-settled contract expires, outstanding positions are liquidated by marking them to cash market prices observed around the time of expiration. For example, the S&P 500 contract is liquidated by marking the position to the value of the S&P 500 index calculated at the open of trading.

CFTC rules require that markets underlying cash-settled futures contracts be "not subject to manipulation or distortion." The rules also require that cash settlement be "based on a cash price series that is reliable, acceptable, publicly available, and timely."

It has been argued by others that cash settlement makes futures markets less subject to manipulation, since it is not necessary for traders to trade in the underlying cash market. For the purposes of this paper, I will define manipulation as an intentional trading strategy undertaken for the purpose of influencing prices so that they do not reflect the "underlying," "natural," or "market" forces of supply and demand. Manipulative strategies include corners and squeezes, dissemination of false rumors designed to influence prices, and dissemination of fake transactions quantities and prices for the purpose of affecting prices. There may be other types of manipulative trading strategies, but that is an issue which goes beyond the scope of this paper.

Does having cash settlement rather than physical delivery make it easier or harder to manipulate a futures contract? Does a cash-settled futures contract create the possibility for manipulative schemes that are distinctly different from ones used with physical delivery? Does the goal of preventing manipulation suggest useful guidelines for constructing cash-settled futures contracts rather than contracts requiring physical delivery? How should cash-settled contracts be constructed?

The purpose of this paper is three-fold.

1

First, I argue that there is a simple and obvious sense in which it does not matter whether a futures contract has cash settlement or physical delivery. Under certain assumptions discussed below, the two different contract forms lead to precisely the same pricing and precisely the same economic outcomes. I call this "The Modigliani-Miller Theorem of Cash Settlement." The concept of arbitrage I have in mind is stronger than the static arbitrage relationship between a spot price and a futures price as it is usually formulated in textbooks. It is much stronger than the Black-Scholes notion of arbitrage, which requires traders to agree on zero-probability events. In addition to various tradability, solvency, and fungibility assumptions about the market, my argument rests crucially on the assumption that arbitragers have the ability to place market orders in the cash market with the property that the orders have guaranteed execution at precisely the cash market prices which will be used for cash settlement. I show that this crucial assumption makes futures contracts with physical delivery exactly equivalent to futures contracts with cash settlement.

Second, I argue that both the academic literature and practice of regulatory enforcement have become confused as a result of not recognizing the significance of this Modigliani-Miller Theorem. The literature confuses trades exerting downward pressure on prices with trades exerting upward pressure on prices. It imputes monopolistic motives to innocuous trading strategies in which traders supply liquidity to one another. The CFTC shares this confusion with the literature, as illustrated by their arguments in the Avista case.[1] I show that when traders agree on fundamental values, the logic of these "monopolistic" efforts to supply to liquidity to one another leads to a market price precisely equal to this fundamental value, with no trade necessary to achieve this since all traders agree what this fundamental value is.

Third, I argue that a useful perspective on futures contract design and futures market regulation is obtained when one starts the analysis from the Modigliani-Miller perspective. Here are some examples:

1.   If the cash market is closed when the cash settlement prices are to be calculated, it is a good idea to defer the calculation to a future date at which the cash market allows guaranteed execution of market orders. This allows market-on-expiration orders to be used to conduct arbitrage with the cash market.

2.   Futures markets are sometimes used as the underlying for cash-settled OTC derivatives. An example includes the NYMEX electricity futures considered in the Avista case. When the settlement price of the futures market is used to cash-settle OTC derivatives, it is a bad idea for this settlement price to be based on an average of trading prices during a brief episode of human, eyeball-to-eyeball trading on a traditional trading floor. Such human trading makes it difficult to guarantee execution of a market-on-expiration at the settlement price.

---

[1] The author worked as an expert witness for one of the individual defendants in the Avista case.

3.   Both market-on-expiration orders which replace futures positions and transfers of positions from the futures market to the cash market in the delivery process of a physically-delivered futures contracts have two properties in common with of wash sales: No risks are transferred, and the price of the transaction does not have economic significance. For physical delivery contracts, deliveries are reported as a separate category of transaction, a good practice which allows traders to distinguish between deliveries and "real" volume in which risk is actually transferred. For cash-settled contracts, it would be a good idea to require offsetting market-on-expiration orders for all cash-settling futures positions to be automatically executed and reported as a separate category of volume, economically equivalent to deliveries. Such reporting would make it possible for traders to measure the volume of transactions in which risk is transferred.

4.   Some traders may intentionally choose to let cash-settled positions expire, in which case the traders transfer risk to other traders by having their positions liquidated according to the cash-settlement price formula. To the extent that traders do this, it facilitates efficient use of the cash-settlement mechanism to structure the cash settlement procedure to resemble a trading strategy which a trader might otherwise choose to carry out without resorting to the cash-settlement mechanism. For example, since some traders like to use VWAP (volume-weighted-average price) strategies, the cash settlement procedure could be structured as a VWAP strategy which averages prices over time rather than rather than taking a snapshot of prices at one point in time. For purposes of practical implementation and clean arbitrage, "time-weighted-average price" might be more practical. For example, rather than establishing a cash-settlement price by calculating the value of an index at a single point in time (say, the open), the index could be averaged over a number of openings or averaged continuously over a trading interval ranging from several minutes to several days.

ASSUMPTIONS

Consider a hypothetical economy in which there exists a potentially large number of assets of various kinds. Among these assets are two futures contracts based on the same underlying asset. Make the following assumptions about these two futures contracts. The first assumption, ability of traders to place "market-on-expiration" orders, is the key assumption. The other assumptions are meant to assure enough tradability, solvency, and fungibility to make the contracts perfectly arbitrage-able against each other. I will refer to these assumptions collectively as "microstructure fungibility."

3

(These assumptions are evolving and may be different in a future, less preliminary, version of this paper.):

1. "Market on Expiration Orders": At any point prior to the execution of the cash-settled contract, it is possible for any trader to place a zero-transactions-cost market order to buy or sell an arbitrary quantity of the deliverable asset in the cash market for guaranteed execution at precisely the cash settlement price. This assumption is satisfied if there is a transparent single-price auction which determines the settlement price, but it may also be satisfied even if the settlement price is determined in a non-transparent manner. This assumption does not require that there are no transactions costs for other types of transactions in the cash market. The cash market may in fact have very high transactions fees, otherwise be non-transparent, illiquid, and even subject to manipulative strategies such as corners or squeezes.

2. Identical Contract Terms: The terms of the two contracts are identical except that one contract calls for cash settlement and the other contract calls for physical delivery. In addition to the same underlying asset, the contracts have the same size, delivery date, mark-to-market conventions, and other similar specifications.

3. No Transactions Costs for Futures Arbitrage Transactions: There are no commissions or exchange fees charged to a trader who makes "arbitrage" transactions in which he buys a physical-delivery contract and simultaneously sells a cash settled contract, or vice versa. The assumption does not rule out commissions or exchange fees on other types of transactions. For example, the traders on the opposite sides of the two "legs" of the arbitrage transactions might incur commissions or exchange fees. This assumption also does not require markets to be particularly liquid. It allows for the ask price to be above the bid price, and it allows market impact to be larger when larger quantities are traded. The bid-asked spread may be wide or tight, and markets may be illiquid or illiquid.

4. Identical Exchange Fees and Commissions. The commissions for transactions in the cash-settled and physical-delivery futures contracts are the same. This assumption is meant to imply that a trader pays the same commissions and fees for a given amount of trading volume, regardless of whether the volume is in the cash-settled or physical-delivery contract. This assumption does not rule out volume discounts, nor does it require all traders to have the same commission rate. For example, a trader who trades 100 contracts of the cash-settled contract and 500 contracts of the physically-delivered contract pays the same fees he would pay if all 600 contracts were cash-settled (or physically-delivered).

5. Fungible Delivery with the Cash Market: The delivery mechanism for the futures contract with physical delivery is identical to the delivery mechanism

4

for the underlying cash market. In particular, when the contract with physical delivery expires, traders with long outstanding positions are required to make delivery to traders with short outstanding positions by entering into the cash market clearing and settlement system transactions in which the longs deliver to the shorts the underlying asset at the final settlement price for the futures contract with physical delivery.

6. Netting of Cash Market Deliveries: If a trader has obligations both to make and take delivery in the cash market at the same delivery date (including deliveries on physically-delivered futures contracts), the quantities to be delivered are netted to the extent possible, and price differences are settled in cash. For example, if a traders has an obligation to deliver 100 contracts at a price of 102 and to take delivery of 60 contracts at a price of 98, then the trader makes a delivery of 40 contracts (100 – 60) at 102 and collects as a cash payment of 240 his profit of 4 (= 102 – 98) on the remaining 60 contracts which are netted.

7. Clearing House Solvency: Both the clearing house for the futures market and the clearing house for the cash market guarantee performance on contracts and never fail themselves. I include in this assumption the idea that the futures exchange will not change the rules of the game while it is being played. For example, the clearing house will not force liquidation of contracts at arbitrary prices or require margins on margin-free positions (to force a member to fail). I view breaking the rules as a form of clearing house failure.

8. Identical Settlement Prices: At the close of trading each day, a settlement price is calculated for futures contracts and all futures positions are marked to market in cash to this settlement price. If the best bid for each of the two contracts is less than the best ask of both of the two contracts, we assume that both contracts are given the same settlement price. For example, if the cash-settled contract is bid at 95 and offered at 101 and the physical delivery contract is bid at 98 and offered at 103, then the settlement price for both contracts is the same. For convenience (although not necessarily required for the proof below), we also assume that the settlement prices respect the bid and the ask at the close. Thus, both contracts might be settled at 98 (the best of the two bids), or both might be settled at 101 (the lowest of the two ask prices), or both might be settled at the same price anywhere between 98 and 103, e.g. at 100. Note that this assumption does not require the futures contracts to be settled at the same price. It merely requires the futures exchange to try to settle the contracts at the same price if that if feasible, in the sense of being consistent with the bids and offers posted at the close of trading.

9. Marking-to-Market: The cash flows associated with marking to market at the final settlement price for both futures contracts and the cash flows associated with deliveries in the cash or futures markets at expiration of the futures

contracts all occur at the same time. This assumption, together with the previous one, makes the final settlement price in the physical-delivery contract a matter of indifference to all traders. If the final physical-delivery settlement price is raised by one dollar, longs will receive an extra dollar when their final position is marked to market and they will pay an extra dollar when delivery is made. Since these two cash flows occur on the same date, they have no economic effect on the trader taking delivery. Analogously, they have no effect on the short trader taking delivery either.

10. Bids and Offers Subject to Immediate Acceptance. If a trader posts a bid or an offer, this bid or offer can be accepted immediately by other traders. Within one contract, a trader is not allowed to post an order inconsistent with existing orders at the time, i.e., a trader is not allowed to offer at 99 when the market is bid at 101. At the close of trading each day, we require that traders be given an opportunity to accept the last bids and offers before remaining unaccepted bids and offer become "final." This assumption allows arbitragers to eliminate arbitrage opportunities which may exist in the market just before the close. Note that this assumption does not require that all traders have equal access to posting or accepting bids and offers. Some traders, e.g. floor traders or traders with faster computer technology, may be able to post, accept, or cancel bids and offers faster than others. Furthermore, this assumption does not require that the futures markets conduct an efficient single-price auction at the open, at the close, or at any other time.

11. Risk-free Legging of Spreads: If a trader wishes to accept simultaneously both a bid and an offer, possibly on different contracts, he can do this without risk that he will be executed on only one leg of the transaction. For example, if a trader sees the cash-settled contract bid at 101 and the physical-delivery contract offered at 99 (a situation we claim below to be an arbitrage opportunity), then the trader is allowed to accept both simultaneously. As a practical matter, this means that the trader can press a button on his computer to try to accept both simultaneously. He will then receive a confirmation either that both legs of the trade executed or that neither leg of the trade executed. (This assumption is meant to eliminate "trade-throughs" of economically identical contracts.)

12. No Position Limits on Arbitrage Positions: There are no position limits on spread positions in which a trader is long the cash-settled contract and short equal quantities of the physical delivery contract when both contracts expire at the same time (or vice versa). There may be position limits on net long or short futures positions or long-short positions involving the cash market.

13. Zero Margin on Futures Arbitrage Positions: Long and short positions in the cash-settled and physical-delivery futures contracts with the same expiration date are netted so that spread positions incur no margin. For example, if a trader is long 20 cash-settled contracts and short 15 otherwise identical

physical-delivery contracts, he is margined as if he were merely long 5 cash-settled contracts (20 – 15 = 5). The netted spread position—long 15 cash-settled contracts and short 15 physical-delivery contracts—has zero required margin.

14. Traders Prefer More to Less: If one transaction generates more economic value than another, the trader will choose the transaction which generates more rather than less. This assumption gives traders an incentive to eliminate arbitrage opportunities between the two futures contracts. This assumption does not require that markets be competitive. Indeed, all traders may recognize that their trading has an effect on market prices, and large traders may recognize that they have significant market power. Traders may attempt to exercise their market power in various ways, for example, by attempting to monopolize the deliverable supply of the asset (as in a corner or squeeze), or quoting wide bid-asked spreads when supplying liquidity to other traders.

THE ECONOMIC EQUIVALENCE BETWEEN CASH SETTLEMENT AND PHYSICAL DELIVERY

Under the assumptions above, I claim that the cash-settled contract and the physical-delivery contract must have identical prices. I do not assume that markets are particularly liquid. Instead, I assume that there is a best bid price and best offer price for each contract. "Identical prices" means that the maximum of the best (highest) bids for both the cash-settled and physically-delivered contracts cannot be greater than the best (lowest) offers for the two contracts.

To prove that prices are identical, suppose that there is an arbitrage opportunity, e.g., the best bid for the physical-delivery contract, say 101 for 20 contracts, is greater than the best offer for the cash-settled contract, say 99 for 5 contracts. An arbitrager will attempt to sell the physical-delivery contract for 101 and simultaneously buy the cash-settled contract for 99 as many contracts as possible, 5 contracts in the case (min(20,5) = 5). Because I have assumed risk-free legging of spreads, this attempt will either result in no transaction (an "unable") or in a transaction in which the trader buys the cash settled contract for 99 (or lower) and sells the physical-delivery contract for 101 (or higher) for some number of contracts between 1 and 5. If the attempt results in no transaction, this means that the arbitrage opportunity disappeared because the arbitrager did not act on it quickly enough. Perhaps another trader got there first.

Suppose the attempt to capture the trade is successful and the arbitrager buys 5 cash settled contracts at 99 and sells 5 physical-delivery contracts at 101. The arbitrager then sits on this position—long 5 cash-settled contracts and short 5 physical-delivery contract—until a shortly before the expiration of the contracts.

To make his attempted arbitrage transaction successful, the arbitrager must unwind his transaction in a manner that locks in an arbitrage profit. He can do this in the

following manner. Shortly before the expiration of the cash-settled contract, the arbitrager places a "market on expiration" order to buy 5 contracts in the cash market at precisely the price at which the expiring cash-settled contract is going to be settled. When the arbitrager places this order, he does not know the price at which this order will be executed. It may be far higher or far lower than prices a few moments before he places the order. Suppose the order executes at a "high" price of 150. Since we have assumed that the order executes at the price used for cash-settlement of the cash settled futures contract, the cash-settled futures contract is settled at the same price of 150.

Now let us examine the cash flows per contract which the arbitrager receives from the transactions he has made. He purchases the cash-settled contract for 99 and sells it for 150. This generates a positive cash flow, or profit, of 51, calculated as $150 - 99 = 51$. He sold (shorted) the physical delivery contract at 101, and this contract had a final settlement price of 150, generating a loss (negative cash flow) or 49, calculated as $101 - 150 = -49$. Thus, his net profit per contract from the futures transactions is 2, calculated as $51 - 49 = 2$, the same profit he "locked in" when he initiated the arbitrage transaction in the first place.

The cash flows from the physical delivery mechanism are zero. On his physical delivery contract, the arbitrager has an obligation to deliver 5 contracts at a price of 150. From the market-on-close orders, the arbitrager has an obligation to take delivery of 5 contracts at a price of 150. According to our assumptions above, the delivery obligations net against each other, leaving the arbitrager with no obligation either to make or to accept delivery. If there were no netting, the arbitrager would be required both to make and to take delivery of 5 contracts. If delivery collateral were scarce, or there were fees associated with making or taking delivery, this would impose costs on the arbitrager. It might create "fails" or defaults in the physical delivery process. Since the prices of the purchase and sale are both identical, 150 in this case, they are netted against each other, and the arbitrager receives a cash payment of precisely zero from the cash-market delivery mechanism.

Although the arbitrager locks in an ultimate profit of 2 per contract on his arbitrage transaction, it is conceptually possible that marking to market differences between the two contracts results in these cash flows occurring at different times. If this were the case, the arbitrager would also have to consider interest on cash flows resulting from different timing of the transactions. We claim that there can be no cash flow differences between the two contracts, because all of the contracts will be marked to market at the same prices on the same dates. To see this, consider first the trading process which determines the final settlement prices of both the cash-settled and physical-delivery contracts precisely when the contracts expire. Since there is only one set of cash flows after this point, and these cash flows occur at the same time, we do not need to worry about differences in timing of cash flows after this point. Our argument in the paragraph above guarantees that arbitrage opportunities will be eliminated at the final settlement point, in the sense that the highest of the best bids in the two contracts will be no higher than the lower of the two offers. This makes it feasible for the exchange to settle the two contracts at identical prices which respect these bids and offers. According

to the assumptions we have made above, they will in fact do so. Thus, at expiration, the final settlement price of the cash-settled contract will be equal to the final settlement price of the physical-delivery contract.

Now, the day before the expiration day, all traders can anticipate that the final settlement prices will be the same on the last day. This implies that the timing of cash flows on any attempted arbitrage transactions (which take place on the day before the final settlement day) will be identical. This arguments extends backwards (by induction) to prove that settlement prices will be the same on each day in the life of the contract. If traders anticipate that cash flow timing will be the same on all future days, forces of arbitrage will induce them to eliminate arbitrage opportunities today in such a way as to make settlement prices the same today as well. When they do so, such traders will not have to worry that they lose interest on money due to differences in marking the two contracts to market. We have thus proved that the timing of cash flows on the two contracts is the same, because settlement prices are the same on every day in the lives of the two contracts. In other words, the trader who buys cash-settled contracts at 99 and sells physical-delivery contracts at 101 receives a cash flow of +2 per contract as soon as the day's trades are marked to market. By assumption, he posts no margin; therefore, he needs no capital to enter into the transactions. Since settlement prices are the same on subsequent days, his cash flows are precisely zero on every subsequent day. At delivery, netting in the cash market the delivery mechanism works in such a manner that he needs no collateral or cash to make or take delivery. Thus, capital requirements for this riskless arbitrage transaction are zero, and any trader can carry out such a strategy.

Although the numerical example is specific, it clearly generalizes to arbitrary quantities and prices. Moreover, the arbitrage strategy operates equally in reverse, with the arbitrager buying the physical-delivery contract and selling the cash-settled contract rather than selling the physical-delivery contract and buying the cash-settled contract.

The Modigliani-Miller Theorem of Cash Settlement. My proof above establishes that, under the assumptions made, the contract with physical delivery is economically exactly equivalent to the contract with cash settlement. In fact, the arbitrage concept we have established is much stronger than the usual arbitrage relationships found in the finance literature. In the finance literature, there are two distinctly different types of arbitrage relationships, which I will refer to as "static arbitrage" and "dynamic arbitrage." To distinguish the type of arbitrage above from the other two, I will call the concept I am using "microstructure arbitrage."

An example of static arbitrage is the usual derivation which shows that a futures contract equals the cash price of the underlying deliverable instrument, adjusted upward for interest and storage costs and adjusted downward for dividends. This arbitrage relationship is usually demonstrated under the assumption that there is one "price" at which both the futures and cash instrument can be bought or sold. It demonstrates that the futures contract is a "redundant" asset, in the sense that it can be economically replaced by a cash instrument with associated borrowing and lending. This arbitrage relationship does not take into account the possibility that futures contracts may improve

economic equilibrium by offering lower trading costs and greater liquidity than the cash market. Instead, it is typically assumed that there are zero trading costs, markets are competitive, there is a single price at which traders can buy and sell, and there is a single interest rate at which traders can borrow and lend.

An example of dynamic arbitrage is the dynamic trading strategy which leads to the Black-Scholes model of option pricing. In addition to the assumptions made for static arbitrage, dynamic arbitrage arguments require that traders agree unanimously that particular zero-probability events will not in fact occur. In the Black-Scholes model, traders must agree that the path of the stock price will have a realized volatility equal to a particular quantity, called "sigma" in the Black-Scholes model. Since volatility is not precisely predictable, the Black-Scholes model does not hold precisely in real-world trading.

The microstructure arbitrage argument I have given above does not require competitive markets, does not require zero transactions cost, does not require one price at which all traders can trade, does not require traders to agree on zero-probability events (other than that the assumptions I have made failing to hold, e.g. all participants agree that the clearing house will not fail), does not require that markets have any particular degree of depth, does not require that the cash market function efficiently, and does not require that borrowing or lending markets function at all. Since I do not assume that credit markets function at all, I have assumed no mechanism that guarantees the usual "convergence" between cash prices and futures prices. I have only assumed that the futures contracts are priced identically to each other, not that they are priced in any particular manner in relationship to the cash market. Since I have assumed that the costs of trading the cash-settled and physical-delivery contracts are the same, these two contracts are fungible in the sense that a trader will achieve exactly the same economic outcome if he does all of his trading in physical delivery contracts, does all of his trading in cash-settled contracts, or uses a combination of both types of contracts. If one of the two contracts is discontinued and positions are replaced by the other type of contract, there will be no effect on the economic allocation of resources. What is required is that the traders understand the assumptions I have made above, which lead to the understanding that the contracts are exactly equivalent for trading purposes, even in imperfect market with imperfect liquidity and exercise of monopoly power affecting prices.

For example, under the assumptions I have made, it is possible that transactions costs in the futures market are far lower than transactions costs in the cash market, as a result of which futures contracts have very high trading volume relative to the cash market. If the futures market were eliminated, this might result in a significant welfare loss to society, due to the higher costs of risk-transfer on trades being forced into an illiquid and perhaps non-transparent cash market. But if only one of the two futures contracts were eliminated (e.g. physical delivery abolished), but not the other (e.g., cash settlement allowed), then all traders could re-label every bid, every ask, and every transaction in the abolished contract with an economically equivalent transaction in the

allowed contract. After this re-labeling, all transactions would be economically equivalent to what happened with trading in both contracts allowed.

For this reason, microstructure arbitrage leads to a strong "Modigliani-Miller Equivalence" results: Under the assumptions I have made, the economic allocation of resources will be equivalent, regardless of whether only cash-settled futures contracts are traded, only physical-delivery futures contracts are traded, or both types of futures contracts are allowed. This equivalence will occur regardless of whether markets are liquid or illiquid, competitive or non-competitive, manipulated or non-manipulated. Whatever the outcome with one set of contracts, it will be the same with the other.

Another way of looking at the result is to realize that the trading platform and clearing house could make the contracts completely fungible by, say, matching a buy of the cash settled contract directly against a sell of the physically delivered contract. Indeed, a trader would not need to indicate whether he wants his contract to be a cash-settlement contract or a physical delivery contract until shortly before expiration. After the trader declares the type of contract he wants, the exchange assigns delivery notices to the traders who want physical delivery and places market-on-expiration orders for traders who want cash settlement. This leaves the clearing house with not net economic exposure since the settlement price of both contracts is the same. The market-on-expiration orders will give the traders who ask for cash settlement something economically equivalent to physical delivery. They can undo the physical delivery by making trades in the cash market on terms identical to traders who have physical delivery contracts, or by making additional trades in the futures contract, at or before expiration. They have the same set of economic opportunities as traders choosing physical delivery.

RISK-TRANSFER AT EXPIRATION

Futures trading is a mechanism for transferring risk. Although some futures markets are highly liquid, real futures markets do not offer a risk-transfer service for free. Instead, traders face costs associated when transferring risks to other traders. There are exchange fees and commissions, positive bid-asked spreads, and market impact or slippage costs associated with larger transactions. In other words, traders who demand liquidity for transferring risk are well-advised think about what they pay for this liquidity. Some traders may receive a reward for supplying liquidity. Usually, we think of an impatient trader as being willing to pay to demand liquidity and a patient trader as intending to profit by supplying liquidity. But thinking about the demand and supply of liquidity is not straightforward (and not the main subject of this paper). Patient traders who demand liquidity may trade very slowly and look like suppliers of liquidity. Impatient traders with private information may appear to demand liquidity, but they intend to make a profit doing so, as a result of which the traders who supplied the liquidity may unexpectedly lose money. This occurs when a market maker is "picked off" by an informed trader with short-lived information.

When a trader uses a market transaction to change his risk exposure, the trader demands or supplies liquidity. For example, a trader long 30 cash-settled contracts and short 10 physical-delivery contracts (with no position in the cash market) can be deemed to have an exposure of +20 contracts (30 – 10 = 20) in the futures market. If he sells 5 of his 30 cash-settled contracts, his exposure is reduced by 5 contracts to +15. The sale of 5 contracts results in a transfer of -5 contracts of risk, the minus sign representing that contract risk was sold, not bought.

For the purpose of measuring exposure to risk, we might also aggregate positions in futures contracts with positions in economically similar cash market instruments. For example, if the same trader long 30 cash-settled futures contracts and short 10 physical-delivery futures contracts is also short cash market instruments similar to 18 futures contracts, his net market position—combining futures and cash positions—is only +2 contracts (30 – 10 – 18 = 2). In addition, he has a "basis position" (cash minus futures) of 18 contracts. He is incurring two sorts of risk, "market risk" and "basis risk." These risks are different. For example, if the positions are in long-term bond futures and long-term cash market bonds of various coupons and maturities, the market risk is associated with overall movements in long-term interest rates. The basis risk is associated with movements in short term interest rates, i.e. the repo rates (including specials) associated with borrowing collateral to finance the short position in cash-market bonds. The trader may be vulnerable to a squeeze if the bond he is short is not the one deliverable into the futures contract. This vulnerability is part of his basis risk. There is no basis risk associated with offsetting positions in otherwise equivalent cash-settled and physical-delivery futures contracts, because we showed above that these contracts always trade at the same equivalent prices.

Making or taking delivery of a physical-delivery futures contract does not result in any transfer of risk. For example, if our trader short the 10 physical delivery contracts chooses to deliver them, his short position moves from the futures market into the cash market, but he is still short these same ten contracts and still has the same market risk.

Similarly, replacing an expiring position in a cash-settled contract by placing market-on-expiration orders to buy equivalent quantities in the cash market for each expiring long positions and to sell in the cash market equivalent quantities of expiring short positions does not result in any transfer of risk. The market risk simply moves from the futures market to the cash market at the moment the contracts are cash-settled, but the overall market risk remains the same.

In the example above, if the trader long 30 cash settled contracts and short 10 physical delivery contracts replaces the 30 cash-settled contracts with market-on-expiration orders and makes delivery on the 10 short physical-delivery contracts, he would wind up with an unchanged cash market position of +2 futures contract equivalents. Note that my assumptions imply that 10 of the 30 cash settled contracts he replaces by buying would be netted against his ten short positions he is obligated to deliver against the physical delivery contract. Thus, he would take delivery of 20 contracts and make delivery of none. If the 18 contract short position in the cash market

were in instruments exactly equivalent to the 20 net futures deliveries, these would also be further netted and the trader would wind up with net deliveries of 2 contracts, and no basis risk. If the 18 contract short position in the cash market were instruments not identical to the cash market instruments underlying the futures contract, the trader would potentially wind up with a long position of 20 contracts of one type of bond and a short position of 18 contracts of another type of bond. He would then have 18 contracts of basis risk involving cash market bonds, but he would still have 2 contracts of market risk. For example, if the trader were being squeezed on his short position in cash market bonds, and if the futures contract allowed substitution of one bond for another in the delivery process, based on a "cheapest-to-deliver" concept, then the squeezed trader might unexpectedly be delivered different bonds on his long futures positions from the ones he was short. This would leave him vulnerable to further financing his short cash-bond position at "special" rates in the repo market.

If a trader with a cash-settled contract wants to change his risk exposure at the time of expiration, he must do something different from replacing his 30-contract long position by placing a market-on-expiration order to buy 30 contracts in the cash market. If he wants to increase his exposure by 10 contracts, he must supplement the 30-contract market on expiration buy order with 10 additional purchases, either in the cash of futures markets. Including the 30 contract market-on-expiration order, he would need to purchase 40 contracts overall If he wants to reduce his exposure by 10 contracts, he must either reduce the 30-contract market-on-expiration order by ten contracts, or supplement the 30-contract market-on-expiration order by selling 10 additional contracts in the cash or futures market. According to our assumptions, the trader pays nothing to keep his underlying exposure to risk the same by placing the 30-contract "market-on-expiration" order to buy. But if the trader chooses to do something different, e.g. by increasing or decreasing his risk exposure 10 contracts, he would be well-advised to think about how much he is paying or being paid to demand or supply 10 contracts of liquidity.

For example, suppose that the trader chooses to reduce his risk exposure by 10 contracts, at exactly the moment the cash-settled contract expires. He could do this by reducing the size of his "market-on-expiration" order in the cash market from 30 contracts to only 20 contracts. Interpreting such a 20-contract "market-on-expiration" order can be confusing. The trader is a net seller of 10 contracts, and not a net buyer at all. Since he is a net seller of 10 contracts, he is putting downward pressure on prices. Yet his order is an order to buy, not an order to sell. This feature of cash-settled futures contracts can lead to confusion.

If the trader places no "market-on-settlement" order at all, he demands liquidity by becoming a net seller of 30 contracts. His long cash-settled futures position of 30 contracts is liquidated at the settlement prices in the cash market, and his market risk exposure is reduced by 30 contracts. In this sense, appearing to "do nothing," by placing no orders, actually represents a demand for market liquidity. The trader is in fact placing downward pressure on prices, and should expect to incur a transactions cost as a result of doing so. A naive trader who expects his 30 contracts to be executed without any adverse effect on prices is likely to be surprised that he has to pay for the liquidity he is

demanding. A trader should not expect a speculative market to offer its risk-transfer service for free.

What kind of trading behavior should we expect to see at the expiration of a cash-settled futures contract? Suppose that at the moment of expiration, there is some single market-clearing price in the cash market at which all traders are happy with their existing positions in the market, and no trader wishes to change its market exposure at that price. Then every trader will have an incentive to place "market-on-expiration" orders to replace their entire futures positions with positions in the cash market. All of the traders with long positions will place buy orders to replace their long positions, and all of the traders with short positions will place sell orders to replace their short positions. Since the number of long positions is always by definition exactly equal to the number of short positions in a futures contract, the buy orders will net out against the sell orders exactly, and there will be no net demand for liquidity in the cash market as a result of all traders replacing their expiring futures positions with cash market positions. These traders will all be faced with settling positions in the cash market. How difficult such settlement is depends on how the underlying cash market operates. For example, if the cash-settled futures contract is S&P 500 futures, then each trader will be faced with the prospect of making or taking delivery on baskets of odd-lots in approximately 500 individual stocks. In my assumptions, I have not assumed that making or taking delivery in the cash market is cheap or easy. If making or taking delivery in the cash market is complex and expensive process, I expect the traders who will choose to place the "market-on-expiration" orders will on average be experienced, sophisticated, well-capitalized entities with good interfaces into the cash market. This expectation is no different from a contract with physical delivery. If making or taking delivery on a physical delivery contract is complex and expensive, I would also expect traders who make or take delivery to be experienced, sophisticated, well-capitalized entities with good interfaces into the cash market as well.

Suppose that some of the traders with positions in the market at expiration are not sophisticated enough to navigate the process of trading in the cash market. Such traders will demand liquidity from the cash market by not placing market-on-expiration orders. The forces of supply and demand will determine prices in the cash market at which the cash-settled contracts will be liquidated. The assumptions I have made in no way suggest that the futures contract will be well arbitraged against the cash market. For example, my assumptions allow high exchange fees and commissions associated with cash market transactions which are not market-on-expiration orders, wide bid-asked spreads and great illiquidity in cash market trading, a non-transparent cash market without price reporting or reporting of current bid and ask prices, unequal access to the market across traders, abuse of market power by various types of traders, and poorly functioning credit for financing arbitrage positions. Under these circumstances, a naïve futures trader who allows his cash settled-futures position to expire and be cash settled may obtain prices he considers poor. In other words, the cash market may supply liquidity to a trader who demands it by not replacing positions with market-on-expiration orders at a very high price. Before having pity on such a trader, notice that a trader who allows his cash-settled futures position to be liquidated by cash settlement, without placing market-on-

14

expiration orders, is in exactly the same position as a trader in an otherwise equivalent physical-delivery contract who waits until after the futures contract expires, takes an unwanted delivery, then dumps the unwanted delivery into the cash market at whatever price the cash market happens to charge for this demand for liquidity. The price for this liquidity may be very high. Both the unsophisticated trader in the cash-settled contract and the unsophisticated trader in the physical-delivery futures contract are well-advised to liquidate their futures positions in the futures market before taking deliveries they do not have appropriate experience to deal with.

THE CONFUSION IN THE ACADEMIC LITERATURE
ON CASH SETTLEMENT

The academic literature has become confused concerning the issue of whether cash-settled contracts offer something different from physical-delivery contracts. On the one hand, there are papers which recognize that cash settled contracts can be arbitraged against the cash market in a manner similarly to the way physical delivery contracts can be arbitraged. Papers by Dutt and Harris (2005), Stoll and Whaley (1997), Pirrong (2001), and Kumar and Seppi (1992) all recognize that there is a sense in which cash settled contracts are equivalent to physical-delivery contracts. All of this paper creates confusion, however, by analyzing cash-settled contracts in a manner which avoids confronting the equivalence which they all recognize and which I discussed above.

Let us consider first an argument which is conceptually equivalent to one by Dutt and Harris (2005). Dutt and Harris examine a cash-settled contract in which the underlying is a weighted index of many assets, and they introduce conversion factors to convert prices to cash amounts and cash amounts to contract units. I will simplify his argument by assuming there is only one underlying asset, and I will ignore the various conversion factors by setting them all to 1. Suppose that just before the cash settled contract expires, a trader has a position of X contracts. Suppose that this trader believes that he faces a linear residual supply curve in the cash market against which he can trade at expiration by placing market-on-expiration orders. Since he thinks he knows the residual supply curve against which he trades, the trader thinks he knows the price at which a market-on-expiration order of any given size will be executed. Suppose the trader believes the fundamental value of the asset is 100, and he believes that if he acquires positions in the cash market by trading at expiration, then he will be able to liquidate these positions at the fundamental value of 100. The slope of the residual supply curve is denoted λ.

Dutt and Harris go on to make the assumption that if the trader does not place a market on close order at all, and thus allows his entire position to cash-settle, then the trader expects the cash settlement price to be equal to the fundamental value of 100. As I will discuss below, I believe this assumption is unreasonable. As a result of this assumption, the trader believes that the cash-market price P, as a function of the size Q of his market-on-expiration order, is given by

$$P = 100 + \lambda*Q.$$

Under these assumptions, the total revenue a trader earns by placing an order of size Q is given by

$$\text{Total Revenue} = (100 + \lambda*Q)*X - (100 + \lambda Q)*Q + 100*Q.$$

This equation says that the trader receives a payment of $100 + \lambda Q$ for each of his X cash-settled contracts, pays $100 + \lambda Q$ for each of the Q contracts he buys in the cash market, and receives 100 for each of the Q contracts he sells later in the cash market. Since total revenue is a quadratic function of Q, it is a simple calculus problem to show that the order size Q which maximizes total revenue is $Q = X/2 = 10$. In fact, for any positive value of the inverse depth parameter λ—regardless of how small—the amount purchased Q is always exactly one-half the size of the futures position which is about to be cash-settled. By placing a market-on-expiration order of size $Q = X/2$ rather than placing no market-on-expiration order at all, the trader increases his profit by λX.

Dutt and Harris refer to choosing Q=X/2 instead of Q=0 as "manipulation" resulting from the trader's exercise of market power. Their intuition—which also seems to be shared by Pirrong (2001), Kumar and Seppi (1992), and Stoll and Whaley (1997), but not by me—is that the trader "manipulates" the cash market for the purpose of benefiting the cash settlement price of his futures position. According to the confusing intuition of these authors, relative to placing no order, the optimal cash market part of the transaction reduces profits by half as much money as the futures part of the transaction increases profits. Kumar and Seppi (1992) refer to this as "punching the settlement price" in the cash market for the purpose of benefiting the futures position.

The CFTC also shared this confusion in the Avista case. In this case, there were OTC derivatives cash-settled against the NYMEX electricity futures contract, so NYMEX electricity futures play the role of the "cash market" and the OTC derivatives play the role of the "futures market" as far as cash settlement is concerned. The CFTC argued as follows: "To the extent that Avista Energy's Traders could distort the price for futures contracts with an order at the Close on Options Expiration Day that was smaller than the positions created by Avista Energy's OTC derivatives contracts, Avista Energy's Traders thought that they might be able to profit via an artificially created increase in the value of its OTC derivatives contracts. This would be so because the value of those contracts would exceed any losses incurred from buying NYMEX Western U.S. electricity futures contracts at an artificially high price or selling futures contracts at an artificially low price." This captures precisely the "punching-the-settlement-price" intuition of Kumar and Seppi (1992) as well as the Dutt and Harris intuition.

"The Code of Market Conduct" in the FSA's Handbook (1.6.15(2)) contains the following example of "market abuse" falling into the category of "manipulating transactions:" "A trader *buys* a large volume of *commodity futures*, which are *qualifying investments*, (whose price will be relevant to the calculation of the settlement value of a *derivatives* position he holds) just before the close of trading. His purpose is to position

the price of the *commodity futures* at a false, misleading, abnormal or artificial level so as to make a profit from his *derivatives* position." This example also seems to capture intuition consistent with the "punching-the-settlement" argument of Kumar and Seppi (1992).

What is wrong with this intuition? To see why this intuition is misleading, let us examine the economic substance of the transaction, taking into account our proof above that cash settled contracts are equivalent to ones with physical delivery. As we showed above, when the trader places a market order to buy Q contracts, he is increasing his market exposure by an amount equivalent to purchasing $Q - X$ contracts if $Q > X$, or decreasing his market position by an amount equivalent to selling $X - Q$ contracts if $Q - X < 0$. If $Q = X$, the trader is neither demanding nor supplying liquidity. Thus, when the trader places a market-on-expiration order to buy only half his long position, $Q = X/2$ contracts ($X > 0$), he is in economic substance a net seller of $X - Q = X/2$ contracts.

What is the correct economic intuition motivating why the trader would like to be a net seller of $X/2$ contracts? If the trader does not demand or supply liquidity at all, he places a market-on-expiration order to buy X contracts, and the cash-settled futures contract settles at a price of $100 + \lambda X$, an amount which the trader perceives to be λX above its fundamental value. Evidently, the rest of the market is placing upward pressure on prices relative to what the all traders believe is likely to prevail later in the cash market. Like any supplier of liquidity, such as a typical large market maker in a normally functioning market, the trader sees an opportunity to make a profit by being a net seller at this attractively high price. The trader knows that the more he sells, the more he depresses the price. The trader maximizes his profits by being a net seller of half the number of contracts necessary to drive the price all the way back to its perceived fundamental value of 100. Thus, what the authors mentioned above refer to as "manipulation" is actually equivalent to ordinary market making by a market maker large enough to supply liquidity so aggressively that it would be feasible for him to drive his own profits to zero. Of course, a market maker will not willingly supply liquidity so aggressively that he tries to drive his profits from doing so to zero. Instead, he rations the amount of liquidity he supplies to maximize his profits. Precisely this issue was discussed in the Indiana Farm Bureau (****) case, where it was ruled (in the context of physical-delivery futures contracts) that it is not manipulative for a trader to seek to make a profit by restricting the amount of liquidity he supplies to the market. If the Indiana Farm Bureau logic is applied to cash settled futures contracts, it is not manipulative for a trader to replace his expiring cash-settled contracts by purchasing X contracts in the cash market. I believe that it is incorrect to refer to the behavior in the example discussed above as manipulation.

The correct intuition becomes clearer when we re-write the problem in notation which consistent with the conventional use of supply and demand curves and conventional microstructure intuition. The relationship $P = 100 + \lambda Q$ looks like a supply curve, but it is not a conventional supply curve. If the trader does not engage at risk-transfer at expiration, then he places a market-on-expiration order to purchase X contracts. Letting P^* denote this expected price with no risk transfer, we have $P^* = 100$

+ λX. Let L denote the traders supply of liquidity, defined by $L = Q - X$. Then the equation $P = 100 + \lambda Q$ can be equivalently re-written in the form of a conventional supply curve as

$$P = P^* + \lambda L.$$

If the trader believes that the positions he acquires L can be liquidated at a price of 100, his profit from acquiring L contracts of risk in the market is given by

$$\text{Profit} = (P - 100)\, L = (P^* + \lambda L - 100)^*L.$$

This equation is also quadratic in L, and the value of L which maximizes it is $L = -(P^* - 100)/(2\lambda)$. Using the definition of P*, it is easy to see that $L = -X/2$, i.e., the trader sells X/2 contracts. Note that number of contracts sold depends only the depth of the market λ and on price difference between the fundamental value of the asset, 100 in this case, and the price of the asset if the trader does not engage in risk transfer, P* in this case. Since the trader believes that the asset will be overvalued by P* - 100 if he does not trade, he supplies liquidity to the market by selling exactly half the quantity which would drive price exactly to its fundamental value. In other words, he acts like any market maker trying to make a profit. There is nothing intrinsically manipulative about such a trading strategy. For example, if the market is very deep, λ is likely to be a tiny number, implying that P* differs only very slightly from the fundamental value of 100. The tiny value of λ may be the result of competition among numerous market makers, none of whom drives the price all the way to fundamental value because all of them want to make a living. Furthermore, the quantity L may also be very tiny, e.g. 1 or 2 contracts in a broad deep market.

The reason that the actual quantity traded X/2 depends on the size of the trader's position is that Dutt and Harris assume in their example that P* differs from 100 by an amount proportional to the size of the trader's initial position. This assumption is based on the underlying assumption that the liquidation price will be equal to the fundamental value of 100 if the trader liquidates his entire position by allowing it to cash-settle without placing a market-on-expiration order. Is this assumption reasonable? I believe that there is a confused regulatory agenda hiding behind this assumption. The regulatory agenda begins with the proposition that a trader with a position in a cash-settled futures contract should liquidate his position by allowing it to cash settle, without placing market-on-expiration orders. This proposition is economically equivalent to the proposition that a trader in a physical-delivery contract should not take delivery. Both propositions are inconsistent with the behavior of a conventional arbitrager. Consider, for example, a conventional arbitrager with a long position in the futures market and an equal an offsetting short position in the cash market. Such a conventional arbitrager would often want to take delivery in a physical-delivery contract or do the equivalent by placing market-on-expiration orders. It is precisely this behavior that enforces an arbitrage relationship between the cash and futures markets. If the principle is adopted that traders should not take deliveries—or, equivalently, should not place market on expiration orders to replace cash settling positions—the link between the cash market and

the futures market is fundamentally broken. Instead, the forces of supply and demand would be replaced by a debate among lawyers and regulators concerning what the fundamental value of the asset should be. Dutt and Harris try to short-circuit this debate by assuming that all traders agree that the fundamental value of the asset is 100.

What happens if we allow each trader to exploit his monopoly power optimally in the context of the Dutt and Harris example? To keep things simple, consider two traders with positions in the market. One trader is long 20 contracts just before expiration and the other trader is short 20 contracts. There are also numerous other traders in the market, but none of them have positions. Suppose all traders agree that the correct, fundamental value of the asset is 100. Let us propose to both traders that they allow their positions to cash settle, and let us assume that all traders agree that if both traders liquidate their positions by allowing them to cash-settle, without placing market-on-expiration orders, that the price will be equal to the fundamental value of 100. In the simple example above—taking as given the assumption that the other trader is going to liquidate half of his position—each of the two traders has an incentive to exploit his monopoly power optimally by allowing only half of his position to cash-settle (replacing the other half with market-on-expiration orders or deliveries). Each trader conjectures that by supplying less liquidity to the other trader, he will be able to transact at favorable prices. To exploit his monopoly power optimally, the trader long 20 contracts chooses to buy 10 contracts and the trader short 20 contracts chooses to sell 10 contracts. In other words, both traders "turn down" the proposal to allow all of their positions to cash settle. What happens if each trader believes that the other trader will liquidate half his position? Then each trader will actually find it optimal to liquidate only half of half his position. The trader who is long 20 contracts will conjecture that the trader short 20 contracts will sell 10 contracts, thus demanding 10 contracts of liquidity. This will lead the trader long 20 contracts to supply 5 contracts of liquidity by placing a market order to buy 15 contracts. Following the same logic, the trader short 20 contracts will sell 15 contracts.

It is easy to see that the only set of conjectures which can be fulfilled in equilibrium are that each trader replaces his entire position, the trader long 20 contracts by buying 20 contracts and the trader short 20 contracts by selling 20 contracts. If each trader demands to transfer half as much risk as the other trader expects, a trader's demands for risk transfer can only be consistent with the other trader's expectations if the demand for risk transfer is zero. Demand for risk transfer is zero when each trader replaces his entire position. In other words, the correct conjecture each trader makes about price formation is

$$P = P^* + \lambda L.$$

Where P^* is defined by $P^* = 100$, not by $P^* = 100 + \lambda X$. In other words, the equilibrium conjectures are for traders to believe that price will equal fundamental value if they replace their entire positions, not if they allow their entire positions to cash settle. If all traders believe that price will equal fundamental value if each replaces his entire position, then each trader will do so, since doing anything else is less profitable.

19

PUBLIC POLICIY IMPLICATIONS

The main "Modigliani-Miller Theorem" in this paper might be stated as follows: **"Market-on-expiration orders" plus "microstructure fungibility" implies "microstructure equivalence" between cash-settled and physically-delivered futures contracts.** Keeping this equivalence in mind provides a useful perspective for thinking about various public policy issues related to cash settlement.

Corners and Squeezes. Let us suppose that the CFTC, through regulatory pressure, successfully encourages traders to let expiring positions cash settle rather than be replaced with market-on-expiration orders. This may have the unintended consequence of making it easier to corner the supply of an underlying asset.

For example, suppose that a trader builds up a large long position in physically delivered OTC instruments large enough to corner the available supply. Suppose that the trader is simultaneously short equivalent cash-settled futures contracts. Such a combination of positions has not net market exposure and does not amount to a corner position. If the trader pursues the market-neutral strategy of replacing his expiring cash-settled futures position by placing market-on-expiration orders to replace the entire position, he will have offsetting long and short positions in the cash market, and these will offset against each other (assuming the same underlying) with no corner or squeeze occurring.

If the trader succumbs to CFTC regulatory pressure and lets his cash settled futures position cash settle, then immediately after expiration of the cash-settled futures positions, the trader will discover that he has cornered the market! In essence, CFTC regulatory pressure tries to create large amounts of liquidity where none is naturally provided. This makes it easy to corner the market by effectively buying a large position at the instant of cash settlement. Since deep liquidity makes it relatively easy to corner the market, such regulatory pressure makes corners easier to implement.

It also may make corners harder to detect, for two reasons. First, the cornering position is actually in physically delivered OTC instruments. It is my understanding that currently the CFTC does not have as easy access to OTC data on positions and trades as the CFTC has to futures data. Thus, discovering that the asset has been cornered would require that the CFTC collect large amounts of data it might not ordinarily collect. Second, the corner is implemented by **not** placing market-on-expiration orders. This makes it difficult for the CFTC to point to "trades" which are manipulative, if the CFTC considers not placing market-on-expiration orders to be non-trades.

Wash Sales and Volume Reporting. The "market on expiration" orders are used by an arbitrager to replace an expiring cash-settled position with a position in the cash market. Since these orders precisely offset the positions about to be cash settled, use of such orders to "replace" expiring positions in cash-settled contracts results in no net risk-transfer taking place. The position in the cash-settled futures contract is replaced by an

20

economically equivalent position in the physical asset. In this sense, the use of market-on-expiration orders to replace expiring positions in the cash market resembles a wash sale. In one of the examples considered above, the arbitrager has a long position of 5 cash-settled contracts which is going to be effectively "sold out" at the settlement price. Simultaneously, he replaces the position by buying the about-to-be-sold-out position by buying 5 physical delivery contracts at the same price. As in a wash sale, the trader does not care what the price of the transaction is, as long as the two prices are the same. As in a wash sale, no net risk is transferred as a result of both transactions.

The process of taking delivery in a physical-delivery contract also looks like a "wash sale" in which the delivery obligation is transferred from obligations in futures contracts intermediated by a futures clearing house into cash market obligations. Other transactions which have equivalent "wash sale" characteristics include exercises on deep-in-the-money call options with physical delivery.

Certain kinds of "wash sales" are illegal. Presumably, wash sales are illegal because they could potentially be used to create the appearance of volume when no underlying risk transfer is taking place. They could also be used as "money-pass" transactions to move money between accounts with different beneficial owners, or to establish prices which could be used to reduce tax liabilities by creating a pretense that risk-transfer occurred when it in fact did not. To deal with these potential abuses associated with wash sales, deliveries in physical-delivery contracts are not counted as part of regular trading volume. Deliveries are instead reported as a unique category of transaction, so that traders can distinguish between deliveries and trading volume in which risk transfer actually occurs. Similarly, option exercises are reported as a separate category of transaction, to avoid confusion with transactions in which risk transfer occurs. In the case of cash-settled contracts, consistent treatment suggests that it would be a good idea to report market-on-expiration orders which "replace" expiring cash settled positions as a separate category, economically equivalent to deliveries in a physical-delivery contract. They could be called something like "delivery replacement" transactions.

To avoid confusion and promote transparency, it would be a good idea to require that all outstanding positions in cash-settled contracts be replaced with market-on-expiration orders. If a trader did not want to replace his position, he could supplement the market-on-settlement order with a regular order which transfers risk. Consider, for example, a trader long 100 contracts who seeks to convert 80 contracts into cash by seeking liquidity from the market and who seeks to replace 20 contracts with cash positions. To promote avoid confusion and promote transparency, it would be a good idea for the trader to place two orders: a "market-on-expiration" order to buy 100 contracts and thus replace all of his position with an equivalent cash-market position; and a regular order to sell at expiration 80 contracts. The 80 contracts would be reported as "regular" volume, accurately indicating to market participants that 80 contracts of risk transfer was taking place. The 100 "delivery replacement" transactions would be reported in a special category, equivalent to deliveries, indicating that these 100 contracts transfer positions from the futures market to the cash market but do not transfer

underlying risks. The alternative to my proposal is to report 20 contracts of volume. Such a report is misleading in that it conceals 80 contracts of risk transfer taking place. It misleadingly makes it look like a trader is selling 20 contracts when he is actually a net seller of 80 contracts.

My proposal would have the intended benefit of making transparent the transaction in which the trader corners the market by allowing his entire large short position to cash settle. To achieve the corner, the trader would have to place two orders: an automatic market-on-expiration order to sell his entire position, plus another large order to buy the position back. The market-on-expiration order to sell would be reported as a separate category of transaction since it has the characteristics of a wash sale. The large order to buy back the position would be reported as a transaction in which real risk transfer takes place. This large transaction would be easily observable to regulators as potentially manipulative. Not requiring market-on-expiration orders to be reported separately allows the large potentially manipulative trade to remain hidden, as an apparent non-trade.

Risk Management. Requiring market-on-expiration orders to replace all expiring positions simplifies and makes transparent risk management of customers' positions. Firms have an incentive to manage the risks of the customers' futures positions in real time. For example, firms which handle customers' orders have an incentive to prevent the customers from placing orders which result in positions exceeding limits based on the amount of capital the customer has in his account. At expiration of a cash-settled contract, requiring separate market-on-expiration for all existing positions would make management of risk somewhat less transparent, by preventing the possibility that a customer, by doing "nothing," could increase the riskiness of his position by letting cash-settling positions expire. Consider, for example, a customer with a large long position in a nearby immediately cash-settling position and short an equivalent position in a deferred contract which will not cash-settle for a long time. A customer who lets his nearby position cash-settle without replacing it may have enough capital to support his former spread position but not enough to support his new naked short position. A risk management system needs to figure out how to prevent the customer from dramatically increasing the riskiness of his position by doing nothing.

Next-Day-Settlement When Markets Are Closed. What should happen when the cash market is closed, due to an unanticipated disruption, at the moment when futures contracts are supposed to cash settle? To reduce risks associated with hedging activities, it is useful to postpone cash-settlement to a later date by postponing automatically the execution of market-on-expiration orders to the next time a single-price auction can conveniently be held, such as the next day. The alternative, canceling market-on-expiration orders and liquidating all outstanding futures positions according to some other prices, has the effect of transferring risk in an unintended manner between market participants.

Eyeball-to-eyeball Pit Trading. While market-on-expiration orders are easy to implement when there an electronic single-price auction, eyeball-to-eyeball pit trading

among human beings can make execution of market-on-expiration orders difficult. For example, suppose that the settlement price is calculated as the arithmetic average of the transactions prices during a one-minute time interval at the end of the day, and all expiring positions are cash-settled at this price. Then it may be impossible for traders to guarantee execution of market orders at the settlement price. This suggests that futures markets would function more smoothly and efficiently if human auctions were replaced by an electronic single-price auction.

CONCLUSION

Cash settlement is pervasive in financial markets. Examples of cash settlement include S&P 500 Futures, single- stock futures, VWAP trades, total return swaps, LIBOR itself, LIBOR futures, loans indexed to LIBOR, interest rate swaps, currency swaps, oil contracts, gas contracts, electricity contracts, OTC derivatives with futures as underlying (e.g. NYMEX Electricity), cotton quality differentials, delivery factors (T-Bond futures), gold fixing, gold spot market contracts, copper contracts, spot transactions indexed to closing prices, cash-settled derivatives, forced liquidations after failed margin calls, live cattle futures, death spiral convertible preferred stock agreements, prediction markets, and mutual fund redemptions (late trading).

Regulators, and some finance literature, have argued that cash settlement can enhance liquidity by allowing positions to cash settle rather than be physically delivered. This paper shows that this notion is an illusion because cash-settled futures contracts are economically identical to physically delivered ones, under the assumption that market-on-expiration orders can be place and certain assumptions associated with microstructure fungibility are satisfied. To the extent that these assumptions are not satisfied, there may be lack of equivalence between cash-settled and physically delivered contracts, but the existing finance literature does not relate its claim of increased liquidity to differences in these assumptions.

REFERENCES

Hans R. Dutt and Lawrence E. Harris
Positions Limits for Cash-Settled Derivative Contracts
The Journal of Futures Markets, lV. 25, No. 10, 94–5965 (2005)

FSA Handbook.
Code of Market Conduct. Chapter 1. MAR 1, release 064, April 2007.

Frank J Jones
The Economics of Futures and Options Contracts Based on Cash Settlement
The Journal of Futures Markets (pre-1986); Spring 1982

Kenneth D Garbade and William L Silber
Cash Settlement of Futures Contracts: An Economic Analysis
The Journal of Futures Markets (pre-1986); Winter 1983

Praveen Kumar; Duane J. Seppi
Futures Manipulation with "Cash Settlement"
The Journal of Finance, Vol. 47, No. 4. (Sep., 1992), pp. 1485-1502

Albert S. Kyle
Notes on Cash Settlement
Unpublished, 1989

Craig Pirrong
Manipulation of Cash-Settled Futures Contracts
The Journal of Business, Vol. 74, No. 2. (Apr., 2001), pp. 221-244.

Hans R. Stoll and Robert E. Whaley, "Expiration-Day Effects of the All Ordinaries Share
Price Index Futures: Empirical Evidence and Alternative Settlement Procedures"
Australian Journal of Management, Vol. 22, No. 2, December 1997,

In re Indiana Farm Bureau Coop. Assn.,
[1982-1984 Transfer Binder]
Comm. Fut. L. Rep. (CCH) 7 21,796 (Dec. 17, 1982)