Essays in liquidity

Sébastien Galy

A Thesis

in

The Department of Finance of

The John Molson School of Business

Presented in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy (Ph.D.) at

Concordia University

Montreal, Quebec, Canada

April 2003

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Mr. Sébastien Galy

Entitled: Essays in Liquidity

And submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (Administration (Finance))

Complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

____________________________________ Chair

____________________________________ External Examiner

____________________________________ Examiner

____________________________________ Examiner

____________________________________ Examiner

____________________________________ Thesis Supervisor

Approved by ____________________________________
Chair of Department or Graduate Program Director

________20__ ______________________________
Dean of Faculty

ABSTRACT

Essays in Liquidity

Sébastien Galy, Ph.D.

Concordia University, 2003

Derivatives markets can quickly become illiquid in periods of high uncertainty. Neither the source of this illiquidity nor its implications are well understood. First, this thesis shows that hedgers' trades have an adverse impact on the futures price creating effectively an endogenous transaction cost increasing in times of uncertainty and acting as the source of illiquidity in these markets. Second, illiquidity is shown to strengthen the wealth effect, which has been proven to be too weak empirically to explain the behavior of prices. The wealth effect is the mechanism through which changes in investors' wealth impact their attitude towards risk. As investors lose wealth, they become more risk averse and ask for a higher compensation to hold a risky asset thereby decreasing its price. Third, illiquidity is shown to potentially explain the shape of the implied volatility function not only as a function of moneyness but also of the options' volume or open interest. These results are derived from models, where producers maximize their expected utility derived from their profits. They seek to hedge the uncertain price at which they will sell their product in the future. They can use futures or put options to reduce this price risk but must pay speculators, defined as having no position in the underlying asset, a premium. This premium disappears normally as trades are assumed to be too small to matter and the risk of trading perfectly shared. Both of these assumptions are relaxed to derive the illiquidity transaction costs and its implications.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS

ENDOGENOUS ILLIQUIDTY TRADING COSTS FROM RISK SHARING

~	is used over variables that are uncertain
G	is the number of producers of commodities
N	is the number of speculators
W	is the investor's initial wealth
$\tilde{q}$	is the amount produced by a given producer
$\tilde{p}$	is the price of this good
h	is its certain cost per unit
ξ	is the number of futures contracts sold by the producer to hedge
f	is the futures price
π	is the futures expected risk premium
S	is the amount held of the non-marketable good
R_m	is the return on the non-marketable good
α	is the degree of risk aversion
$\tilde{C}$	is the investor's consumption
$U(\tilde{C})$	is the investor's utility
$\tilde{Q}$	is the aggregate demand for the production good
η_d	is the price elasticity
M	is a constant controlling the importance of the random demand shock $\tilde{\delta}_d$
$\tilde{\delta}_d$	is a random demand shock

ILLIQUIDITY AND THE WEALTH EFFECT

G	is the number of producers of commodities
N	is the number of speculators
π_t	are the profits from production
$V(\pi_{t+j})$	is the investor's utility function for $j = 1,...,\infty$
$U(\pi_t)$	is the investor's instantaneous utility function for $j = 1,...,\infty$
β	is the investor's preference for time
α	is the degree of non-separability of preferences
M_t	is the marginal rate of substitution
σ_t	is the elasticity of the marginal rate of substitution
ε_{F_0,f_0}	is the price elasticity of futures contracts
ρ_t	is the investor's degree of risk aversion
$\rho_{V_t}^A$	is the investor's degree of risk aversion based on his utility $V(\pi_{t+j})$
$\rho_{U_t}^A$	is the investor's degree of risk aversion based on his instantaneous utility
$p_{t,i}$	is the investor's degree of prudence
y_0	is the amount produced at time 0
$p_t(y_0)$	is the price of that good
f_0	is the quantity of futures contracts
F_0	is the price of the futures contract
$c(y_0)$	is the cost of production at time 0

Illiquidity and the Graph of the implied volatility Functions

G is the number of producers of commodities

N is the number of speculators

π_t are the profits from production

$V(\pi_{t+j})$ is the investor's utility function for $j = 1,...,\infty$

ρ_t is the investor's degree of risk aversion

υ_t is the investor's degree of prudence

y_0 is the amount produced at time 0

$p_t(y_0)$ is the price of that good

f_0 is the quantity of futures contracts

F_0 is the price of the futures contract

$c(y_0)$ is the cost of production at time 0

P is the subjective probability measure

Q is the risk neutral probability measure

W is the probability measure used for pricing the futures contract

V is the probability measured used for pricing the put option

s is the state of the world

Ω is the world of all possible events s

n_0 is the number of put options

P_0 is the price of the put options

K is the put option's strike price

t is the maturity of the option

M_t is the marginal rate of substitution

INTRODUCTION

The results of empirical research on futures and options markets suggest strongly that the next step in the research on derivatives pricing is to address the actual microstructures or trading in these markets. In this thesis, I link the microstructure literature with the risk-sharing literature, which addresses the hedgers' needs to share their risk. The microstructure literature shows how the trading mechanism of a given asset impacts its price by creating endogenously an illiquidity trading cost generally in the form of a bid-ask spread. However, this approach, which is based on imbalances of supply and demand, cannot be extended to derivatives that are priced almost exclusively using equilibrium models. More recently, illiquidity trading costs have been added exogenously to equilibrium models to study their impact on the investors' decisions. I derive endogenously an illiquidity trading cost in an equilibrium model by linking the microstructure literature with the risk sharing literature. To achieve this, I introduce a new trading mechanism that accounts for the fact that hedgers enter the derivatives markets to share their risks with other hedgers and speculators. The results that are derived from this potentially explain several empirical puzzles concerning the distribution of futures prices and the shape of the implied volatility as a function of the options' strike price and volume. Detailed explanations of the results achieved in this thesis are explained in what follows.

Derivatives markets can quickly become illiquid in periods of high uncertainty. Neither the source of this illiquidity nor its implications are well understood. First, this thesis shows that hedgers' trades have an adverse impact on the futures price creating effectively an endogenous transaction cost increasing in times of uncertainty and acting as the source of illiquidity in these markets. Second, illiquidity is shown to strengthen the wealth effect, which has been proven to be too weak empirically to explain the behavior of prices. The wealth effect is the mechanism through which changes in investors' wealth impact their attitude towards risk. As investors lose wealth, they become more risk averse and ask for a higher compensation to hold a risky asset thereby decreasing its price. Third, illiquidity is shown to potentially explain the shape of the implied volatility function not only as a function of moneyness but also of the options' volume or open interest.

This thesis shows how, theoretically, illiquidity creates an endogenous transaction cost increasing with the variance of the spot price and the volume of trades in the futures market generated by hedging pressures. High uncertainty represented by high volatility in the spot market drives out liquidity in the futures market the larger the trades. This transaction cost comes from the trader's inability to share risk freely with the rest of the futures market. Even in its absence, futures markets will be illiquid if its price setting mechanism allows for multiple prices. This suggests that a single price mechanism increases liquidity in the futures markets by forcing the sharing of risks, abstracting from traditional trading costs that would create effectively a bid-ask spread. These conclusions are derived assuming that the investor's degree of risk aversion does not increase as they lose wealth. This assumption is relaxed in the second paper.

Investors' attitudes towards risk and the resulting impact on prices in financial markets are determined by changes in their wealth. This wealth effect, however, provides a poor explanation of the mean, skewness and kurtosis of the observed distribution of futures prices especially in the derivatives market for reasonable values of the degree of risk aversion. The second paper shows that illiquidity in the futures market, modeled endogenously as a trading cost, increases the strength of the wealth effect for the same degree of risk aversion. The resulting distribution of futures prices presents a more pronounced left fat tail and left skewness than would have been implied by the wealth effect alone. This model is extended in the third paper to consider the impact of illiquidity on put options.

Illiquidity in the put option's market can potentially jointly explain the empirical puzzles concerning the graph of the implied volatility as a function of moneyness and as a function of the volume or open interest. Illiquidity proves to be closely related to volatility in the underlying spot price for options close to or at the money. It can therefore be partially controlled by hedging the Vega risk of a change in volatility. This could explain why Bollen and Whaley (2002) found that abnormal returns from selling put options disappear when controlling for Vega risk. Put options, as well as futures contracts, are priced using the technique of changes in the subjective probability measure normally reserved for risk neutral pricing. In addition, we explore how changes in the hedging period, size of the cash good being hedged and profits change the demand for put options and futures contracts.

REVIEW OF THE LITERATURE

Illiquidity is studied within the microstructure literature and the asset pricing literature. In this thesis, we try to explain known empirical puzzles in the pricing of derivatives by linking the illiquidity literature with the risk sharing literature in incomplete markets.

1. ILLIQUIDITY IN THE MICROSTRUCTURE AND ASSET PRICING LITERATURE:

The microstructure literature is based on market makers who match the supply and demand of a given security. This creates transaction costs whose properties measure[1] the degree of illiquidity in that market. The asset pricing literature takes a more macroeconomic approach by using models of equilibrium. These equate the supply and demand of a given security and take illiquidity as given instead of generating it from the markets' microstructures.

In the microstructure literature, the market maker provides immediacy in the market by holding stocks in inventory to cover imbalances in the buy and sell orders of traders. He sets bid-ask spreads to cover his costs and risks of trading with investors. Risk is assumed away as the market maker is typically assumed to be risk neutral. Demsetz (1968) introduced order-processing costs, while Garman (1976) posited inventory costs as the main source of costs for the market maker. Glosten and Milgrom (1985) show how trading costs can appear endogenously by introducing trades from potential insiders who may have an information advantage (see O'Hara (1994) for a comprehensive review of the microstructure field).

Insider trading and its accompanying cost, as described by Glosten and Harris (1985), is probably the main reason why single underlying stock futures have had little success, as they

[1] Kyle (1984, 1985) introduced the two traditional properties of illiquidity: tightness measured by the bid-ask spread and depth measured by the ability to make a trade without seeing the bid or ask move against oneself. Other measures exist in practice such as volume, volatility or for example the average daily ratio of absolute stock return to dollar volume used by Amihud (2002).

allow a leveraged bet by an insider. Simons[2] reported in 2001 that only two out of seven exchanges have had any success with single underlying stock futures. In the United States, this potential for manipulation of single stock futures markets led in part to the 1982 Shad Johnson agreement between the CFTC and the SEC preventing the introduction of single underlying futures. Therefore, most active futures markets are based almost exclusively on either a financial index or a commodity and suffer comparatively less from manipulation and insider trading than individual stocks. In addition, manipulation by squeezing the underlying spot market is difficult because of their large size, with some notable exceptions such as the attempted manipulation of the copper market by a trader from Sumitomo during 1995 and 1996. Bid-ask spreads are in practice generally quite small in commodity and index futures markets under normal conditions but are known to increase significantly when these markets come under stress.

Inventory costs are non-existent by definition in futures market as a futures contract is created and reversed without the need to hold it in reserve for delivery. Hence, there is no need to ask for a premium to cover such a cost. However, there may be a cost in transferring risk to a party unwilling to bear it, as in the case considered in this paper, especially as it is a leveraged position. In the Nikkei 225 index futures contracts traded on the SIMEX, Kim, Ko and Noh (2000) find that asymmetric information costs are very small and that inventory costs, the holding of undesired risky positions, account for a very large proportion of the bid-ask spread[3]. Hasbrouk (2002) finds a decomposition of long-run price volatility into trade and non-trade related components. He finds that trades in the pork belly, Euro and British pound futures contracts on the Chicago Mercantile Exchange account for about one half of their volatility.

Theoretically, in addition to the microstructure literature, the asset pricing literature studies illiquidity at a more macroeconomic level by taking illiquidity as exogenous and using

[2] Howard Simons, 24/01/2001, What is the Future for Single Stock Futures, Street.com

[3] Chueh and Yen (2002) find that order processing costs and asymmetric information costs are equally important to explain the bid-ask spread of the Nikkei 225. Neither paper considers all three sources of costs at the same time.

equilibrium models equating supply and demand. For example, Ericsson and Renault (2001) price bonds with liquidity and credit risk assuming that illiquidity is an exogenous random trading cost. This approach is philosophically different from the microstructure literature as illiquidity is not an endogenous result of microstructures and the price is determined by equilibrium of supply and demand or equivalently arbitrage, as markets are implicitly assumed to be complete. Jacoby, Fowler and Gottesman (2000) use this approach to derive a liquidity-adjusted version of the Capital Asset Pricing Model (CAPM). They find results suggesting that the measure of the bid-ask spread has an impact on both expected returns and systematic risk and find some empirical evidence of this in their 2002 paper. Acharya and Pedersen (2002) develop a four beta CAPM with persistent illiquidity and study its impact on the investors' decision-making. This paper shows that illiquidity trading costs can be created endogenously by relaxing the assumption that markets are incomplete and that risk can be shared costlessly.

Empirical and theoretical work on illiquidity in the stock market does not transfer fully to most active futures market with some exceptions. Empirically, under the notion of commonality, Hasbrouck and Seppi (2001) find, for example, that there is empirical evidence that a given market cannot freely absorb trades when most stock markets are heading in the same direction. If market risk is defined as the risk of a price change, then commonality is the interaction between illiquidity and systematic market risk. Speculators who have lost a large part of their marketable wealth in a market downturn may become more unwilling to become the counterparty to a new futures trade without compensation, creating larger transaction costs.

In addition to the methodological difference between the microstructure literature and the asset pricing literature, there is no agreement on a single definition of liquidity. For example, O'Hara (1994) defines liquidity as the ability to trade essentially costlessly, while Ericsson and Renault (2001) define illiquidity as the discount incurred for the immediacy of a trade. This paper adopts the definition of Friend and Blume (1975), which includes that of O'Hara (1994). Illiquidity is the inability of traders to buy or sell any quantity at no cost. Order processing transaction costs, as in Constantinides (1986) and Amihud and Mendelson (1986), are not considered nor are delays in transactions brought about by market closures. The two traditional pillars of microstructures, inventory costs and asymmetries of information, are not considered. Yet illiquidity will be generated as an endogenous trading cost from the inability of traders to share a common risk perfectly. Each trade requires a transfer of risk to another trader for the payment of a transaction cost. The bid-ask spread is therefore generated endogenously as is the case when insider information is introduced.

2. THE RISK SHARING LITERATURE

In this thesis, we study illiquidity in equilibrium, linking the illiquidity literature with the risk sharing literature in incomplete markets. Risk sharing is concerned with the possibility for economic agents to exchange risks through financial markets, using derivatives as risk transfer vehicles. In the asset pricing literature, Dumas (1989) shows that when investors have different degrees of risk aversion and suffer a common risk, they must share the aggregate risk. Wang (1994) considers investors that are heterogeneous both in their investment opportunities and access to information. In the absence of information asymmetries, according to Wang (1994), selling by an investor increases the volume and decreases the price, increasing its expected return, as the asset's expected payoff has not changed. This leads other investors to buy the asset so that its price may remain independent of its volume. Risk sharing becomes of interest only if investors cannot correctly assess the asset's expected payoff. This happens when investors have heterogeneous beliefs as in Detemple and Murphy (1994) or asymmetries of information as in Wang (1994). Risk sharing will change depending on the source of the friction, such as non-insurable labor income shocks in Constantinides and Duffie (1996), and frictions or constraints imposed on the investors, such as the lower bound imposed by Grossman and Zhou (1996) on the hedger's wealth which forces them to hedge so as to respect it. This sharing of risks creates a demand for derivatives to shift risk between those who are willing to take on more of risk for a premium and those who must reduce their risk. Risk sharing comes under the broad denomination of ‘portfolio insurance demand’ in the derivatives literature.

In the derivatives literature, risk sharing among heterogeneous investors creates therefore a role for derivatives as risk transfer vehicles. Grossman and Zhou (1996) studied such an exchange of risk in complete markets and continuous time where one type of hedger is constrained not too lose a given fraction of his initial wealth, creating an asymmetric need to share risk and hence a demand for put options. Franke, Stapleton and Subrahmanyam (1998) show that the degree to which investors face non-hedgeable background risks, such as labor income shocks or shocks to non traded assets, determines the exchange of risks. Finally, Bates (2001) considers the sharing of crash risk or negative stock market jumps and shows that it explains partially why stock options tend to overpredict volatility and the risk of a price jump.

When markets are incomplete, i.e. traders do not have enough uncorrelated assets to hedge all risk sources (Harrison and Pliska (1983) and Duffie and Huang (1985)), the sharing of risks is hampered. In such a context, the supply and demand of financial assets is imperfectly elastic as pointed out for example by Leisen (2002), which implies therefore a transfer of risk when trading. Leisen pointed out that derivatives would not be traded if the price of their underlying asset at the next trading date is assumed locally normal and markets are incomplete. Magill and Quinzii (1995) following Keynes (1930) define market incompleteness as a failure of the market to coordinate activities as all futures contract trades cannot be made at a predetermined price due to frictions. This creates a demand for cash to hedge against price uncertainty in addition to futures. The portfolio insurance demand literature extends this result to show that it creates a demand for options.

3. EMPIRICAL PUZZLES

Risk sharing and market microstructures are only beginning to appear in the field of derivatives. Researchers are increasingly looking if the nature or structure of exchanges can explain known limitations in current pricing models, specifically the graph of the implied volatility as a function of the strike price and volume. The graph of the implied volatility decreases monotonically as a function of the strike price or moneyness. Judd and Leisen (2002) report that fixed maturity plots of a call option's open interest across strike prices peaks for the at-the-money contract. Bollen and Whaley (2002) find that the time variation (term structure) of the implied volatility of an option series is directly is a function of net buying pressure from public order flow using S&P500 index put options and twenty individual stocks. Net buying pressure is defined as the number of contracts traded at or above the prevailing bid/ask midpoint less the number of contracts traded below the prevailing midpoint. Finally, Bates (2001) studies a market where negative jumps (crashes) can occur. Less crash-averse investors insure the more risk-averse through the options market. Bates finds that heterogeneity amplifies the impact of jumps (crashes). Option prices then predict as in the data that option prices overpredict volatility and jump risk due to the wealth effect. Finally, This paper shows that net buying pressure creates illiquidity in the derivatives market in the form of an endogenous illiquidity trading cost that amplifies the wealth effect. This can potentially explain empirical puzzles related to the graph of the implied volatility functions as a function of moneyness, volume or open interest and time structure.

ENDOGENOUS ILLIQUIDITY TRADING COSTS FROM RISK SHARING

ABSTRACT

In general, index and commodity futures markets are considered to be highly liquid. Yet these markets can quickly become illiquid in periods of high uncertainty. So far, there exists no theoretical explanation as to why liquid futures markets can become illiquid in these periods of high uncertainty. This paper shows how, theoretically, illiquidity creates an endogenous transaction cost increasing with the variance of the spot price and the volume of trades in the futures market generated by hedging pressures. High uncertainty represented by high volatility in the spot market drives out liquidity in the futures market the larger the trades. This transaction cost comes from the trader's inability to share risk freely with the rest of the futures market. Even in its absence, futures markets will be illiquid if its price setting mechanism allows for multiple prices. This suggests that a single price mechanism increases liquidity in the futures markets by forcing the sharing of risks, abstracting from traditional trading costs that would create effectively a bid-ask spread.

1. INTRODUCTION

In general, index and commodity futures markets are considered to be highly liquid. Yet these markets can quickly become illiquid in periods of high uncertainty characterized by high volatility in the underlying spot market. Illiquidity is a risk common to financial markets as typified by the 1987 stock market crash. Financial institutions used portfolio insurance programs that automatically sold stocks as their prices were dropping thereby reducing these institutions' exposure to further drops in the stock price. The combined selling pressure is considered one of the reasons for the large amplitude of the crash and forced financial institutions to sell positions far below market values. As this example shows, illiquidity or the inability of traders to buy or sell any quantity at no cost, rises drastically in periods of high uncertainty. While this mechanism is widely recognized in practice, it remains poorly understood in theory.

This paper shows how long traders are forced to pay above the market-clearing price and short traders are forced to accept below the market-clearing price in the futures market, when their trades are not negligible and the market-clearing price corresponds to the case of negligible trades. Trading creates therefore an endogenous cost that increases the more uncertain the spot market is and the larger the trade. Intuitively, each trade is a transfer of risk from an investor unwilling to bear it to another one willing to take it at a cost. Equivalently, a short trade is a shift of the supply curve to the right and a long trade is a shift of the demand curve to the right creating, by their impact on the price, effectively a trading cost.

Typically, these trading costs disappear when deriving the equilibrium price that equates the aggregate supply and demand of futures. The implicit assumptions, that will be relaxed, are that the futures price is independent of the trade's size and that the risks that motivate the trades are perfectly shared. Risk is perfectly shared in the sense that the trading costs compensating the risk transfers, disappear in the aggregation of each investor's trades to find the market-clearing price. This could mean, for example, that for each hedger wanting to reduce his risk there is a speculator looking to increase his, so that in the aggregate there is no trading cost arising from risk sharing. This result of illiquidity arising from risk sharing is confirmed by showing that even small trades can be illiquid if the futures market does not force risk sharing by imposing a single trading price.

These trading costs are derived from the traders' inability to share risk freely irrespective of the trade's size when markets are incomplete. Harrison and Pliska (1983) and Duffie and Huang (1985) defined markets as incomplete if traders do not have enough uncorrelated assets to hedge all risk sources. This forces investors to share risk when trading, as the risk free replication of the traded asset is no longer possible. Equivalently, the supply and demand for the traded asset is no longer perfectly elastic as pointed out by Scholes (1972) and therefore each trade or shift in the supply and demand curves implies the previously discussed payment of a risk premium or trading cost that would not be present if they were perfectly elastic. The concept of illiquidity is therefore equivalent to the inelasticity of supply and demand. The notion of market incompleteness is compounded by the presence of a non-marketable good that cannot be replicated and a finite supply of speculators willing to absorb any hedging pressure in the futures market. Thus, this trading cost is a compensation for risk shifting when markets are incomplete.

To show how this endogenous trading cost can arise, I assume that the size of the trade depends on the market price in the investor's optimization problem. This has some similarities with assuming that the price of a good depends on the quantity produced in a manufacturer's optimization problem, as would be the case for oligopolies. As is well known, the profits generated by their pricing power attract other manufacturers until the profits disappear in the limit and the market becomes one of perfect competition. This supposes of course that manufacturers do not interact strategically.

The fact that one party must initiate an action before the other one reacts is introduced by assuming that hedgers are the ones initiating the trades and that speculators enter the futures market attracted by the premiums or transaction costs offered to shift risk to them. Hence, the assumption of perfect risk sharing is relaxed, as it is now costly. In the futures market, the number of traders is limited and each trader acts as a manufacturer so that imperfect competition translates into an inelastic demand as well as inelastic supply. Instead of producing goods, they produce 'bads' as they offer risks that other traders will not accept without compensation. Traders therefore have the opposite of pricing power, a production or trading cost.

Illiquidity appears as an endogenous trading cost compensating for the sharing of risks measured here by the spot market volatility. It is different from the trading cost generated by the microstructure literature on insider trading. In this literature, the market maker covers himself against the risk of insider trading by offering a bid-ask spread. An investor must therefore buy and sell his asset from the market maker at a different price creating effectively a trading cost.

A typical critique of illiquidity modeled as a trading cost is that it is in essence nothing more than a trading cost model. Adding an extra source of risk or trading cost linearly to the derivatives' expected return is not considered a major improvement. However, in this model, the trading cost is derived endogenously from a definition of liquidity. The illiquidity trading cost does appear as an extra source of risk or trading cost but adds richness to the futures expected return by introducing the expected spot volatility and the futures' volume. Finally, a critique is that illiquidity as defined in this model is market risk[4]. Indeed illiquidity is an overlooked facet of market risk. Shifts in the price created by changes in the supply and demand curves are introduced directly in the price setting mechanism, in addition to using the fact that hedgers must initiate trades first before speculators can enter to take advantage of the premium or trading cost offered.

Confirming that illiquidity can be generated by the inability to share risk freely, we find that futures markets can be illiquid without assuming that the futures price depends on the

[4] Market risk is defined as the risk of losses due to movements in market prices. It represents therefore the risk associated with movements in the supply and demand of the asset considered.

quantity traded. It suffices that the market clearing mechanism allows for multiple prices. Each futures trade has then potentially a different price dependent on its size and the volatility in the spot market. This suggest that a market clearing mechanism, generating a single futures price, makes the market more liquid by forcing the investors to share risks. The conclusion abstracts, of course, from traditional trading costs that would otherwise generate a bid-ask spread and therefore multiple prices.

2. PREVIOUS WORK ON LIQUIDITY

Illiquidity is studied within the microstructure literature and the asset pricing literature. The microstructure literature is based on market makers who match the supply and demand of a given security. This creates transaction costs whose properties measure[5] the degree of illiquidity in that market. The asset pricing literature takes a more macroeconomic approach by using models of equilibrium. These equate the supply and demand of a given security and take illiquidity as given instead of generating it from the markets' microstructures.

In the microstructure literature, the market maker provides immediacy in the market by holding stocks in inventory to cover imbalances in the buy and sell orders of traders. He sets bid-ask spreads to cover his costs and risks of trading with investors. Risk is assumed away as the market maker is typically supposed to be risk neutral. Demsetz (1968) introduced order-processing costs, while Garman (1976) posited inventory costs as the main source of costs for the market maker. Glosten and Milgrom (1985) show how trading costs can appear endogenously by

[5] Kyle (1984, 1985) introduced the two traditional properties of illiquidity: tightness measured by the bid-ask spread and depth measured by the ability to make a trade without seeing the bid or ask move against oneself. Other measures exist in practice such as volume, volatility or for example the average daily ratio of absolute stock return to dollar volume used by Amihud (2002).

introducing trades from potential insiders who may have an information advantage (see O'Hara (1994) for a comprehensive review of the microstructure field).

Insider trading and its accompanying cost, as described by Glosten and Harris (1985), is probably the main reason why single underlying stock futures have had little success, as they allow a leveraged bet by an insider. Simons[6] reported in 2001 that only two out of seven exchanges have had any success with single underlying stock futures. In the United States, this potential for manipulation of single stock futures markets led in part to the 1982 Shad Johnson agreement between the CFTC and the SEC preventing the introduction of single underlying futures. Therefore, most active futures markets are based almost exclusively on either a financial index or a commodity and suffer comparatively less from manipulation and insider trading than individual stocks. In addition, manipulation by squeezing the underlying spot market is difficult because of their large size, with some notable exceptions such as the attempted manipulation of the copper market by a trader from Sumitomo during 1995 and 1996. Bid-ask spreads are in practice generally quite small in commodity and index futures markets under normal conditions but are known to increase significantly when these markets come under stress.

Inventory costs are non-existent by definition in futures market as a futures contract is created and reversed without the need to hold it in reserve for delivery. Hence, there is no need to ask for a premium to cover such a cost. However, there may be a cost in transferring risk to a party unwilling to bear it, as in the case considered in this paper, especially as it is a leveraged position. In the Nikkei 225 index futures contracts traded on the SIMEX, Kim, Ko and Noh (2000) find that asymmetric information costs are very small and that inventory costs, the holding of undesired risky positions, account for a very large proportion of the bid-ask spread[7]. Hasbrouk (2002) finds a decomposition of long-run price volatility into trade and non-trade related

[6] Howard Simons, 24/01/2001, What is the Future for Single Stock Futures, Street.com

[7] Chueh and Yen (2002) find that order processing costs and asymmetric information costs are equally important to explain the bid-ask spread of the Nikkei 225. Neither paper considers all three sources of costs at the same time.

components. He finds that trades in the pork belly, Euro and British pound futures contracts on the Chicago Mercantile Exchange account for about one half of their volatility.

Theoretically, in addition to the microstructure literature, the asset pricing literature studies illiquidity at a more macroeconomic level by taking illiquidity as exogenous and using equilibrium models equating supply and demand. For example, Ericsson and Renault (2001) price bonds with liquidity and credit risk assuming that illiquidity is an exogenous random trading cost. This approach is philosophically different from the microstructure literature as illiquidity is not an endogenous result of microstructures and the price is determined by equilibrium of supply and demand or equivalently arbitrage, as markets are implicitly assumed to be complete. Jacoby, Fowler and Gottesman (2000) use this approach to derive a liquidity-adjusted version of the Capital Asset Pricing Model (CAPM). They find results suggesting that the measure of the bid-ask spread has an impact on both expected returns and systematic risk and find some empirical evidence of this in their 2002 paper. Acharya and Pedersen (2002) develop a four beta CAPM with persistent illiquidity and study its impact on the investors' decision-making. This paper shows that illiquidity trading costs can be created endogenously by relaxing the assumption that markets are incomplete and that risk can be shared costlessly.

Empirical and theoretical work on illiquidity in the stock market does not transfer fully to most active futures market with some exceptions. Empirically, under the notion of commonality, Hasbrouck and Seppi (2001) find, for example, that there is empirical evidence that a given market cannot freely absorb trades when most stock markets are heading in the same direction. If market risk is defined as the risk of a price change, then commonality is the interaction between illiquidity and systematic market risk. Speculators who have lost a large part of their marketable wealth in a market downturn may become more unwilling to become the counterparty to a new futures trade without compensation, creating larger transaction costs.

In addition to the methodological difference between the microstructure literature and the asset pricing literature, there is no agreement on a single definition of liquidity. For example, O’Hara (1994) defines liquidity as the ability to trade essentially costlessly, while Ericsson and Renault (2001) define illiquidity as the discount incurred for the immediacy of a trade. This paper adopts the definition of Friend and Blume (1975), which includes that of O’Hara (1994). Illiquidity is the inability of traders to buy or sell any quantity at no cost. Order processing transaction costs, as in Constantinides (1986) and Amihud and Mendelson (1986), are not considered nor are delays in transactions brought about by market closures. The two traditional pillars of microstructures, inventory costs and asymmetries of information, are not considered. Yet illiquidity will be generated as an endogenous trading cost from the inability of traders to share a common risk perfectly. Each trade requires a transfer of risk to another trader for the payment of a transaction cost. The bid-ask spread is therefore generated endogenously as is the case when insider information is introduced.

The paper is organized as follows. In section 3, the theoretical model is presented and the trading costs are derived by relaxing the assumption that the quantity traded is independent of its price. In section 4, the equilibrium demand and supplies are aggregated across investors. Depending on the aggregation assumptions, liquidity is a property of equilibrium or disequilibrium.

3. THE DEMAND FOR FUTURES CONTRACTS AND ILLIQUIDITY TRADING COST

I assume a one period model where at the beginning of the period, uncertain end of period variables are noted by ~. There are three groups of utility maximizing agents, G producers of commodities and N speculators, both with utility functions that exhibit Constant Absolute Risk Aversion (CARA), as well as M risk averse consumers whose preferences remain unspecified.

The $G+N$, $i=1,...,G,...,G+N$, producers and speculators may have different degrees of risk aversion α_i. Uncertainty exists at the beginning of the period and disappears at the end.

The agent consumes $\widetilde{C}$ of the single type of non-durable good produced in the economy. He starts with an initial wealth W and generates revenue from producing $\widetilde{q}$ of the non-durable good sold at the price $\widetilde{p}$ for the certain cost per unit h. As both price and quantity are uncertain the agent sells ξ futures contracts to hedge against the uncertainty in his expected revenues of production. The futures contracts delivery period coincides with the one period needed to produce and then sell the good. The producer agrees today to sell in one period a production good for the fixed price f. He is therefore selling in advance his production to hedge himself. It is assumed that agents hold S of the non-marketable commodity R_m making markets incomplete. Markets are also incomplete if there are more sources of risk than assets to hedge them with. The agent is said to be a hedger if he is actually producing the good and a speculator otherwise ($\widetilde{q}=0$). Speculators have a fixed cost t of entering into the futures market so that it can be reasonably assumed that speculators do not all enter into the futures market as it is too expensive for some to enter. The agent is assumed to leave no wealth at the end of the period (no bequest). He therefore consumes all his remaining initial wealth with his revenue from production and his holdings of the non-marketable good.

$$\widetilde{C}=\begin{cases} W+(\widetilde{p}-h)\widetilde{q}+(f-\widetilde{p})\xi+S\widetilde{R}_M & \textit{For a Hedger} \\ W-t+(f-\widetilde{p})\xi+S\widetilde{R}_M & \textit{For a Speculator} \end{cases} \tag{1}$$

A classic requirement is that the hedger be at least as risk averse as the speculators. The hedger wants to reduce his exposure to price changes in his production good by using futures because he is risk averse and has a large price risk to hedge. However, the speculator (a hedge fund for example) has no such risk and would enter into such a contract only if he gained from it.

Each investor maximizes the expected utility of his consumption $\widetilde{C}$. They prefer higher returns and dislike increasing the variance of their consumption. The relative importance of return versus variance is controlled by the parameter α, which controls the degree of risk aversion.

$$U = E(\widetilde{C}) - \frac{\alpha}{2} Var(\widetilde{C}) \qquad (2)$$

The agents make decisions at the beginning of the period on their consumption, quantity of futures and of non-marketable goods. These decisions are rational as defined by Muth (1961) and Lucas (1973, 1976). Agents use all available information to make their decisions and perceive the true probability of events.

3.1 THE CONSUMPTION-PRODUCTION GOOD MARKET

Deriving the equilibrium in the spot or consumption good market gives results of little interest, so that in practice the dynamics of the spot market is assumed as in Tien (2002). This approach is more appropriate for liquid financial spot goods. We follow Hirshleifer (1988, 1989) in assuming a demand curve for the physical commodity introducing the demand shock as a single source of risk. The aggregate demand $\widetilde{Q}$ for the production good is imposed exogenously and is an indirect representation of the consumer's preferences. The consumers in the aggregate are assumed to pay less when more of the consumption good is produced. Mathematically, the log of the consumers' demands when aggregated is assumed downward sloping as a function of the price of the production-consumption good $\widetilde{p}$.

$$\ln(\widetilde{Q}) = \ln(M\widetilde{\delta}_d) + \eta_d \ln(\widetilde{p}) \qquad (3)$$

The negative slope $\eta_d < 0$ is the price elasticity and the intercept $\ln(M\tilde{\delta}_d)$ is the overall level of aggregate demand. M is a constant controlling the importance of the random demand shock $\tilde{\delta}_d$. The aggregate demand is therefore uncertain at the beginning of the period.

This market is assumed to be competitive, so that when maximizing the utility (2) of the producer over his production, given his wealth constraint given in (1), the usual condition of pricing the good at the marginal cost is found. In this case, it is the expected spot good price that equals a fixed marginal cost h.

$$E\tilde{p} = h \tag{4}$$

In equilibrium the aggregate supply from the G producers each selling their output $\tilde{q}$ equals the aggregate demand for that good (3), so that

$$\ln(G\tilde{q}) = \ln(M\tilde{\delta}_s) + \eta_s \ln(\tilde{p})$$

$$G\tilde{q} = M\tilde{\delta}(\tilde{p})^{\eta} \tag{5}$$

This equation can be rearranged to have the price of the consumption good as a function of its supply in equilibrium.

$$\tilde{p} = \left(\frac{G\tilde{q}}{M\tilde{\delta}}\right)^{\frac{1}{\eta}} = k\left(\frac{\tilde{q}}{\tilde{\delta}}\right)^{\frac{1}{\eta}}, \quad k > 0 \tag{6}$$

The price of the consumption good is a decreasing function of the production and an increasing function of the demand shock. There are therefore two sources of risk in the model, the demand shock and the return on the non-tradable asset.

3.2 THE FUTURES MARKET

The agent maximizes his utility (2) under a budget constraint (1) by deciding on his holdings of the non-marketable good S and his holding of futures contracts ξ. The equations for the optimal allocation of the non-marketable good (7) and the futures contract (8) for each agent are given by:

$$E(\widetilde{R}_M) = \alpha(SVar(\widetilde{R}_M) + Cov((\widetilde{p} - h)\widetilde{q} + \xi\widetilde{\pi}, \widetilde{R}_M)) \tag{7}$$

$$\pi = \alpha_i \left(\xi_i Var(\widetilde{\pi}) + Cov(\widetilde{\pi}, (\widetilde{p} - h)\widetilde{q}) + S_i Cov(\widetilde{\pi}, \widetilde{R}_M) \right) \tag{8}$$

Where, $\pi = E\widetilde{\pi} = E(\widetilde{p} - f)$ is the expected premium asked to hold the futures contract rather than the spot or consumption-production good. As the futures contract is of only one period and the spot good is assumed not traded by others until it is sold, the premium does not take into account changes in the prices of the spot and futures markets in the meantime.

3.3 HEDGING DEMAND

The commodity producers sell futures to hedge their production by locking in, at the beginning of the period, the price at which they will sell the good at the end of that period. At the end of the period, the price of one unit of good hedged "naively" with one futures contract would be $f - p + p = f$, if the end of the period coincides with the delivery date of the futures contract. In a mean-variance setting, the commodity producer still hedges by going short but with a different hedge ratio of futures contracts to commodity goods as his purpose is not only to minimize his risk as measured by variance, but also maximize his returns as measured by the expectation. Rewriting equation (8), denoting each agent by the index i, we can isolate the number of futures contracts ξ_i.

$$\xi_i = \frac{\frac{\pi}{\alpha_i} - Cov((\widetilde{p} - h)\widetilde{q} + S_i\widetilde{R}_M, \widetilde{\pi})}{Var(\widetilde{\pi})} \tag{9}$$

As the speculator is defined as having no position in the underlying commodity $\widetilde{q} = 0$ and risk aversion $\alpha_i > 0$ that may be smaller than that of hedgers, it follows from equation (9) that he will be long in futures contracts if he expects a premium for doing so and he has no holdings of the non traded good. Now that the individual demand and supply for futures contracts have been derived, these must be aggregated across investors to find the volume and price in equilibrium. Note that the supply or demand of futures is inelastic in this incomplete market economy. First, however I will show that relaxing the assumption that a trade in the futures market has no impact on the futures price creates an illiquidity trading cost.

3.4 ILLIQUIDITY AS AN ENDOGENOUS TRADING COST

Until now, we have assumed implicitly in the optimization that the size of the futures trade was of a negligible size so that the price of the futures contract f was unaffected by the size of the trade. Now that we relax this assumption, the futures price $f(\xi)$ will now behave differently in the investor's optimization problem. The optimization technique makes an incremental change in the investor's futures contracts position such that there is no positive incremental change to his utility. This differentiation technique is the mathematical representation of the following trading strategy: the investor enters the futures market and trades futures contracts such that he maximizes his utility.

As the futures price depends on the number of futures contracts the investor will trade, the value of a portfolio of futures contracts $\xi(f(\xi)-\tilde{p})$ when traded or equivalently optimized, given in equation (10), will be equal to the futures premium $f(\xi)-\tilde{p}$, the difference between the futures price of a spot sale and its value at maturity when the position is closed out plus an additional term. Using the derivation rules, the optimization or equivalently trading of the futures portfolio is:

$$\frac{\partial[\xi_i(f(\xi_i)-\tilde{p})]}{\partial\xi_i} = f(\xi_i)-\tilde{p}+\left(\frac{\partial(f(\xi_i)-\tilde{p})}{\partial\xi_i}\right)\xi_i = \tilde{\pi}+\left(\frac{\partial f(\xi_i)}{\partial\xi_i}\right)\xi_i \qquad (10)$$

The term on the right in equation (10) was not present before as we had assumed that the futures price was independent of the futures position. Note that the term on the right $\left(\partial f(\xi_i)/\partial\xi\right)\xi_i$ is further simplified as the spot price at the delivery date is expected and assumed not to provoke a natural price squeeze[8]. In other words, the spot market is assumed liquid. It is straightforward to verify that the futures premium under this relaxed assumption will equal the previous in equation (8) plus this additional term in expectation:

$$\pi + \left(\partial f(\xi_i)/\partial\xi_i\right)\xi_i = \alpha_i\left(\xi_i Var(\tilde{\pi}) + Cov(\tilde{\pi},(\tilde{p}-h)\tilde{q}) + S_i Cov(\tilde{\pi},\tilde{R}_M)\right) \qquad (11)$$

The additional term that appears on the right $TC_i = (\partial f(\xi_i)/\partial\xi_i)\xi_i$ is a trading cost. It equals the change in the futures price brought about by trading multiplied by the quantity of futures traded. Graphically, selling (buying) is a shift to the right of the supply (demand) curve creating a downward (upward) pressure on the price. A seller (buyer) will therefore get (pay) a premium below (above) the market price at which supply equals demand when trades are assumed as before in equation (8) to be too small to matter. This trading cost is endogenous as it depends on the investor's preferences. Note that if markets were complete, there would be no such illiquidity trading cost, as the slope would be equal to zero.

In the derivatives and asset pricing literature, this trading cost, when considered, is imposed exogenously (Jacoby, Fowler and Gottesman (2000) or Acharya and Pedersen (2002)). In this paper, the trading cost is created endogenously in the futures market. The expected trading cost of delivering the spot good is ignored so as not to introduce the problems of illiquidity in the spot market. However, this trading cost better known as a natural squeeze could be derived from equation (6).

This approach is different from the microstructure literature as in Glosten and Milgrom (1985). Trading does not involve explicitly a market maker who provides a service of immediacy at a cost covered by the bid-ask spread at which traders can buy and sell. It is a model of

[8] Proof 5 considers the delivery risk of an illiquid spot market at delivery.

equilibrium rather than one of matching buy and sell orders. The cost of trading considered here does not come from insider trading, inventory costs or order processing costs, but from the inability to share risk costlessly. In addition, microstructure models assume that the market maker is risk neutral so that he would absorb the transfers of risk without cost.

As the slope in equation (10) of the trading cost is not known, it is derived from equation (8). In that equation, the size of the trade was assumed implicitly to be too small to matter. We use the fact that a large trade is an aggregation of smaller insignificant ones. The producer's trading cost (TC_i) equals therefore his risk aversion multiplied by the volatility of the spot and the size of the trade.

$$TC_i = (\partial f(\xi_i) / \partial \xi_i)\xi_i = \alpha_i Var(\widetilde{p})\xi_i \tag{12}$$

See proof 1 in the appendix for more details

Intuitively, volatility represents a measure of risk, which is shared when trading. Risk aversion measures the sensitivity to that risk while the trade size gives the trader's overall exposure to that risk. As the spot market becomes more volatile, there is more risk being passed around than investors willing to take it cheaply depending on the trade size and risk aversion of the investor.

The illiquidity trading cost is associated with the variance of the spot price. Illiquidity is therefore linked to variance, something that had to be assumed previously as in Bangia, Diebold, Schuerman and Stroughair (1998). Illiquidity is also associated with volume when all trades are aggregated, a result common to the microstructure literature (i.e. Roll (1984)). Finally, Ericsson and Renault's (2001) assumption that illiquidity-trading costs are dependent on the number of traders in the market is verified here.

4. AGGREGATION AND MARKET CLEARING IN THE FUTURES MARKET

Market clearing mechanisms have been extensively studied especially in the stock market (see O'Hara (1994)). Specialist or market makers provide immediacy but at the cost of a bid-ask spread that will depend heavily on insider trades and inventory costs. In the case of futures markets, most major markets are based on indices or physical commodities leaving little leeway for insider trading and inventory costs are non-existent by definition. There are however fees to cover the access to electronic platforms, such as GLOBEX in the Chicago Mercantile Exchange, and minimal delays in executing trades physically on the floor.

The aggregation mechanism that is proposed here is based on a variant of the Walrasian auction that clears supply and demand in much of economics. The Walrasian auctioneer aggregates the demand and supply of futures to find a market-clearing price and adjusts the price until supply equals demand. This mechanism is the one used implicitly when equating any supply or demand curve to find the equilibrium or market-clearing price in that market and is therefore the foundation of equilibrium economics. The fact this mechanism does not exist explicitly in practice has not limited its dominance outside the microstructure literature.

In this model, the auctioneer starts with the clearing market price of a Walrasian auction that assumes implicitly that all trades are negligible. He has now available a schedule of the supply and demand linking the price each would like to pay for a given quantity of futures. A trade, assumed to be non-negligible, shifts the supply (demand) curve. Equivalently, a trade moves the market in price and quantity. The auctioneer then must find those willing to take that trade on the demand (supply) curve at a discount (premium). Hence, risk shifting is introduced in the market clearing mechanism, where volatility is the risk being shifted from one party to a futures contract to another. This variant of the Walrasian auction recognizes therefore that one party must initiate a trade and offer a compensation for others to accept the trade. Shifts in the supply or demand curve are therefore introduced directly in the price setting mechanism. As the Walrasian auction is just the equilibrium of supply and demand, this variant of the Walrasian auction is the equilibrium of supply and demand after the seller or buyer has initiated a trade, shifting the supply curve when he sells and the demand curve when he buys.

We aggregate the investors' demands and supplies of futures contracts. In equilibrium, each contract has two parties, so that the aggregate net supply or demand must be equal to zero and the market is said to clear. From this, we derive the price of futures in this classical Walrasian auction equilibrium. The first two sections study a market where trades are assumed too small to have an impact on the price before relaxing this assumption in the last section. We will see that the type of aggregation will have a very important impact on risk sharing and liquidity. First, the usual aggregation with a single futures clearing price is derived and illiquidity disappears in equilibrium. Then, the assumption of a single price is relaxed to see that risk sharing is forced in the usual aggregation and that otherwise illiquidity exists when markets are in equilibrium. This shows that the market mechanism chosen to match supply and demand in the futures market has a direct impact on the liquidity in that market.

4.1 AGGREGATION AND CLASSICAL MARKET CLEARING EQUILIBRIUM

The individual hedging demand and supply were defined in equation (9). These must be aggregated to find the equilibrium price of the futures contract. In equilibrium, the demand and supply of futures contracts must balance out. Summing them up for the $\hat{N}$ positions of speculators (hold no position in the underlying asset) and G hedgers,

$$\sum_{i=1}^{\hat{N}+G}\xi_i = 0 \tag{13}$$

Noting that the first of the i investors are hedgers and the others speculators and assuming that there is a unique futures price such that the futures market is cleared, then the premium can be separated into two components.

$$\pi = Cov(\widetilde{p},(\widetilde{p}-h)\widetilde{q})\sum_{i}^{G}\alpha_i + Cov(\widetilde{p},\widetilde{R}_M)\sum_{i}^{\hat{N}+G}\alpha_i S_i \qquad (14)$$

This equation is very similar to the Capital Asset Pricing Model of Hirshleifer (1989) and Mayers (1973, 1976). The futures premium is made up of a non-systematic risk that depends on the market return ($\widetilde{R}_m$) and hedged production $((\widetilde{p}-h)\widetilde{q})$. Clearly the futures premium is independent of the futures trades so that illiquidity is not present in this model. Now that we have shown the classical Walrasian auction equilibrium as a benchmark, let us consider what happens when a single futures price is not imposed to clear the market.

4.2 AGGREGATION IN ALTERNATIVE MARKET CLEARING EQUILIBRIUM

The equilibrium was derived by implicitly assuming that the futures price was unique. The equilibrium or clearing price is therefore forced in the aggregation process. This fact is illustrated in the following example.

Assume that the equilibrium in the futures market is determined by the exchange of futures premiums[9] between investors. Summing the demands in equation (9) would not be helpful as the futures' premium and risk aversion cannot be separated. Hence, I sum each investor's premium in equation (8) and impose there the market clearing condition of equation (13).

Consider a market with two hedgers $(i = 1,2)$ one short and one long in the commodity, but otherwise identical except for their risk aversion. Then imposing the market clearing

[9] the difference between the price at which they will buy and sell the spot good

condition of equation (13) on the premium equation (8), the aggregate premium is different from zero and depends on the volume of trade:

$$f_1 - f_2 = (\alpha_1 - \alpha_2)\xi_1 Var(\widetilde{p}) + (\alpha_2 - \alpha_1)\Big(Cov(\widetilde{p},(\widetilde{p} - h)\widetilde{q}) + S_1 Cov(\widetilde{p},\widetilde{R}_M)\Big) \qquad (15)$$

See proof 2 in the appendix for more details

As can be seen from equation (15), there is no such thing as a single futures price in equilibrium $f_1 - f_2 \neq 0$ if the producers differ only on their risk aversion $\alpha_1 \neq \alpha_2$ when sharing a common commodity price risk. Rather there are several prices as if a market maker was present. Each futures price is determined by the size of the trade or potential trade.

Now assume that both hedgers have the same risk aversion, then the total premium equals zero. Hence, in that case all risk is transferred via the premium and there is a unique price for the futures contract. More generally, assuming different degrees of risk aversion and this type of equilibrium shows how agents transfer risk to others. This approach provides an interesting way to introduce illiquidity in equilibrium as arising from the limitation in the ability to transfer risk by trading. Unsurprisingly, the risk transfer takes a very similar form to the transaction cost derived in equation (12). Let us generalize our argument.

By aggregating among consumers without assuming that the futures' price is unique and independent of the aggregation, we obtain a liquidity premium since it depends on the quantity of futures traded in equilibrium. Let us aggregate the premiums of the long and short traders of equation (8) in this alternative equilibrium, impose the market clearing condition (13) and assume that the variance and covariances of the premium are the same for all agents, then the aggregate premium transfer is:

$$\begin{aligned}&\sum_{i=1}^{N_1}(f_i - \widetilde{p}) - \sum_{j=1}^{N_2}(\widetilde{p} - f_i) = Var(\widetilde{p})\left[\sum_{i=1}^{N_1}\alpha_i\xi_i - \sum_{j=1}^{N_2}\alpha_j\left(\sum_{i=1}^{N_1}\xi_i + \sum_{k=1,k\neq j}^{N_2}\xi_k\right)\right] + \\ &\left[\sum_{j=1}^{N_1}\alpha_j S_j - \sum_{i=1}^{N_2}\alpha_i S_i\right]Cov(\widetilde{p},\widetilde{R}_M) - \sum_{i=1}^{G}\alpha_i\big(Cov(\widetilde{p},(\widetilde{p} - h)\widetilde{q})\big)\end{aligned} \qquad (16)$$

See proof 3 in the appendix for more details

Where the index i represents the N_1 short traders, j and k the N_2 long traders. Illiquidity disappears in equilibrium if investors have the same degree of risk aversion (see proof 3). Irrespective of this, there will be no single futures price clearing the market as there is a risk exposure to the spot market and non-marketable good that remains unshared. This equilibrium could be characterized as one of imperfect in the market for risk sharing.

In practice, this means that choosing a single price market forces the sharing of risk. A futures market would naturally evolve to one of market makers otherwise who can offer different quantities at different prices creating liquidity costs without the added benefit of immediacy offered in the stock market. It also means that the larger the amount of futures contracts one wants to sell, the more costly it can be. Now that we have shown that illiquidity can exist in equilibrium even for atomistic[10] trades, let us consider a variant of the classical Walrasian auction where trades are no longer assumed atomistic.

4.3 AGGREGATION AND MARKET CLEARING EQUILIRBIUM WITH TRADING COSTS

Let us consider the impact of illiquidity transactions costs in the aggregate. The market clearing mechanism is a variant of the Walrasian auction. First the auctioneer observes the prices at which each would be willing to buy or sell and finds a market-clearing price assuming that trades are atomistic. When a trade of non-negligible size is made, the supply (or demand) curve shifts. The auctioneer finds the price that will clear the market given this shift. Using this mechanism, we derived the illiquidity trading cost in equation (12) for a single investor.

When trades are large, the futures premium is no longer defined by equation (12). The investor's optimization problem introduces a trading cost (see equation (10)) for the producers.

$$\tilde{\pi}_i + I_i\alpha_i Var(\tilde{\pi}_i)\xi_i = \alpha_i\left(\xi_i Var(\tilde{\pi}_i) + Cov(\tilde{\pi}_i,(\tilde{p}-h)\tilde{q}) + S_i Cov(\tilde{\pi}_i,\tilde{R}_M)\right) \qquad (17)$$

[10] A trade is said to be atomistic if it is too small to have any impact on the market.

$I_i = 1$ if i is a producer (hedger) and 0 otherwise (speculator)

If the trade was of negligible size, the trading cost would disappear and the equation would revert to equation (12) where trades were small. From equation (17), the demand for futures is therefore:

$$(1 - I_i)Var(\tilde{\pi}_i)\xi_i = \frac{\pi_i}{\alpha_i} - Cov((\tilde{p} - h)\tilde{q} + S_i\tilde{R}_M, \tilde{\pi}_i) \tag{18}$$

To find the equilibrium, we aggregate the demand across investors and impose the market clearing condition of a single price. As only the producers are initiating trades by selling futures, they shift risk, measured as a function of the spot variance, to the other investors. For these investors to accept these sales, they must pay a discount in the form of a trading cost.

$$\pi = \frac{1}{k}\left[Var(\tilde{p})\sum_{i=1}^{G}(-\xi_i) + \sum_{i=1}^{G} Cov((\tilde{p} - h)\tilde{q}, \tilde{p}) + \sum_{i=1}^{\hat{N}+G} S_i Cov(\tilde{p}, \tilde{R}_M) \right] \tag{19}$$

$k = \left(\sum_{i=1}^{\hat{N}+G} \frac{1}{\alpha_i} \right)$ is the average tolerance in the futures market

See proof 4 in the appendix for more details

The futures contract's risk premium is a function of the covariance of the spot price respectively with the revenue generated by sales and the position in the non-marketable asset as was the case without the illiquidity transaction cost in equation (14). The illiquidity transaction cost TC, defined for a single investor in equation (12), creates a risk premium (20) measured by the volatility of the futures premium and the number of trades that hedgers use to shift risk to speculators.

$$TC = Var(\tilde{p})\sum_{i=1}^{G}|\xi_i| \tag{20}$$

The illiquidity premium under these preferences is therefore the volume of futures multiplied by the variance of the spot market. In practice, only the volume of speculator's trades should be considered, as long hedgers will have opposite liquidity trading costs. Hence, it is the volume of trades generated by hedging pressures (trades for which there is no hedger as a counterparty) that generates in practice illiquidity trading costs. The illiquidity trading cost is therefore defined by equation (21).

$$TC = Var(\widetilde{p}) \times \text{Future's volume of trades by hedging pressure} \tag{21}$$

Illiquidity changes the way risk is shared by the market, as risk premiums are now a nonlinear function of risk aversion (k), while it was in equation (14) a linear one. Risk is now equally deflated by the average measure of risk tolerance in the market. The less risk tolerant investors are, the bigger the futures' premium will be.

The illiquidity premium deflates even more the futures' premium, as hedgers must offer a discount to speculators to accept the shift in risk embedded in the trade. Speculators act therefore implicitly as market makers by buying futures contracts from the hedgers in exchange for a transaction cost. These transaction costs create in effect an ask price. If long hedgers were present we would therefore have a bid-ask spread.

Risk transfers are therefore more complex when taking illiquidity into account. Illiquidity increases the futures' premium and makes it depend on the trade size and the expected volatility of the spot market. The dynamics of the futures return are therefore richer with illiquidity. The natural question is how could we manage this risk. While this is not the point of this paper, there are ways to at least reduce that risk ex-ante. One might in practice cut a large futures trade into a series of smaller ones. One could also use options on the underlying spot good to cover against the expected adverse movement in the futures price due to illiquidity in the underlying spot market at delivery.

Using CARA preferences in a one period model has the advantage of showing clearly how illiquidity creates an endogenous transaction cost without order processing and inventory costs or asymmetries of information. This comes at a cost. Under these preferences, risk aversion remains constant irrespective of the investors' wealth. In a more general utility setting, one would anticipate in a crash, a greater unwillingness from speculators to bear risk as they lose much of their wealth, with illiquidity compounding the wealth effect. Therefore, each trade should become more costly as speculators become less willing to accept risk shifting from hedgers. In addition, using a one period model does not allow for the risk in reversing a given position.

CONCLUSION

Illiquidity is introduced in the optimisation or trading problem of the investor as an inability to trade and share risk without changing the market price. This creates an endogenous trading cost and bid-ask spread without the need for informational asymmetries, inventory or order processing costs. This trading cost is a linear function of the variance of the spot market multiplied by the volume generated by hedging pressures, making the futures price dynamics richer. Illiquidity will exist in equilibrium even with very small trades if there is no price mechanism to force market clearing at a single futures price confirming that risk sharing is a source of illiquidity.

ANNEX

Proof 1

I show that the slope of the demand curve for futures contracts $\partial f(\xi)/\partial\xi$ under the constrained optimization (atomistic trade) is the same as the slope $\partial f(\xi_i)/\partial\xi_i$ in the unconstrained optimization where the futures price depends on the size of the trade (large trade).

Let the large futures trade $\xi = \int_1^N \xi_j dj$ made by a given trader be the aggregate sum of atomistic or negligible trades ξ_j taken by identical investors initiating a trade. Making a total differential of the futures price when trades are large and thus a function of these atomistic trades, I find that:

$$df(\xi) = df(\xi_1,...,\xi_N) = \int_1^N \frac{\partial f(\xi_i)}{\partial\xi_i} d\xi_i di \tag{22}$$

Differentiating (22) as a function of the large trade and using the fact that each atomistic trade is identical and therefore has the same slope $\partial f(\xi_i)/\partial\xi_i$, I find that:

$$\frac{\partial f(\xi)}{\partial\xi} = \int_1^N \frac{\partial f(\xi_i)}{\partial\xi_i}\frac{d\xi_i}{d\xi} di = \frac{\partial f(\xi_i)}{\partial\xi_i}\int_1^N \frac{d\xi_i}{d\xi} di \tag{23}$$

Integrating the small trades into the large one, we find that the slope for the large trade is the same as that for an atomistic or very small trade,

$$\frac{\partial f(\xi)}{\partial\xi} = \frac{\partial f(\xi_i)}{\partial\xi}\frac{\int_1^N d\xi_i di}{d\xi} = \frac{\partial f(\xi_i)}{\partial\xi}\frac{d\xi}{d\xi} = \frac{\partial f(\xi_i)}{\partial\xi} \tag{24}$$

Using equation (8), and noting that the futures price is fixed at the beginning of the period, the slope for the large trade equals therefore,

$$\frac{\partial f(\xi)}{\partial\xi} = \alpha\xi Var(\tilde{\pi}) = \alpha\xi Var(\tilde{p}) \tag{25}$$

Proof 2

I show that there is no unique price in the futures market even under the simplest setup where the only difference between the one who buys the futures contract and the one who sells is the degree of risk aversion.

Suppose that there are only two producers with exactly opposite hedging needs one selling a good and the other buying it, with different degrees of risk aversion, being otherwise perfectly identical. It is straightforward to determine their expected futures premium as in equation (8):

$$\pi_1 = \alpha_1 \left(\xi_1 Var(\tilde{\pi}_1) + Cov(\tilde{\pi}_1, (\tilde{p} - h)\tilde{q}) + S_1 Cov(\tilde{\pi}_1, \tilde{R}_M) \right) \tag{26}$$

$$\pi_2 = \alpha_2 \left(\xi_2 Var(\tilde{\pi}_2) + Cov(\tilde{\pi}_2, (h - \tilde{p})\tilde{q}) + S_2 Cov(\tilde{\pi}_2, \tilde{R}_M) \right) \tag{27}$$

From equation (8), the supply of futures is offered by producer 1 and demanded by producer 2. The futures premium of producer 1 is therefore opposite to that of producer 2 $(sign(\pi_1) = -sign(\pi_2))$. Using the futures premium definition, we can write,

$$\pi_1 = \alpha_1 \left(\xi_1 Var(f_1 - \tilde{p}) + Cov(f_1 - \tilde{p}, (\tilde{p} - h)\tilde{q}) + S_1 Cov(f_1 - \tilde{p}, \tilde{R}_M) \right) \tag{28}$$

$$\pi_2 = \alpha_2 \left(\xi_2 Var(\tilde{p} - f_2) + Cov(\tilde{p} - f_2, (h - \tilde{p})\tilde{q}) + S_2 Cov(\tilde{p} - f_2, \tilde{R}_M) \right) \tag{29}$$

Using the fact that the futures prices are determined at time 0, we can simplify further equations (28) and (29),

$$\pi_1 = \alpha_1 \left(\xi_1 Var(\tilde{p}) - Cov(\tilde{p}, (\tilde{p} - h)\tilde{q}) - S_1 Cov(\tilde{p}, \tilde{R}_M) \right) \tag{30}$$

$$\pi_2 = \alpha_2 \left(\xi_2 Var(\tilde{p}) + Cov(\tilde{p}, (h - \tilde{p})\tilde{q}) + S_2 Cov(\tilde{p}, \tilde{R}_M) \right) \tag{31}$$

Summing the supply (30) and demand (31) equations, and using the assumption that both producers hold the non-marketable good in the same amount, we find that:

$$\pi_1 + \pi_2 = (\alpha_1 \xi_1 + \alpha_2 \xi_2) Var(\tilde{p}) + (\alpha_2 - \alpha_1) \left(Cov(\tilde{p}, (\tilde{p} - h)\tilde{q}) + S_1 Cov(\tilde{p}, \tilde{R}_M) \right) \tag{32}$$

Using the clearing condition ($\xi_1 + \xi_2 = 0$) and the definition of the premiums,

$$f_1 - f_2 = (\alpha_1 - \alpha_2)\xi_1 Var(\widetilde{p}) + (\alpha_2 - \alpha_1)\Big(Cov(\widetilde{p},(\widetilde{p}-h)\widetilde{q}) + S_1 Cov(\widetilde{p},\widetilde{R}_M)\Big) \tag{33}$$

It can therefore be concluded that:

$$f_1 - f_2 \neq 0 \quad if \quad \alpha_1 \neq \alpha_2 \tag{34}$$

Hence, there is no unique futures price even in a very simplified model when the only difference between the seller and the buyer is their risk aversion.

Proof 3

This is a generalization of proof 2. I show that there is no unique price in the futures market in this model. The futures market is illiquid in equilibrium if investors have different degrees of risk aversion.

I consider first the premium asked by the N_1 short traders (35) and then that of the N_2 long traders (36). It is straightforward to show as in equation (8) that the futures premiums of the short and long traders are,

$$f_i - \widetilde{p} = \alpha_i \Big(\xi_i Var(\widetilde{p}) - Cov(\widetilde{p},(\widetilde{p}-h)\widetilde{q}) - S_i Cov(\widetilde{p},\widetilde{R}_M)\Big) \tag{35}$$

$$\widetilde{p} - f_i = \alpha_i \Big(\xi_i Var(\widetilde{p}) - Cov(\widetilde{p},(\widetilde{p}-h)\widetilde{q}) - S_i Cov(\widetilde{p},\widetilde{R}_M)\Big) \tag{36}$$

Assuming that N_1 agents are short and N_2 are short, I sum both equations over their type of agents,

$$\sum_{i=1}^{N_1}(f_i - \widetilde{p}) = Var(\widetilde{p})\sum_{i=1}^{N_1}\alpha_i\xi_i - \sum_{i=1}^{N_1}\alpha_i\Big(Cov(\widetilde{p},(\widetilde{p}-h)\widetilde{q}) + S_i Cov(\widetilde{p},\widetilde{R}_M)\Big) \tag{37}$$

$$\sum_{i=1}^{N_2}(\widetilde{p} - f_i) = Var(\widetilde{p})\sum_{i=1}^{N_2}\alpha_i\xi_i - \sum_{i=1}^{N_2}\alpha_i\Big(Cov(\widetilde{p},(\widetilde{p}-h)\widetilde{q}) + S_i Cov(\widetilde{p},\widetilde{R}_M)\Big) \tag{38}$$

Summing up the aggregate premiums (37) and (38), and noting that only the G hedgers have a position in the spot good, I find that:

$$\sum_{i=1}^{N_1}(f_i - \tilde{p}) + \sum_{j=1}^{N_2}(\tilde{p} - f_i) = Var(\tilde{p})\left[\sum_{i=1}^{N_1}\alpha_i \xi_i + \sum_{j=1}^{N_2}\alpha_j \xi_j\right] + \left[\sum_{j=1}^{N_1}\alpha_j S_j - \sum_{i=1}^{N_2}\alpha_i S_i\right]Cov(\tilde{p}, \tilde{R}_M) - \sum_{i=1}^{G}\alpha_i\left(Cov(\tilde{p},(\tilde{p}-h)\tilde{q})\right) \tag{39}$$

Using the market clearing constraint $\sum_{j=1}^{N_1}\xi_j = -\sum_{i=1}^{N_2}\xi_i \Rightarrow \xi_j = -\sum_{i=1}^{N_1}\xi_i - \sum_{k=1,k\neq j}^{N_2}\xi_k$ on equation (39),

$$\sum_{i=1}^{N_1}(f_i - \tilde{p}) - \sum_{j=1}^{N_2}(\tilde{p} - f_i) = Var(\tilde{p})\left[\sum_{i=1}^{N_1}\alpha_i \xi_i - \sum_{j=1}^{N_2}\alpha_j\left(\sum_{i=1}^{N_1}\xi_i + \sum_{k=1,k\neq j}^{N_2}\xi_k\right)\right] + \left[\sum_{j=1}^{N_1}\alpha_j S_j - \sum_{i=1}^{N_2}\alpha_i S_i\right]Cov(\tilde{p}, \tilde{R}_M) - \sum_{i=1}^{G}\alpha_i\left(Cov(\tilde{p},(\tilde{p}-h)\tilde{q})\right) \tag{40}$$

If risk aversion is constant, it is easy to see using equation (39) that the liquidity premium will disappear (simply factor out the risk aversion parameter and apply the market clearing condition). Prices will remain unequal because of differing exposure to the spot good and the non-marketable asset as can be seen more clearly in equation (40). Hence, there will be no single futures price to clear the futures market.

Proof 4

I find the futures premium when transactions are not assumed atomistic and the hedgers initiate the futures trades.

Summing the individual futures demand given in equation (18) assuming a single clearing futures price, I find that:

$$-Var(\tilde{\pi})\sum_{i=1}^{G}\xi_i = \sum_{i=1}^{\hat{N}+G}\frac{\pi}{\alpha_i} - \sum_{i=1}^{\hat{N}+G}Cov((\tilde{p}-h)\tilde{q} + S_i\tilde{R}_M, \tilde{\pi}_i) \tag{41}$$

Therefore, the futures premium, the expected difference between the buy and sell price for the spot is:

$$\pi = \frac{1}{k}\left[-Var(\tilde{\pi})\sum_{i=1}^{G}\xi_i + \sum_{i=1}^{\hat{N}+G}Cov((\tilde{p}-h)\tilde{q} + S_i\tilde{R}_M, \tilde{\pi}_i)\right] \tag{42}$$

Noting that only hedgers have a position in the spot good since they produce it, equation (40) further simplifies to:

$$\pi = \frac{1}{k}\left[-Var(\tilde{\pi})\sum_{i=1}^{G}\xi_i + \sum_{i=1}^{G} Cov((\tilde{p}-h)\tilde{q},\tilde{\pi}_i) + \sum_{i=1}^{\hat{N}+G} S_i Cov(\tilde{R}_M,\tilde{\pi}_i)\right] \tag{43}$$

The futures premium is therefore given by:

$$\pi = \frac{1}{k}\left[-Var(\tilde{\pi})\sum_{i=1}^{G}\xi_i + \sum_{i=1}^{G} Cov((\tilde{p}-h)\tilde{q},\tilde{p}) + \sum_{i=1}^{\hat{N}+G} S_i Cov(\tilde{p},\tilde{R}_M)\right] \tag{44}$$

$k = \left(\sum_{i=1}^{\hat{N}+G}\frac{1}{\alpha_i}\right)$ is the average tolerance in the futures market

Proof 5

I allow a natural squeeze in proof 4 by allowing in equation (11), the spot price to depend on the quantity of futures traded, that is delivery is done in a spot market which is no longer assumed liquid.

First, I derive the impact of the futures trade on the price of the spot good at delivery. The natural squeeze or price pressure of delivery is found from equation (6):

$$\frac{\partial\tilde{p}}{\partial\tilde{q}} = \frac{k}{\eta\tilde{\delta}}\left(\frac{\tilde{q}}{\tilde{\delta}}\right)^{\frac{1-\eta}{\eta}} \tag{45}$$

This is introduced into the unrestricted version of equation (10).

$$\frac{\partial\xi_i(f-\tilde{p})}{\partial\xi_i} = f - \tilde{p} + \left(\frac{\partial(f-\tilde{p})}{\partial\xi}\right)\xi_i = \tilde{\pi} + \left[\alpha_i\,\mathrm{var}(\tilde{p}) - \frac{k}{\eta\tilde{\delta}}\left(\frac{\xi_i}{\tilde{\delta}}\right)^{\frac{1-\eta}{\eta}}\right]\xi_i \tag{46}$$

It is straightforward to show that the investor's problem when both the spot and futures market are assumed illiquid gives the following futures premium:

$$\pi + \xi_i E\left[\alpha_i\,\mathrm{var}(\tilde{p}) - \left(k/\eta\tilde{\delta}\right)\left(\xi_i/\tilde{\delta}\right)^{\frac{1-\eta}{\eta}}\right]$$

$$= \alpha_i \Big(\xi_i Var(\tilde{\pi}) + Cov(\tilde{\pi}, (\tilde{p} - h)\tilde{q}) + S_i Cov(\tilde{\pi}, \tilde{R}_M) \Big) \tag{47}$$

INTRODUCTION TO ILLIQUIDITY AND THE WEALTH EFFECT

Hedger’s trades have an adverse impact on the futures price creating effectively an endogenous transaction cost increasing in times of uncertainty and acting as the source of illiquidity in these markets. This trading cost assumes a utility with Constant Absolute Risk Aversion so that the investor’s degree of risk aversion does not increase as he loses wealth. This wealth effect is the mechanism through which changes in investors’ wealth impact their attitude towards risk. As investors lose wealth, they become more risk averse and ask for a higher compensation to hold a risky asset thereby decreasing its price. In the next paper, illiquidity is shown to strengthen the wealth effect, which has been proven to be too weak empirically to explain the behavior of prices.

ILLIQUIDITY AND THE WEALTH EFFECT

ABSTRACT

Investors' attitudes towards risk and the resulting impact on prices in financial markets are determined by changes in their wealth. This wealth effect, however, provides a poor explanation of the mean, skewness and kurtosis of the observed distribution of futures prices especially in the derivatives market for reasonable values of the degree of risk aversion. This paper shows that illiquidity in the futures market, modeled endogenously as a trading cost, increases the strength of the wealth effect for the same degree of risk aversion. The resulting distribution of futures prices presents a more pronounced left fat tail and left skewness than would have been implied by the wealth effect alone.

INTRODUCTION

Although illiquidity is recognized in practice as a major disrupting factor in the functioning of financial markets, its theoretical foundation remains in doubt. This paper proposes a model, with a very general utility framework, in which illiquidity results from the inability of economic agents to share risk at no cost and takes the form of an endogenous trading cost whose importance increases as financial markets come under stress. Illiquidity tends to strengthen the wealth effect, which has traditionally been used to explain the behaviour of risk prices under stress. The wealth effect is the mechanism through which changes in the investors' wealth affect their attitude towards risk and thus prices on the financial markets. This mechanism provides a poor explanation (see Jackwerth and Brown (2001)) of the mean, skewness and kurtosis of the observed distribution of prices especially in the derivatives market for reasonable values of the degree of risk aversion. This paper shows that illiquidity in the futures market, modeled endogenously as a trading cost, increases the strength of the wealth effect by acting as a risk lever. Investors become more risk averse as their wealth falls and therefore ask for ever-higher risk premiums to become the counterparty to a futures contract. With illiquidity, for any decline in wealth, there is more risk to be shared and less willingness to assume it without a lower price and thus a higher return. The increase in futures risk premiums, reflecting greater illiquidity trading costs, is associated with a more pronounced left fat tail and left skewness in the distribution of futures prices in a market dominated by short hedgers than would have been implied by the wealth effect alone. Risk transfers are further studied using comparative statics.

Futures contracts are standardized instruments offered on organized markets. Investors can use them to hedge their risk stemming from changes in the price of the good underlying the contract, while speculators are willing to assume those risks in anticipation of a possible gain. The risks from having a position in the underlying asset are shared through the futures market as well as other derivatives markets. For example, a fast food chain can buy pork belly futures to hedge against the risk of an increase in the price of pork bellies, while a pork producer would sell pork belly futures. If both the producer and the fast food chain need to hedge the same amount for the same period, they could enter into a forward agreement, mimicking the futures contract available on organized markets, to exchange pork bellies at a predetermined forward price. Their risk is then defined as perfectly shared and there is no pressure on the futures market assuming that their wealth is solely determined by the price of pork. However, should they have different quantities to hedge, there will be a pressure[11] on the futures price so as to attract speculators willing to become the counterparty to the excess hedging supply or demand for futures. Hedging pressures are therefore transfers of risks for a price.

The idea that these risk transfers are costly was considered in Galy (2002) by relaxing the assumption that an investor can initiate a trade of a non-negligible size without having an impact on the market price. This creates endogenously an illiquidity trading cost that increases with the size of the futures trade and the volatility in the underlying spot price. This result was limited to investors with utility functions that exhibit constant absolute risk aversion. This paper generalizes this result by allowing the investor's degree of risk aversion to change as a function of his wealth thereby introducing the wealth effect. The trading cost increases with the size of the futures trade and by, approximately, the volatility in the underlying spot price. Furthermore, adding illiquidity to the wealth effect changes the distribution of futures prices.

Illiquidity trading costs make the distribution of futures prices more fat in the left tail, left skewed and dependent on the size of futures trades. If futures prices are above their expected value (or arbitrage price), as implied mainly by the conditions prevailing on the underlying good market, then speculators have little wealth at risk and are quite willing to bear that risk cheaply. By contrast, as the underlying good or the rest of their portfolio loses some value, speculators tend to ask for increasing premiums and the distribution of futures prices becomes thicker on the left for low futures prices than on the right for high futures prices. The illiquidity trading cost

[11] This assumes that there are limits to the ability to arbitrage

tends to aggravate the distortions produced by the wealth effect on the distribution of futures prices.

One might argue that Keynes (1930) thoroughly described how speculators become more fearful as they lose their wealth and flee the futures market. However, his analysis neglected the enhancing effect of illiquidity. This paper shows that as frightened speculators leave the market these become increasingly illiquid.

The paper is organized as follows. Section 1 briefly reviews the illiquidity and risk sharing literatures. Section 2 presents the theoretical model under two central assumptions: (i) futures prices are independent from the quantity traded; and (ii) markets are incomplete. The model is further developed, by relaxing the first assumption in Section 3 and the second assumption in Section 4. Section 5 draws on the model findings to analyse the risk sharing mechanism behind the wealth effect and changes in illiquidity.

1. PREVIOUS WORK ON ILLIQUIDITY AND RISK SHARING

Illiquidity[12] is derived endogenously within the microstructure literature generally in the form of a bid-ask spread or transaction costs, stemming mainly from inventory costs introduced by Demsetz (1968), market order processing costs from Garman (1976) and insider trading from Glosten and Milgrom (1985). These models assume the presence of market makers matching the supply and demand who establish bid-ask spreads to cover their operating costs. O'Hara (1994) provides a comprehensive overview of this field. These matching models, however, are of limited use to the vast array of equilibrium and arbitrage models used for pricing and risk management.

Equilibrium or arbitrage models, such as in Ericsson and Renault (2001), take illiquidity as given in the form of an exogenous trading cost and study how it impacts the decisions of investors. Acharya and Pedersen (2002) for example study a four beta CAPM with persistent illiquidity and study its impact on the investors' decision-making. Illiquidity and risk sharing were until now separate lines of research.

Risk sharing is concerned with the possibility for economic agents to exchange risks through financial markets, using derivatives as risk transfer vehicles. In the asset pricing literature, Dumas (1989) shows that when investors have different degrees of risk aversion and suffer a common risk, they must share the aggregate risk. Wang (1994) considers investors that are heterogeneous both in their investment opportunities and access to information. In the absence of information asymmetries, according to Wang (1994), selling by an investor increases the volume and decreases the price, increasing its expected return, as the asset's expected payoff has not changed. This leads other investors to buy the asset so that its price may remain independent of its volume. Risk sharing becomes of interest only if investors cannot correctly assess the asset's expected payoff. This happens when investors have heterogeneous beliefs as in Detemple and Murphy (1994) or asymmetries of information as in Wang (1994). Risk sharing will change depending on the source of the friction, such as non-insurable labor income shocks in Constantinides and Duffie (1996), and frictions or constraints imposed on the investors, such as the lower bound imposed by Grossman and Zhou (1996) on the hedger's wealth which forces them to hedge so as to respect it. This sharing of risks creates a demand for derivatives to shift risk between those who are willing to take on more of risk for a premium and those who must reduce their risk. Risk sharing comes under the broad denomination of 'portfolio insurance demand' in the derivatives literature.

In the derivatives literature, risk sharing among heterogeneous investors therefore creates a role for derivatives as risk transfer vehicles. Grossman and Zhou (1996) studied such an exchange of risk in complete markets and continuous time where one type of hedger is constrained not too lose a given fraction of his initial wealth, creating an asymmetric need to share risk and hence a demand for put options. Franke, Stapleton and Subrahmanyam (1998)

[12] See Galy (2002) for a more comprehensive review of this field

show that the degree to which investors face non-hedgeable background risks, such as labor income shocks or shocks to non traded assets, determines the exchange of risks. Finally, Bates (2001) considers the sharing of crash risk or negative stock market jumps and shows that it partially explains why stock options tend to overestimate volatility and the risk of a price jump.

When markets are incomplete, i.e. traders do not have enough uncorrelated assets to hedge all risk sources (Harrison and Pliska (1983) and Duffie and Huang (1985)), the sharing of risks is hampered. In such a context, the supply and demand of financial assets is imperfectly elastic as pointed out for example by Leisen (2002), which implies therefore a transfer of risk when trading. Leisen pointed out that derivatives would not be traded if the price of their underlying asset at the next trading date is assumed locally normal and markets are incomplete. Magill and Quinzii (1995) following Keynes (1930) define market incompleteness as a failure of the market to coordinate activities as all futures contract trades cannot be made at a predetermined price due to frictions. This creates a demand for cash to hedge against price uncertainty in addition to futures. The portfolio insurance demand literature extends this result to show that it creates a demand for options.

In a previous paper (Galy (2002)), I linked risk sharing with illiquidity and showed how illiquidity trading costs arise endogenously in an equilibrium model from the inability to share risk freely when investors have heterogeneous constant absolute risk aversion. This result is further generalized to a broad class of utilities in this paper allowing for the wealth effect and thereby altering the distribution of futures prices.

2. MODEL

I assume an infinite horizon model with two groups of utility maximizing agents, G producers of commodities and N speculators. Each agent maximizes the utility he expects from consuming the profits π_t generated by the sale of a good decided at time 0 but realized only at time t. Therefore, the agent faces a problem[13] each period whereby his cash flows are decided now but realized only t periods in the future. The utility function $V(\pi_{t+j}), j = 1,...,\infty$ is assumed to be increasing and concave in the investor's wealth. The $G + N$, $i = 1,...,G,...,G + N$, producers and speculators may have different degrees of risk aversion.

The hedger produces at time 0 the quantity y_0 of goods that will be sold at time t at an uncertain price $p_t(y_0)$ that he hedges by selling the quantity f_0 of futures contracts at a price of F_0 . Producing the good costs him $c(y_0)$ at time 0 but these costs are only realized at time t so that the cash flow from the production of goods decided at time 0 occurs t periods later. The investor's profits at time t equal the revenues generated by the sale of the consumption good minus the costs of production and the cost of hedging the production good by taking short futures contracts positions at time 0 and reversing the position at time t. The profits at time t are therefore given by equation (1):

$$\pi_t = p_t(y_0)y_0 - c(y_0) + f_0(F_0 - F_t) \qquad (1)$$

The investor chooses the number f_0 of futures contracts so as to maximize his concave utility $V(\pi_{t+j}), j = 1,...,\infty$. I will use the notation $V_t^{(i)}$ for the ith derivative with respect to profits of the utility function $V(\pi_{t+j}), j = 1,...,\infty$ at time t. The first derivative represents the marginal value of an additional dollar for the investor, the second represents the utility's curvature. It can be interpreted as a proxy for risk as can be seen in the Arrow-Pratt measure of risk aversion as well as other global measures of risk.

Differentiating $V(\pi_{t+j}), j = 1,...,\infty$ with respect to f_0 under the constraint of equation (1) yields equation (2). The futures' price is such that the expected value at time 0 of a dollar

[13] If preferences are separable, it is straightforward to see that the model reverts to a classical maximization of terminal wealth. The fact that the maximization is repeated every period does not change anything as no wealth is accumulated.

measured in terms of marginal utility $E_0 V_t^{(1)}$ invested in the futures contract has a zero return $\left(E_0 V^{(1)}((F_t - F_0)/F_0) = 0\right)$.

$$E_0 V_t^{(1)}(F_0 - F_t) = 0 \tag{2}$$

A simple transformation $E_0\,(V_t^{(1)} / V_0^{(1)})(F_t - F_0) = E_0 M_{t,0}(F_t - F_0) = 0$ shows that equation (2) can be rewritten in terms of the marginal rate of substitution $M_{t,0} = V_t^{(1)} / V_0^{(1)}$ between time 0 and time t. The marginal rate of substitution gives the price at which the agent would be willing to wait and consume later. It is a classic result that if the investor is allowed to trade t period risk free bonds with an interest of r_t, then the marginal rate of substitution will equal the risk free discount rate $E_0 M_{t,0} = 1/(1 + r_t)$. The marginal rate of substitution is therefore widely interpreted as the investor's personal risky discount rate $M_{t,0} = 1/(1 + \tilde{r}_t)$. Equivalently, it is the cash flow deflator used for present valuation but it may be risky as it depends on the investor's preferences

The futures demand and supply f_0, thereafter referred to as the futures demand, is found by differentiating equation (2) as a function of the futures price F_0 at time 0. Hedgers offer futures contracts to speculators who have the same utility maximization problem, but are long in the future contract and have no position in the underlying asset. The resulting futures demand is given by the following equation (3):

$$f_0 = \frac{E_0 V_t^{(1)}}{E_0 V_t^{(2)}(F_t - F_0)} \tag{3}$$

The futures demand is a function of the investor's utility. It is an increasing function of the marginal value of profits when hedging and an inverse function of the futures risk premium $F_t - F_0$.

To better understand the demand for futures contracts, I re-express it to introduce the familiar Arrow-Pratt risk aversion. Using the definition of the conditional expectation, conditional covariance[14] and that of local risk aversion $\rho_t = -V_t^{(2)} \big/ V_t^{(1)}$, the demand for futures contracts (3) becomes a function of the error in predicting the futures price $e_t^F = E_0 V_t^{(1)} (F_0 - F_t) - V_t^{(1)} (F_0 - F_t)$:

$$f_0 = \frac{E_0 V_t^{(1)}}{E_0 \rho_t E_0 V_t^{(1)} (F_0 - F_t) - Cov_0\left(\rho_t, e_t^F\right)} \tag{4}$$

The demand is, as before, an increasing function of the marginal value of profits for the investor $E_0 V^{(1)}$. It is also an inverse function of the futures price change $F_t - F_0$. This relation depends on whether markets are in equilibrium or not. When the futures price is in equilibrium, defined by equation (2), the first element of the denominator disappears from the equilibrium futures demand in equation (4). The second element of the denominator is the conditional covariance of the error in predicting the futures contract value with the Arrow-Pratt measure of risk aversion. The more the investor's risk aversion is correlated with the error in pricing e_t^F, the higher the hedging demand. In other words, the more fearful the investor becomes of pricing errors, the less he is willing to use futures contracts. Therefore, fear of mispricing is an additional factor driving the demand for futures. These risks are shared when investors trade either to enhance or reduce their exposure. As will be seen in the next section, trading creates endogenously illiquidity in the futures market in the form of a trading cost.

3. ILLIQUIDITY AS AN ENDOGENOUS TRADING COST

In this section, we relax the assumption that agents are able to buy or sell any quantity of futures contracts while leaving the futures price unchanged. Friend and Blume (1975) define such

[14] $Cov_0(X_t, Y_t) = E_0(X_t Y_t) - E_0(X_t) E_0(Y_t)$

markets as illiquid. This changes two things, first the futures price is no longer assumed to be independent of the quantity traded, and secondly each producer initiates a trade thereby creating a price pressure on the futures market that will attract speculators, defined as having no position in the futures contract's underlying asset[15].

3.1 MODEL WITH LARGE TRADES

Relaxing the assumption that the futures price is independent of the quantity traded in the producer's optimization problem changes the futures price. Compared to equation (2) where this assumption was imposed, an additional trading cost appears endogenously. It equals the quantity of futures traded multiplied by the expected change in the futures price created by the trade.

$$E_0 V_t^{(1)}(F_t - F_0) + f_0 E_0 V_t^{(1)} \left(\frac{\partial (F_t - F_0)}{\partial f_0} \right) = 0 \qquad (5)$$

This trading cost $TC = f_0 E_0 V_t^{(1)} \left(\partial (F_t - F_0) / \partial f_0 \right)$ is paid by the producer who sells futures for hedging purposes to the speculator buying the futures as a compensation for the sharing in the producer's risk.

3.2 DERIVING THE TRADING COST

The derivative $\partial (F_t - F_0) / \partial f_0$ in equation (5) of the trading cost is unknown. It can be determined directly from equation (2) where it was assumed implicitly to be too small to matter. This assumes that trading a large amount of futures contracts does not change the shape of the demand curve. Differentiating equation (2) as a function of the quantity of futures traded f_0, I find that the expected value $E_0 V_t^{(1)} \, \partial (F_t - F_0) / \partial f_0$ of the differential in terms of marginal utility equals the expected value in terms of risk of a deviation in the futures price $(F_t - F_0)^2$.

[15] See the proof of proposition 2 to see why this condition must hold.

$$E_0 V_t^{(1)} \frac{\partial (F_t - F_0)}{\partial f_0} = -E_0 \rho_t \left(V_t^{(1)} (F_t - F_0)^2 \right) = E_0 \left(V_t^{(2)} (F_t - F_0)^2 \right) \qquad (6)$$

where risk is measured by the Arrow Pratt measure of risk aversion. Note that the risk of a change in the futures price was shown before to be a determinant of the futures demand.

Having derived the trading cost, it is straightforward to find the price equation of the futures contract by introducing equation (6) into (5). This modified price equation of the futures contract equals the original equation (2) plus an additional trading cost $TC = f_0 E_0 \rho_t \left(V_t^{(1)} (F_t - F_0)^2 \right)$ on the right that lowers the futures price.

$$E_0 V_t^{(1)} (F_t - F_0) - f_0 E_0 \rho_t \left(V_t^{(1)} (F_t - F_0)^2 \right) = 0 \qquad (7)$$

The demand and supply of futures contracts must now be aggregated to find the futures price in the aggregate equilibrium in the presence of illiquidity trading costs.

3.3 EQUILIBRIUM IN AN ILLIQUID FUTURES MARKET

To find the futures price in the aggregate equilibrium, we must aggregate the supply of and demand for futures across investors assuming a single price will clear the futures market. The G producers initiate the futures contracts and are therefore the ones paying a trading cost to speculators, who do not have a position in the futures contract's underlying asset. Equation (7) which prices the futures contract can therefore be rewritten to introduce a logical operator I_i to separate hedging trades from speculative trades.

$$E_0 V_{t,i}^{(1)} (F_t - F_0) - I_i f_{0,i} E_0 \rho_{t,i} \left(V_{t,i}^{(1)} (F_t - F_0)^2 \right) = 0 \qquad (8)$$

$I_i = 1$ if i is a producer (hedger) and 0 otherwise (speculator)

Using the market clearing condition $\sum_{i=1}^{N+G} f_{0,i} = 0$ on the futures on equation (8), we find the equation (9) pricing futures contracts in the presence of illiquidity.

$$E_0\left(\sum_{i=1}^{N+G} V_{t,i}^{(1)}\right)(F_t - F_0) - \sum_{i=1}^{G} f_{0,i} E_0 \rho_{t,i}\left(V_{t,i}^{(1)}(F_t - F_0)^2\right) = 0 \tag{9}$$

The first term on the left of equation (9) is the futures contract's value in the market and the second term on the left is the aggregate illiquidity trading cost that must paid by hedgers in order for speculators to accept such trades. The illiquidity trading cost depends, among other things, on the hedgers' trades and their degree of risk aversion. Equation (9) can be rewritten as equation (10) to show how illiquidity and risk shifting are related using the definition of the conditional covariance to separate the futures risk premium from the aggregate marginal utility.

$$E_0(F_t - F_0) = -\frac{Cov_0\left(\sum_{i=1}^{N+G} V_{t,i}^{(1)}, F_t - F_0\right)}{E_0\left(\sum_{i=1}^{N+G} V_{t,i}^{(1)}\right)} + \frac{\sum_{i=1}^{G} f_{0,i} E_0 \rho_{t,i}\left(V_{t,i}^{(1)}(F_t - F_0)^2\right)}{E_0\left(\sum_{i=1}^{N+G} V_{t,i}^{(1)}\right)} \tag{10}$$

The futures risk premium or drift $E_0(F_t - F_0)$ on the left of equation (9) depends on the futures contract's ability to reduce or increase the investors' risk in the futures market. This is measured in equation (9) by the covariance $Cov_0\left(\sum_{i=1}^{N+G} V_{t,i}^{(1)}, F_t - F_0\right)$ on the right side of the equation between the futures risk premium or drift $F_t - F_0$ and the marginal value of a dollar for investors in that market $\sum_{i=1}^{N+G} V_{t,i}^{(1)}$. To shift the financial risk of producing with futures contracts, the producers must pay a trading cost (last element on the right of equation (10)) dependent on the size of the futures trades and approximately (see proof 3) the expected variance of the futures price for the producers adjusted for their expected degree of risk aversion.

3.4 FUTURES PRICE DISTRIBUTION

Introducing illiquidity makes the futures price distribution at time t F_t fatter on the left tail, more left skewed and dependent on the size of trades by strengthening the wealth effect.

As was seen in equations (9) and (10) and proof 3, illiquidity creates an endogenous trading cost that increases with the size of the futures trades, the expected degree of risk aversion and a measure related to the expected variance of the futures price. Illiquidity increases the strength of the wealth effect, which determines how shocks to producers affect the futures price, and thereby determines its distribution. Following a negative shock to the producer's profit, the wealth effect states that the producer's degree of risk aversion $\rho_{t,i}$ increases (see proof 1), as he is assumed to feel more vulnerable and is said to have a prudential motive ($V_{t,i}^{(3)} \leq 0$). This triggers a fall in the futures price as it becomes more desirable to short it for hedging and consequently an increase in the futures expected risk premium $E_0(F_t - F_0)$ and the marginal value of his profits $V_{t,i}^{(1)}$.

This, in turn, increases the illiquidity trading cost in the aggregate equilibrium $TC = \sum_{i=1}^{G} f_{0,i} E_0 \rho_{t,i} \left(V_{t,i}^{(1)} (F_t - F_0)^2 \right)$ as its components, the degree of risk aversion $\rho_{t,i}$, marginal value of profits $V_{t,i}^{(1)}$ and especially the element $(F_t - F_0)^2$, all increase. This last element is the square of the fall in the futures price resulting from the wealth effect, so that for every fall in the investor's profits, the trading cost increases by even more. The illiquidity trading cost increases therefore the strength of the wealth effect.

Illiquidity pushes therefore the futures price distribution to the left. As for every shock affecting the producer's profits, the futures price is lower than it would be without illiquidity. This assumes that in practice producers, selling futures to hedge, initiate a majority of trades. The futures price distribution is therefore more skewed and fatter on the left tail, than would be the case with only the wealth effect.

The properties of skewness and left fat tail can be deduced mathematically from equation (10). Skewness is a measure of the bias in the expectation and is a scaled function of the third moment of the futures price distribution $\mu_3 = (F_0 - E_0 F_t)^3$. From equation (10), the third moment is:

$$\mu_3 = \left[E_0 \left(\sum_{i=1}^{N+G} V_{t,i}^{(1)} \right) \right]^{-3} \left(Cov_0 \left(\sum_{i=1}^{N+G} V_{t,i}^{(1)}, F_t - F_0 \right) - \sum_{i=1}^{G} f_{0,i} E_0 \rho_{t,i} \left(V_{t,i}^{(1)} (F_t - F_0)^2 \right) \right)^3 \quad (11)$$

The futures price distribution is skewed $\mu_3 \neq 0$, as the elements on the right of equation (11) are different from zero. Intuitively, the distribution must be skewed to the left as the futures price today trades at a large discount to its expected price $F_0 < E_0 F_t$ under pressure from hedgers. The sign of the conditional covariance on the right hand side can be determined by deriving the marginal value of profits as a function of the futures premium conditional on the information available at time 0, all else equal.

$$\partial \left(\sum_{i=1}^{N+G} V_{t,i}^{(1)} \right) \Big/ \partial (F_t - F_0) = -\sum_{i=1}^{N+G} V_{t,i}^{(2)} f_{0,i} \quad (12)$$

The conditional covariance is therefore positive, as the utility is concave $V_{t,i}^{(2)} \leq 0$, and this more so for the G producers selling futures contracts. The trading cost is positive and the entire right side of equation (11) is therefore negative. The futures price distribution is therefore skewed to the left $F_0 < E_0 F_t$.

In addition to being skewed, the distribution is fatter on the left tail as described by the fourth moment of the futures distribution $\mu_4 = (F_0 - E_0 F_t)^4$. This corresponds to equation (11) with the powers changed from three to four. The fourth moment is, as the third, different from zero implying that the futures price distribution has a fat tail. The fourth moment is larger the higher the futures risk premium implying that the distribution has a fatter left tail.

The futures price distribution is left skewed with a fat left tail even if markets are liquid. These properties of the futures price distribution are enhanced by the presence of illiquidity. In essence, a Keynesian run (investors become increasingly fearful as their losses accumulate and leave the market) out of the futures market takes the liquidity out of the futures market. This implies in practice that one needs a lower degree of risk aversion to obtain these properties of the futures price distribution. These results depend however on the assumption that markets are incomplete.

4. WHEN ILLIQUIDITY TRADING COSTS CEASE TO MATTER: CONVERGENCE TO COMPLETE MARKETS

In this section, we show that illiquidity is a property of incomplete markets, which disappears as competing hedging products complete the financial markets, assuming that there are an infinite number of risk sources as well as an infinite number of assets available to complete the market.

Illiquidity trading costs are relevant only if the supply and demand curves are not perfectly elastic. As can be seen in equation (5), the derivative $\partial(F_t - F_0)/\partial f_0$ would then equal zero by definition and the trading cost would disappear. If a perfect substitute could be found by a replication or arbitrage strategy, then it is a classic result that the supply and demand must be perfectly elastic, as investors would switch from one good to the other whenever one market would come under trading pressure. The futures price would therefore be independent of the quantity traded. Illiquidity trading costs would then disappear, as the derivative $\partial(F_t - F_0)/\partial f_0$ within the definition of the trading cost $TC = f_0 E_0 V_t^{(1)}\left(\partial(F_t - F_0)/\partial f_0\right)$ equals zero. A futures market is therefore liquid when the supply and demand curves are perfectly elastic. Equivalently, there is no perfect substitute for the futures contract that could replicate its payoff, as would be the case by taking opposite positions on a call and a put with the same strike price on the same underlying commodity.

This ignores however that other substitute products are available for hedging. Proposition 1, derived in the appendix, shows that, as more options are available for hedging or speculation, the elasticity of the supply and demand curves increases as for every increase in the futures price, one can buy other hedging products as imperfect substitutes.

Proposition 1: Convergence to market completeness

Let N be the number of options available in the market. Let the number of risk sources that influence spot prices be infinite. For a given futures price F_0 and quantity of futures f_0:

a) The futures' supply and demand curves become more elastic as more options become available for hedging. Equivalently, the slope increases as the price and quantity are taken as given.

b) The slope of the futures curve is indeterminate when there is an infinite source of options to hedge an infinite source of risks.

$$\lim_{N\to\infty} \partial f_0 / \partial F_0 = \text{Indeterminate}$$

The slopes of the supply and demand curves become indeterminate as illustrated in figure 1, in the limit when markets are complete, and the supply and demand become impervious to futures price changes. This translates graphically below into curves that increasingly flatten as the markets become more complete.

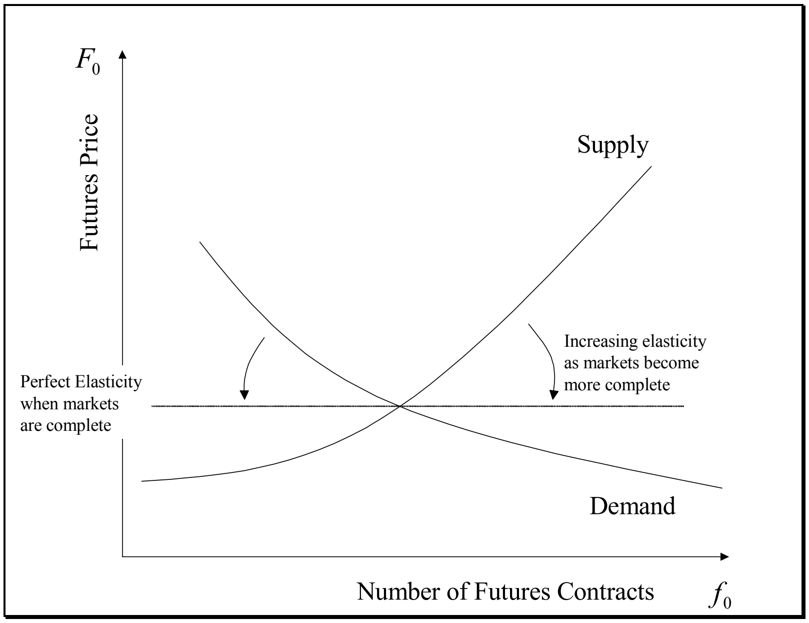
F_0
Futures Price
Supply
Increasing elasticity as markets become more complete
Perfect Elasticity when markets are complete
Demand
Number of Futures Contracts
f_0

Figure 1: Irrelevance of supply and demand when markets are complete

The illiquidity trading cost declines as financial markets become more complete and sharing risk becomes less relevant. The concept of risk sharing and consecutive illiquidity trading costs are studied using comparative statics in the next section.

5. RISK SHARING PROPOSITIONS

In this section, I study risk sharing or changes in the attitude towards risk that generate prices changes through the wealth effect and the illiquidity trading costs. The attitude towards risk will be measured by the degree of risk aversion and prudence. Risk aversion $\rho_{t,i} = -V_{t,i}^{(2)} \big/ V_{t,i}^{(1)}$ measures the investor's willingness to take risks when faced with uncertain profits. It measures the curvature of the utility as a function of profits and is positive, as the utility is assumed concave as a function of profits. The more curved it is, the more certain outcomes are preferred to uncertain ones. The degree of prudence $p_{t,i} = -V_{t,i}^{(3)} \big/ V_{t,i}^{(2)}$, sometimes called precaution, measures the investor's willingness to bear risk as his profits or wealth changes (see proof 1).

When the utility function is specified, risk sharing can be studied by comparing graphically how the payoff of a derivative varies as a function of the underlying asset price. If the graph is nonlinear and high for low states, that is these states are more expensive, then there is clearly an excess demand to hedge against these states, or so argues the portfolio insurance literature. As the utility function remains unspecified beyond the hypothesis that it is an increasing and concave function of profits, I use comparative statics[16] to see how risk alters both the price of and demand for futures contracts.

As we have seen, producers have risks from production, which cannot be hedged away and must be shared with the market. This implies a wealth effect and illiquidity trading costs. In this section, we will show in proposition 2 that this risk sharing puts pressure on the futures risk premium implying that they will tend to decrease (contango) or increase (normal backwardation) over time[17], with investors willing to buy or sell depending on their level of prudence. We will show in proposition 3 that this problem becomes acute in a high-risk situation such as a crash as there is an ever-increasing demand for hedging as hedgers find themselves more at risk. If investors in the futures market have non-separable preferences, in that the utility derived from tomorrow's profits cannot be separated from today's, then futures contracts become more expensive as investors care about how risk is resolved, thereby reducing the use of futures contracts (proposition 4).

5.1. COMPENSATION FOR RISK SHARING

Proposition 2, derived in the appendix, shows that risk transfers create an upward or downward trend for futures prices known as contango and normal backwardation respectively.

[16] See Varian (1992) for examples of the comparative statics method

[17] Keynes (1930) first developed the argument that the unwillingness to bear risk creates contango or normal backwardation in the futures prices.

Proposition 2: Relation between prudential and risk tolerant motive with contango or normal backwardation

$$F_t \geq F_0 \Rightarrow \quad f_0 \leq 0 \Rightarrow Cov_0\left(p_{V_t} \middle/ V_t^{(1)}, V_t^{(1)}(F_0 - F_t)\right) \leq 0 \qquad \text{for a speculator}$$

$$F_t \leq F_0 \Rightarrow \quad f_0 \geq 0 \Rightarrow Cov_0\left(p_{V_t} \middle/ V_t^{(1)}, V_t^{(1)}(F_0 - F_t)\right) \geq 0 \qquad \text{for a hedger}$$

Under normal backwardation, futures prices tend to increase over time $F_t \geq F_0$. Speculators are then willing to share the hedgers' risk $f_0 \leq 0$ as they are compensated by an increase in the futures price. This implies that a speculator becomes less prudent for every dollar invested as the value of the futures contract increases $Cov_0\left(p_{V_t} \middle/ V_t^{(1)}, V_t^{(1)}(F_0 - F_t)\right) \leq 0$. Hedgers push the futures price down by selling futures at time 0 and increase it at time t by reversing their positions thus offering a risk premium to speculators.

Under contango, futures prices tend to decrease $F_t \leq F_0$. Hedgers are more than willing to use futures contracts to hedge $f_0 \geq 0$, as they can transfer risk to speculators and receive a risk premium for it. They become more prudent for every dollar invested as the value of the futures contract increases $Cov_0\left(p_{V_t} \middle/ V_t^{(1)}, V_t^{(1)}(F_0 - F_t)\right) \geq 0$. This unlikely situation $F_t \leq F_0$ is possible, but selling pressures from hedgers will push the futures price down immediately. Therefore, one cannot have a downward trend or contango without allowing for an excess of long hedgers in the model to push the current futures price upwards.

5.2. NONLINEAR RISK SHARING

Proposition 2 shows that contango or normal backwardation means that investors are willing to share risk if they are compensated. The result that risk sharing creates a pressure on futures prices was confirmed in section 3 and is further studied in the proposition 3.

Proposition 3, derived in the appendix, shows how the supply and demand of futures contracts changes with the degree of risk aversion.

Proposition 3: Risk aversion mechanism

$$sign\left\{\partial f_0 \big/ \partial \rho_{V_t}\right\} = -sign\left\{E_0 V_t^{(1)} (F_0 - F_t)\right\}$$

$$sign\left\{\partial^2 f_0 \big/ \partial \rho_{V_t}^2\right\} = sign\left\{E_0 \rho_{V_t} V_t^{(1)} (F_t - F_0)\right\}$$

When futures prices increase over time $F_t \geq F_0$ (normal backwardation), hedgers sell futures contracts as their degree of risk aversion increases $sign\left\{\partial f_0 \big/ \partial \rho_{V_t}\right\} \geq 0$ (as a result of the wealth effect for example) and this at a decreasing rate $sign\left\{\partial^2 f_0 \big/ \partial \rho_{V_t}^2\right\} \leq 0$. Hence, the price pressure on the futures markets will be strongest, when hedgers feel the most vulnerable or equivalently have a high degree of risk aversion. In the language of Keynes (1930), proposition 3 states that speculators are frightened and flee the market during a crisis. As was shown in section 3.4, the futures market becomes less liquid at an increasingly fast pace through the illiquidity trading cost.

5.3. DYNAMIC RISK SHARING

The attitude towards risk has been described by the degree of risk aversion to uncertainty in profits at a given point in time and is therefore static. In the next section, we add a temporal or dynamic dimension to the attitude towards risk by introducing non-separability in the investor's preferences. The investor then cares both about uncertainty in his profits and how this uncertainty resolves itself or equivalently its dynamics. This dynamic component of the attitude towards risk is shown to effectively increase the investor's degree of risk aversion, limiting his willingness to share risk, and therefore strengthening the wealth effect.

Non-separability of preferences increases the investor's degree of risk aversion by introducing a dynamic component (proposition 4.1) and decreases the investor's willingness to shift risk through time (proposition 4.3). The difference between the degree of risk aversion under non-separable preferences and that under separable preferences is a function of the elasticity of the marginal rate of substitution through time σ and the parameter α, which controls the degree to which preferences are non-separable[18] (proposition 4.4). Therefore, the speculator requires a greater compensation for entering into a futures contract, as the expected value of the futures contract at time t is greater. The contract becomes more expensive through these two mechanisms so that the demand for futures contracts falls.

The attitude towards risk is said to be dynamic if the investors' preferences are not an additive sum of his instantaneous utility. Until now, the investors were assumed to have preferences that were additive or separable so that each period the producer took decisions independently of how it would impact his future instantaneous utility and hence future decisions. The investor's utility function $V(\pi_{t+j}), j = 1,...,\infty$ under non-separable preferences is chosen as a nonlinear sum of instantaneous utilities (13) that reverts to the linear or separable case used before when the parameter $\alpha = 0$. The investor's attitude towards time is given by the parameter β discounting his instantaneous utilities, while his time horizon is given by the parameter T_1.

$$V_t = \left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i}) \right]^{1-\alpha} \tag{13}$$

The non-separability of the utility of profits across time introduces a dynamic to risk, the investor taking into account how a risky investment will evolve. For example, managers face great anxiety or elation today when undertaking a high-risk project, as they are uncertain not only how it will work out today and tomorrow, but how the risk of that project will evolve as time goes by. The degree of risk aversion under separable preferences corresponds therefore to the one

[18] If $\alpha = 0$ the utility function reverts to one of separable preferences.

based on the instantaneous utility $\rho_{U_t} = -U_t^{(2)} \big/ U_t^{(1)}$, while under non-separable preferences it is based on the function of the instantaneous utilities $\rho_{V_t} = -V_t^{(2)} \big/ V_t^{(1)}$.

Proposition 4.1, proved in the annex, shows that the degree to which preferences are non-separable influences the demand for futures contracts.

Proposition 4.1: Dynamic risk

$\frac{\partial f_0}{\partial \alpha} \leq 0$ if $F_t \leq F_0$ the futures price is decreasing and $\rho_V^2 \geq \left((1+\alpha)\theta\right)^{-1}$

From proposition 2, we know that this proposition concerns a hedger selling futures contracts as he can both shift his risk and be compensated for it. An increase in dynamic risk, controlled by the parameter α, decreases hedging demand for a given level of the instantaneous degree of risk aversion $\rho_{V_t}^2 \geq \left((1+\alpha)\theta\right)^{-1}$ where θ is the difference between the degree of risk aversion under non-separable and separable preferences. An increase in the importance of dynamic risk controlled by α increases the importance of past and future profits on the utility $V(\pi_{t+j}), j = 1,...,\infty$ derived by the investor from his profits. The hedger uses fewer futures contracts in the presence of dynamic risk even though he can shift his risk and be compensated for it.

Proposition 4.2 shows that introducing a dynamic component to the investor's attitude towards risk increases the investor's degree of risk aversion, thereby strengthening the wealth effect.

Proposition 4.2: Relation between static and dynamic risk aversions measures

$$\rho_{V_t} = \rho_{U_t} + \frac{\alpha U_t^{(1)}}{\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i})}$$

The Arrow-Pratt measure of risk aversion $\rho_{V_t} = -V_t^{(2)} \big/ V_t^{(1)}$ of the utility with non-separable preferences exceeds the one with separable preferences $\rho_{U_t} = -U_t^{(2)} \big/ U_t^{(1)}$ by a factor $\alpha U_t^{(1)} V_t^{-1/1-\alpha}$ that is positive if $\alpha \geq 0$ and $V(\pi_{t+j}, j = 1,...,\infty) \geq 0$. When preferences are separable $\alpha = 0$, the two measures of risks are equal. Under these conditions, assuming separable preferences, and $0 < \alpha \leq 1$ for the utility to be concave (see proof 2), is therefore equivalent to increasing the investor's degree of risk aversion. Note that risk aversion under separable preferences can be said to be dynamic as it depends on future utilities or states, while the traditional measure is static.

Proposition 4.3 shows that the increase in risk aversion comes from an unwillingness to shift risk through time when preferences are non-separable.

Proposition 4.3: Relation between the elasticity of the marginal rate of substitution through time σ_t (EMRST) and the parameter α:

$$\sigma_t = \frac{\partial \ln V_t}{\partial \ln \pi_t} = (1-\alpha) \frac{\pi_t U_t^{(1)}}{\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i})}$$

The elasticity of the marginal rate of substitution through time (EMRST) is a decreasing function of the parameter α. The more the investor is concerned about risk resolution, the less he is willing to shift risk through time and hence the higher the risk premium he will require to enter into a futures contract.

The previous set of propositions showed that risk sharing becomes more difficult when introducing a dynamic component to risk. Proposition 4.4 summarizes these results.

Proposition 4.4: Relation between the degrees of absolute risk aversion $\rho_{V_t}^A$, $\rho_{U_t}^A$ and the EMRST σ

$$\rho_{V_t}^A = \rho_{U_t}^A + \sigma \frac{\alpha}{1-\alpha}$$

Proposition 4.4 follows directly from propositions 4.2 and 4.3 It shows the dynamic risk aversion $\rho_{V_t}^A = -V_t^{(2)}\pi_t \big/ V_t^{(1)}$ difference with the static risk aversion $\rho_{U_t}^A = -U_t^{(2)}\pi_t \big/ U_t^{(1)}$ is a function of the EMRST (σ) and the parameter α. When the parameter α increases, the ratio $\alpha \big/ 1-\alpha$ increases and the EMRST σ decreases (proposition 4.2). The net impact is more clearly seen in proposition 4.2, where the premium increases linearly with α. Hence, dynamic risk as measured by α increases the investor's degree of risk aversion by decreasing his willingness to shift risk through time. This dynamic component disappears when preferences are separable $\alpha = 0$.

CONCLUSION

Risk sharing creates an illiquidity trading cost that strengthens the wealth effect. This in turn increases the fatness of the left tail and skewness of the distribution of futures prices beyond that created by the wealth effect. Risk sharing becomes increasingly difficult as investors find themselves at risk, creating a pressure on the futures prices for speculators to accept the risk unloaded by hedgers. In the presence of non-separable preferences, this mechanism is again strengthened as investors worry about how uncertainty will resolve itself.

ANNEX

Proposition 1: Convergence to market completeness

Let N be the number of options available in the market. Let the number of risk sources that influence spot prices be infinite. For a given futures price F_0 and quantity of futures f_0:

a) The futures' supply and demand curves become more elastic as more options become available for hedging. Equivalently, the slope increases as the price and quantity are taken as given.

b) The slope of the futures curve is indeterminate when there is an infinite source of options to hedge an infinite source of risks.

$\lim\limits_{N\to\infty} \partial f_0 / \partial F_0 =$ Indeterminate

Proof proposition 1:

To prove proposition 1, we must first find the slope of the futures demand. The demand function (3) is derived as a function of the futures price.

$$\frac{\partial f_0}{\partial F_0} = \frac{-f_0 E_0 V_t^{(2)}(E_0 V_t^{(2)}(F_t - F_0)) - E_0 V_t^{(1)}(f_0 E_0 V_t^{(3)}(F_t - F_0) - E_0 V_t^{(2)})}{\left(E_0 V_t^{(2)}(F_t - F_0)\right)^2} \tag{14}$$

$$(3) \Rightarrow \frac{\partial f_0}{\partial F_0} = \frac{-E_0 V_t^{(1)}(E_0 V_t^{(2)}) - E_0 V_t^{(1)}(f_0 E_0 V_t^{(3)}(F_t - F_0) - E_0 V_t^{(2)})}{\left(E_0 V_t^{(2)}(F_t - F_0)\right)^2} \tag{15}$$

$$(3) \Rightarrow \frac{\partial f_0}{\partial F_0} = f_0 \frac{-E_0 V_t^{(2)} - f_0 E_0 V_t^{(3)}(F_t - F_0) + E_0 V_t^{(2)}}{\left(E_0 V_t^{(2)}(F_t - F_0)\right)} \tag{16}$$

$$\frac{\partial f_0}{\partial F_0} = -f_0^2 \frac{E_0 V_t^{(3)}(F_t - F_0)}{E_0 V_t^{(2)}(F_t - F_0)} \tag{17}$$

using equation (40), derived as part of the proof of proposition 2, we have therefore:

$$\frac{\partial f_0}{\partial F_0} = -2 f_0 \; \frac{E_0 V_t^{(2)}}{E_0 V_t^{(2)} (F_t - F_0)} \tag{18}$$

$$\frac{\partial f_0}{\partial F_0} = -2 \frac{E_0 V_t^{(1)} E_0 V_t^{(2)}}{\left(E_0 V_t^{(2)} (F_t - F_0) \right)^2} \tag{19}$$

The hedger has now a portfolio of N derivatives on the underlying product giving each a different payoff $g_i(p(y_t),..)$ at maturity. The sources of uncertainty on the spot price remain unspecified and may be multiple.

$$\pi_{t,N} = p(y_t) y_0 - c(y_0) - f_0 (F_t - F_0) - \sum_{i=1}^{N} f_{0,i} g_i (p(y_t),..) \tag{20}$$

Proof of proposition 1.a):

if $(F_t - F_0) \leq 0$ we just showed that $0 \geq V_t^{(2)}(\pi_{t,N+1}) \geq V_t^{(2)}(\pi_{t,N})$

Using equation (19), the inverse of the price elasticity is

$$\varepsilon_{F_0, f_0}^{-1} = \frac{\partial f_0}{\partial F_0} \frac{F_0}{f_0} = -2 F_0 \; \frac{E_0 V_t^{(2)}}{E_0 V_t^{(2)} (F_t - F_0)} \tag{21}$$

using the equilibrium demand of futures (3) into (21)

$$\varepsilon_{F_0, f_0}^{-1} = \frac{\partial f_0}{\partial F_0} \frac{F_0}{f_0} = -2 F_0 \, f_0 \; \frac{E_0 V_t^{(2)}}{E_0 V_t^{(1)}} \tag{22}$$

Note the more options are available, the more the utility changes so that the demand for futures contracts, which depends on the first and second degree of the utility, changes as other options become available. For a given price and quantity of futures in equilibrium at time 0, I find how the elasticity of the curve changes. The introduction of a new good changes not only the shape of the curve as measured by elasticity but its position.

$$\frac{\left. \varepsilon_{F_0, f_0}^{-1} \right|_{N+1}}{\left. \varepsilon_{F_0, f_0}^{-1} \right|_{N}} = \frac{E_0 V_t^{(1)}(\pi_{t,N})}{E_0 V_t^{(1)}(\pi_{t,N+1})} \frac{E_0 V_t^{(2)}(\pi_{t,N+1})}{E_0 V_t^{(2)}(\pi_{t,N})} \tag{23}$$

As the utility is concave $V_t^{(2)}(.) \le 0$, then $V_t^{(1)}(.)$ is a decreasing function so that:

$$E_0 V_t^{(1)}(\pi_{t,N+1}) \ge E_0 V_t^{(1)}(\pi_{t,N}) \qquad (24)$$

Whether $(F_t - F_0) \ge 0$ or $(F_t - F_0) \le 0$, I show in the demonstration of proposition 1.1 that it implies: $E_0 V_t^{(2)}(\pi_{t,N+1}) \le E_0 V_t^{(2)}(\pi_{t,N})$

So that equation (23) => $\dfrac{\varepsilon^{-1}_{F_0,f_0}\Big|_{N+1}}{\varepsilon^{-1}_{F_0,f_0}\Big|_{N}} \le 1 => \varepsilon_{F_0,f_0}\Big|_{N+1} \ge \varepsilon_{F_0,f_0}\Big|_{N}$ (25)

The elasticity of the supply or demand curve increases the more options are available.

Proof of proposition 1.b):

To find the limit of the slope given by equation (19), I first find the limit of the denominator and numerator. The proofs of propositions 1.a and 1.b assume the price and quantity to be constant.

1) I find the limit of the denominator $\lim\limits_{N\to\infty} E_0(V_t^{(2)}(\pi_{t,N+1}))(F_t - F_0) = ?$

$$E_0(V_t^{(2)}(\pi_{t,N+1}))(F_t - F_0) = E_0(V_t^{(2)}(\pi_{t,N} - f_{0,N+1}g_{N+1}(.)))(F_t - F_0)$$

- if $(F_t - F_0) \le 0$ then by proposition 2, $V_t^{(3)} \le 0$, then $V_t^{(2)}(.)$ is a decreasing function of profits so that, supposing that the additional options are used for speculating ($f_{0,N+1}g_{N+1}(.) \le 0$):

  $V_t^{(2)}(\pi_{t,N+1}) \le V_t^{(2)}(\pi_{t,N}) \le 0$

  => $\lim\limits_{N\to\infty} E_0(V_t^{(2)}(\pi_{t,N+1})) = -\infty$

  and as $(F_t - F_0) \le 0$ =>

  $$E_0(V_t^{(2)}(\pi_{t,N+1}))(F_t - F_0) \ge E_0(V_t^{(2)}(\pi_{t,N}))(F_t - F_0) \ge 0 \qquad (26)$$

Hence, the denominator of equation (17) is an increasing function of the number of substitute options available for speculating. In the limit, the denominator therefore converges to infinity.

$$\lim_{N\to\infty} E_0(V_t^{(2)}(\pi_{t,N+1}))(F_t - F_0) = \infty \qquad (27)$$

- if $(F_t - F_0) \geq 0$ then by proposition 2, $V_t^{(3)} \geq 0$, then $V_t^{(2)}(.)$ is an increasing function of profits so that, supposing that the additional options are used for hedging ($f_{0,N+1}g_{N+1}(.) \geq 0$):

  $V_t^{(2)}(\pi_{t,N+1}) \leq V_t^{(2)}(\pi_{t,N}) \leq 0$

  $=> \lim_{N\to\infty} E_0(V_t^{(2)}(\pi_{t,N+1})) = -\infty$

  and as $(F_t - F_0) \geq 0 =>$

  $$E_0(V_t^{(2)}(\pi_{t,N+1}))(F_t - F_0) \leq E_0(V_t^{(2)}(\pi_{t,N}))(F_t - F_0) \leq 0 \qquad (28)$$

  Hence, the denominator of equation (17) is a decreasing function of the number of substitute options available for hedging. In the limit, the denominator therefore converges to zero.

$$\lim_{N\to\infty} E_0(V_t^{(2)}(\pi_{t,N+1}))(F_t - F_0) = 0 \qquad (29)$$

2) I find the limit of the numerator: $\lim_{N\to\infty} E_0 V_t^{(1)} E_0 V_t^{(2)} = ?$

If $(F_t - F_0) \leq 0$ or $(F_t - F_0) \geq 0$, we just showed that:

$V_t^{(2)}(\pi_{t,N+1}) \leq V_t^{(2)}(\pi_{t,N}) \leq 0 => \lim_{N\to\infty} E_0(V_t^{(2)}(\pi_{t,N+1})) = -\infty =>$

$\lim_{N\to\infty} E_0(V_t^{(2)}(\pi_{t,N+1}))E_0(V_t^{(1)}(\pi_{t,N+1})) = -\infty.0 = ID$ (Indeterminate)

Since the utility is concave $V_t^{(2)}(.) \leq 0$ then $V_t^{(1)}(.)$ is a decreasing function.

If $(F_t - F_0) \geq 0$ => $0 \leq V_t^{(1)}(\pi_{t,N+1}) \leq V_t^{(1)}(\pi_{t,N})$ => $\lim_{N\to\infty} E_0(V_t^{(1)}(\pi_{t,N+1})) = 0$ (30)

If $(F_t - F_0) \leq 0 \Rightarrow 0 \leq V_t^{(1)}(\pi_{t,N}) \leq V_t^{(1)}(\pi_{t,N+1}) \Rightarrow \lim_{N\to\infty} E_0(V_t^{(1)}(\pi_{t,N+1})) = \infty$ (31)

Using the results of 1), 2) and equation (19), we have:

$$\text{if } (F_t - F_0) \leq 0 \Rightarrow \lim_{N\to\infty} \frac{\partial f_0}{\partial F_0} = -\frac{-\infty.\infty}{\infty} = ID \tag{32}$$

$$\text{if } (F_t - F_0) \geq 0 \Rightarrow \lim_{N\to\infty} \frac{\partial f_0}{\partial F_0} = -\frac{0.(-\infty)}{0} = ID \tag{33}$$

The slope of the supply or demand curves are therefore indeterminate in the limit. In the limit, the two are perfectly elastic or flat so that the quantity and hence the slope cannot be determined.

Proposition 2: Relation between prudential and risk tolerant motive with contango or normal backwardation

$F_t \geq F_0 \Rightarrow \quad f_0 \leq 0 \Rightarrow Cov_0\left(p_t / V_t^{(1)}, V_t^{(1)}(F_0 - F_t)\right) \leq 0$ for a speculator

$F_t \leq F_0 \Rightarrow \quad f_0 \geq 0 \Rightarrow Cov_0\left(p_t / V_t^{(1)}, V_t^{(1)}(F_0 - F_t)\right) \geq 0$ for a hedger

Proof of Proposition 2:

To find the demand for futures contracts, I use the demand for futures (3) and expand its denominator using the definition of the covariance.

$$f_0 = \frac{E_0 V_t^{(1)}}{E_0 \rho_t E_0 V_t^{(1)}(F_0 - F_t) + Cov_0\left(\rho_t, V_t^{(1)}(F_0 - F_t)\right)} \tag{34}$$

Using the first order condition of equilibrium (2), the futures demand in equilibrium is therefore given by,

$$f_0^{equilibrium} = \frac{E_0 V_t^{(1)}}{Cov_0\left(\rho_t, V_t^{(1)}(F_0 - F_t)\right)} \tag{35}$$

Equation (35) shows an investor in equilibrium will hold futures contracts long if his degree of risk aversion moves in the same direction as the value of the futures contract and short otherwise.

$$sign\left\{f_0\right\} = sign\left\{Cov_0\left(\rho_t, V_t^{(1)}(F_0 - F_t)\right)\right\} \tag{36}$$

Therefore, the producer hedges because the value of the futures contract increases when the expected spot price of his goods falls. An investor without a position in the underlying asset is less risk averse when the value of the futures contract increases and is therefore willing to act as a speculator.

If the futures prices are increasing (normal backwardation), then as can be seen in equation (35) using the result of (36), then the speculator is willing to take a long position. If the futures prices are increasing (contango), then as can be seen in equation (35) using the result of (36), then the hedger is willing to take a short position. These results are summarized below:

$F_0 - F_t \leq 0 \Rightarrow \quad f_0 \leq 0$ for a speculator

$F_0 - F_t \geq 0 \Rightarrow \quad f_0 \geq 0$ for a hedger

To find the impact of the prudential motive on risk taking, I derive equation (2) noted here (36), pricing the futures contract as a function of the futures price F_t , as a function of the futures position f_0 and then the futures price F_0 , I obtain equation (37) and then (38):

$$E_0 V_t^{(1)} (F_t - F_0) = 0 \tag{37}$$

$$E_0 V_t^{(2)} (F_t - F_0)^2 = 0 \tag{38}$$

$$E_0 V_t^{(2)} 2(F_0 - F_t) + E_0 V_t^{(3)} f_0 (F_0 - F_t)^2 = 0 \tag{39}$$

Equation (38) can be rewritten in the following manner:

$$2E_0 (-V_t^{(2)})(F_0 - F_t) = E_0 V_t^{(3)} f_0 (F_0 - F_t)^2 \tag{40}$$

$$2E_0 \rho_t V_t^{(1)} (F_0 - F_t) = f_0 E_0 p_t \rho_t V_t^{(1)} (F_0 - F_t)^2 \tag{41}$$

$$f_0^2 = 2E_0 V_t^{(1)} \Big/ Cov_0 \left(\frac{p_t}{V_t^{(1)}} \rho_t V_t^{(1)} (F_0 - F_t), V_t^{(1)} (F_0 - F_t) \right) \tag{42}$$

From the result of equation (36), equation (42) can only hold if the hedger becomes more prudent for every dollar invested as the value of the futures contract increases. Conversely, a speculator becomes less prudent for every dollar invested as the value of the futures contract increases.

$$Cov_0 \left(p_t / V_t^{(1)}, V_t^{(1)} (F_0 - F_t) \right) \geq 0 \quad f_0 \geq 0 \tag{43}$$

$$Cov_0 \left(p_t / V_t^{(1)}, V_t^{(1)} (F_0 - F_t) \right) \leq 0 \quad f_0 \leq 0$$

The proposition 2 therefore holds.

Proposition 3: Risk aversion mechanism

$$sign\left\{\partial f_0 / \partial \rho_t\right\} = -sign\left\{E_0 V_t^{(1)}(F_0 - F_t)\right\}$$

$$sign\left\{\partial^2 f_0 / \partial \rho_t^2\right\} = sign\left\{E_0 \rho_t V_t^{(1)}(F_t - F_0)\right\}$$

Proof of Proposition 3:

Deriving the futures demand (supply) of equation (3) as a function of the second derivative of the utility $V_t^{(2)}$, I find that:

$$\frac{\partial f_0}{\partial \rho_t} = \frac{-E_0(V_t^{(1)}) E_0 V_t^{(1)}(F_0 - F_t)}{(E_0 \rho_t V_t^{(1)}(F_0 - F_t))^2} \tag{44}$$

Introducing equation (3) into equation (44), the equation simplifies to:

$$\frac{\partial f_0}{\partial \rho_t} = -\frac{f_0^2}{E_0 V_t^{(1)}} E_0 V_t^{(1)}(F_0 - F_t) \tag{45}$$

The utility being increasing, it follows therefore that:

$$sign\left\{\partial f_0 / \partial \rho_t\right\} = -sign\left\{E_0 V_t^{(1)}(F_0 - F_t)\right\} \tag{46}$$

Deriving equation (47) as a function of the degree of risk aversion:

$$\frac{\partial^2 f_0}{\partial \rho_t^2} = \frac{2\left(E_0 V_t^{(1)}\right)\left(E_0 V_t^{(1)}(F_0 - F_t)\right)^2}{(E_0 \rho_t V_t^{(1)}(F_0 - F_t))^3} \tag{48}$$

The utility being increasing, it follows therefore that:

$$sign\left\{\partial^2 f_0 / \partial \rho_t^2\right\} = sign\left\{E_0 \rho_t V_t^{(1)}(F_0 - F_t)\right\} \tag{49}$$

Hence, from the results of (46) and (49), proposition 3 is proved.

Proposition 4.1: Dynamic risk

$\frac{\partial f_0}{\partial \alpha} \leq 0$ if $F_t \leq F_0$ the futures price is decreasing and $\rho_{V_t}^2 \geq \left((1+\alpha)\theta\right)^{-1}$

Note that the index for risk aversion now depends on whether we are considering the instantaneous utility function $U(\pi_{t+j}), j = 1,...,\infty$ or the utility function $V(\pi_{t+j}), j = 1,...,\infty$.

Proof of Proposition 4.1:

To find the impact of dynamic risk on the use of futures contracts $\partial f_0 / \partial \alpha$, I must first find $\partial V_t^{(1)} / \partial \alpha$ and $\partial V_t^{(2)} / \partial \alpha$. The intertemporal utility function under non-separable preferences is given by equation (50). α represents the degree to which present and future utilities are linked for the investor when taking a decision today. When it equals zero preferences are said to be separable as tomorrow's utility is added to today's. β is the standard parameter describing the agent's preference for time and T_1 is the horizon of the investor

$$V_t = \left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i})\right]^{1-\alpha} \tag{50}$$

The utility's first derivative is given by equation (51) and its second derivative by equation (53):

$$V_t^{(1)} = (1-\alpha) U_t^{(1)} \left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i})\right]^{-\alpha} = (1-\alpha) U_t^{(1)} V_t^{\frac{-\alpha}{1-\alpha}} \tag{51}$$

$$V_t^{(2)} = (1-\alpha)\left(U_t^{(2)} \left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i})\right]^{-\alpha} - \alpha \left(U_t^{(1)}\right)^2 \left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i})\right]^{-\alpha-1} \right) \tag{52}$$

$$V_t^{(2)} = (1-\alpha)\left[U_t^{(2)} V_t^{\frac{-\alpha}{1-\alpha}} - \alpha \left(U_t^{(1)}\right)^2 V_t^{-\frac{1+\alpha}{1-\alpha}} \right] \tag{53}$$

1) $\partial V_t^{(1)} / \partial \alpha$: Deriving the first derivative of the utility (51)as a function of the parameter α

$$\frac{\partial V_t^{(1)}}{\partial \alpha} = -U_t^{(1)} \left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i})\right]^{-\alpha} + (1-\alpha) U_t^{(1)} \left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i})\right]^{-\alpha} \ln\left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i})\right] \tag{54}$$

$$\frac{\partial V_t^{(1)}}{\partial \alpha} = U_t^{(1)} V_t^{\frac{-\alpha}{1-\alpha}} \left(-1 + (1-\alpha)\ln\left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i}) \right] \right) \tag{55}$$

2) $\partial V_t^{(2)} \big/ \partial \alpha$: Deriving the second derivative of the utility (53) as a function of the parameter α

$$\begin{aligned}
\frac{\partial V_t^{(2)}}{\partial \alpha} &= \\
&-\left(U_t^{(2)} \left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i}) \right]^{-\alpha} - \alpha \left(U_t^{(1)} \right)^2 \left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i}) \right]^{-\alpha-1} \right) \\
&+ (1-\alpha)\left(-U_t^{(2)} V_t^{\frac{-\alpha}{1-\alpha}} \ln V_t^{\frac{-\alpha}{1-\alpha}} - \left(U_t^{(1)} \right)^2 \left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i}) \right]^{-\alpha-1} + \alpha \left(U_t^{(1)} \right)^2 V_t^{\frac{-\alpha-1}{1-\alpha}} \ln V_t^{\frac{-\alpha-1}{1-\alpha}} \right)
\end{aligned} \tag{56}$$

$$\frac{\partial V_t^{(2)}}{\partial \alpha} = \begin{aligned}
&-\left(U_t^{(2)} V_t^{\frac{-\alpha}{1-\alpha}} - \alpha \left(U_t^{(1)} \right)^2 V_t^{\frac{-\alpha-1}{1-\alpha}} \right) \\
&+ (1-\alpha)\left(-U_t^{(2)} V_t^{\frac{-\alpha}{1-\alpha}} \ln V_t^{\frac{-\alpha}{1-\alpha}} - \left(U_t^{(1)} \right)^2 V_t^{\frac{-\alpha-1}{1-\alpha}} + \alpha \left(U_t^{(1)} \right)^2 V_t^{\frac{-\alpha-1}{1-\alpha}} \ln V_t^{\frac{-\alpha-1}{1-\alpha}} \right)
\end{aligned} \tag{57}$$

$$\frac{\partial V_t^{(2)}}{\partial \alpha} = \begin{aligned}
&V_t^{\frac{-\alpha}{1-\alpha}} \left[-U_t^{(2)} + (1-\alpha)(-U_t^{(2)} \ln V_t^{\frac{-\alpha}{1-\alpha}}) \right] + \\
&V_t^{\frac{-\alpha-1}{1-\alpha}} \left[\alpha \left(U_t^{(1)} \right)^2 + (1-\alpha)(-\left(U_t^{(1)} \right)^2 + \alpha \left(U_t^{(1)} \right)^2 \ln V_t^{\frac{-\alpha-1}{1-\alpha}}) \right]
\end{aligned} \tag{58}$$

$$\frac{\partial V_t^{(2)}}{\partial \alpha} = \begin{aligned}
&-U_t^{(2)} V_t^{\frac{-\alpha}{1-\alpha}} \left[1 + (1-\alpha)(\ln V_t^{\frac{-\alpha}{1-\alpha}}) \right] \\
&+ \left(U_t^{(1)} \right)^2 V_t^{\frac{-\alpha-1}{1-\alpha}} \left[\alpha + (1-\alpha)(-1 + \alpha \ln V_t^{\frac{-\alpha-1}{1-\alpha}}) \right]
\end{aligned} \tag{59}$$

3) $\partial f_0 \big/ \partial \alpha$: The demand for futures equation given by equation (3) is differentiated as a function of the parameter α

$$
\begin{aligned}
\frac{\partial f_0}{\partial \alpha} = & \frac{E_0\left[U_t^{(1)}V_t^{-\frac{\alpha}{1-\alpha}}\left(-1+(1-\alpha)\ln V_t^{\frac{1}{1-\alpha}}\right)\right]E_0V_t^{(2)}(F_{t_2}-F_0)}{(E_0V^{(2)}(F_{t_2}-F_0))^2} \\
& \frac{-E_0(V_t^{(1)})E_0\left[(F_{t_2}-F_0)\begin{pmatrix} -U_t^{(1)}V_t^{\frac{-\alpha}{1-\alpha}}\left[1+(1-\alpha)(\ln V_t^{\frac{-\alpha}{1-\alpha}})\right] \\ +\left(U_t^{(2)}\right)^2 V_t^{\frac{-\alpha-1}{1-\alpha}}\left[\alpha+(1-\alpha)(-1+\alpha\ln V_t^{\frac{-\alpha-1}{1-\alpha}})\right] \end{pmatrix}\right]}{(E_0V^{(2)}(F_{t_2}-F_0))^2}
\end{aligned}
\tag{60}
$$

$$
\begin{aligned}
\frac{\partial f_0}{\partial \alpha} = & \frac{E_0\left[U_t^{(1)}V_t^{-\frac{\alpha}{1-\alpha}}\left(-1+\ln V_t\right)\right]E_0V_t^{(2)}(F_{t_2}-F_0)}{(E_0V_t^{(2)}(F_t-F_0))^2} \\
& \frac{-E_0(V_t^{(1)})E_0\left[(F_{t_2}-F_0)\left(-U_t^{(1)}V_t^{\frac{-\alpha}{1-\alpha}}\left[1-\alpha\ln V_t\right]-\left(U_t^{(2)}\right)^2V_t^{\frac{-\alpha-1}{1-\alpha}}\left[1+\alpha(1+\alpha)\ln V_t\right)\right]\right)\right]}{(E_0V_t^{(2)}(F_t-F_0))^2}
\end{aligned}
\tag{61}
$$

$$
\begin{aligned}
\frac{\partial f_0}{\partial \alpha} = & \frac{E_0\left[U_t^{(1)}V_t^{-\frac{\alpha}{1-\alpha}}\left(-1+\ln V_t\right)\right]E_0V_t^{(2)}(F_t-F_0)}{(E_0V_t^{(2)}(F_t-F_0))^2} \\
& \frac{-E_0(V_t^{(1)})E_0\left[(F_t-F_0)V_t^{\frac{-\alpha}{1-\alpha}}\left(-U_t^{(1)}\left[1-\alpha\ln V_t\right]-\left(U_t^{(2)}\right)^2V_t^{\frac{-1}{1-\alpha}}\left[1+\alpha(1+\alpha)\ln V_t\right)\right]\right)\right]}{(E_0V_t^{(2)}(F_t-F_0))^2}
\end{aligned}
\tag{62}
$$

$$
\begin{aligned}
\frac{\partial f_0}{\partial \alpha} = & \frac{E_0\left[U_t^{(1)}V_t^{-\frac{\alpha}{1-\alpha}}\left(-1+\ln V_t\right)\right]E_0V_t^{(2)}(F_t-F_0)}{(E_0V_t^{(2)}(F_t-F_0))^2} \\
& \frac{+E_0(V_t^{(1)})E_0\left[(F_t-F_0)U_t^{(1)}V_t^{\frac{-\alpha}{1-\alpha}}\left(\left[1-\alpha\ln V_t\right]+\frac{1}{U_t^{(1)}}\left(U_t^{(2)}\right)^2V_t^{\frac{-1}{1-\alpha}}\left[1+\alpha(1+\alpha)\ln V_t\right)\right]\right)\right]}{(E_0V_t^{(2)}(F_t-F_0))^2}
\end{aligned}
\tag{63}
$$

Using equation (71) of proposition 4.4

$$\frac{\partial f_0}{\partial \alpha} = \frac{E_0\left[U_t^{(1)} V_t^{-\frac{\alpha}{1-\alpha}} \left(-1 + \ln V_t\right)\right] E_0 V_t^{(2)} (F_t - F_0)}{(E_0 V_t^{(2)} (F_t - F_0))^2}$$
$$\frac{+ E_0 (V_t^{(1)}) E_0 \left[(F_t - F_0) U_t^{(1)} V_t^{\frac{-\alpha}{1-\alpha}} \left(\left[1 - \alpha \ln V_t\right] + U_t^{(1)} \rho_{U_t}^2 V_t^{\frac{-1}{1-\alpha}} \left[1 + \alpha(1+\alpha) \ln V_t\right] \right) \right]}{(E_0 V_t^{(2)} (F_t - F_0))^2} \tag{64}$$

Using equation (68) $\alpha U_t^{(1)} V_t^{-\frac{1}{1-\alpha}} = \rho_{V_t} - \rho_{U_t} \equiv \theta \geq 0$ (proof of proposition 4.4)

$$\frac{\partial f_0}{\partial \alpha} = \frac{1}{\alpha} \frac{E_0 \left[\theta V_t \left(-1 + \ln V_t\right)\right] E_0 V_t^{(2)} (F_t - F_0)}{(E_0 V_t^{(2)} (F_t - F_0))^2}$$
$$\frac{+ E_0 (V_t^{(1)}) E_0 \left[(F_t - F_0) \theta V_t \left((1 + \theta \rho_{U_t}^2) + \alpha \ln V_t \, (-1 + (1+\alpha) \theta \rho_{U_t}^2 \right) \right]}{(E_0 V_t^{(2)} (F_t - F_0))^2} \tag{65}$$

$\frac{\partial f_0}{\partial \alpha} \leq 0$ if the futures contracts' prices are decreasing and $\rho_V^2 \geq \frac{1}{(1+\alpha)\theta}$

Hence proposition 4.1 is proved.

Proposition 4.3: Relation between the elasticity of the marginal rate of substitution through time σ and the parameter α:

$$\sigma = \frac{\partial \ln V_t}{\partial \ln \pi_t} = \frac{1-\alpha}{\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i})} \pi_t U_t^{(1)} \tag{66}$$

Proof of proposition 4.3:

Using the definition of the utility function from equation (13) it is straightforward to find:

$$\sigma = \frac{\partial \ln V_t}{\partial \ln \pi_t} = (1-\alpha) \frac{\pi_t U_t^{(1)}}{\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i})} \tag{67}$$

Proposition 4.4: Relation between the degrees of absolute risk aversion $\rho_{V_t}^A$, $\rho_{U_t}^A$ and the EMRST σ

$$\rho_{V_t}^A = \rho_{U_t}^A + \sigma \frac{\alpha}{1-\alpha}$$

Note that A means absolute

Proof of proposition 4.4:

Using the definition of the utility function from equation (13) it is straightforward to find:

$$V_t^{(1)} = (1-\alpha) U_t^{(1)} \left[\sum_{i=0}^{T_1} \beta^i U(\pi_{t+i}) \right]^{-\alpha} = (1-\alpha) U_t^{(1)} V_t^{\frac{-\alpha}{1-\alpha}} \tag{68}$$

$$V_t^{(2)} = (1-\alpha) \left[U_t^{(2)} V_t^{\frac{-\alpha}{1-\alpha}} - \alpha \left(U_t^{(1)} \right)^2 V_t^{-\frac{1+\alpha}{1-\alpha}} \right] \tag{69}$$

We can derive therefore the risk aversion for the utility function from these two equations as:

$$\rho_{V_t} = -\frac{V_t^{(2)}}{V_t^{(1)}} = -\frac{U_t^{(2)}}{U_t^{(1)}} + \alpha U_t^{(1)} V_t^{-\frac{1}{1-\alpha}}$$

Using equation (70), the conclusion of the demonstration follows. Note that:

$$\Rightarrow \rho_{V_t} = \rho_{U_t} + \alpha U_t^{(1)} V_t^{-\frac{1}{1-\alpha}} \Rightarrow \theta \equiv \rho_{V_t} - \rho_{U_t} \geq 0 \text{ if } \alpha \geq 0 \text{ and } V_t \geq 0$$

$$\alpha U_t^{(1)} V_t^{-\frac{1}{1-\alpha}} = \rho_{V_t} - \rho_{U_t} \qquad V_t = (1-\alpha)\ln(\frac{\rho_{V_t} - \rho_{U_t}}{\alpha U_t^{(1)}}) \tag{71}$$

Minor sets of proofs:

Proof 1: Risk aversion is an increasing function of profits if $V^{(3)} \leq 0$ (prudential motive).

$$\frac{\partial \rho_t}{\partial \pi_t} = -\frac{V_t^{(3)} V_t^{(1)} - V_t^{(2)2}}{V_t^{(1)2}} = -\frac{V_t^{(3)}}{V_t^{(1)}} + \left[-\frac{V_t^{(2)}}{V_t^{(1)}} \right]^2 = -\frac{V_t^{(3)}}{V_t^{(1)}} + \rho_t^2 \neq 0 \qquad (72)$$

Proof 2:

Assume that the utility function V_t defined in equation (13) is positive. The instantaneous utility functions U_t and the utility function V_t are assumed to be concave. The second derivative of the investor's utility function $\left(V_t^{(2)}\right)$ is negative if $0 \leq \alpha \leq 1$. Hence, the parameter α must be bounded between 0 and 1 for the hypothesis that the utility is concave to be true.

$$V_t^{(2)} = (1-\alpha)\left[U_t^{(2)} V_t^{\frac{-\alpha}{1-\alpha}} - \alpha \left(U_t^{(1)}\right)^2 V_t^{-\frac{1+\alpha}{1-\alpha}} \right] \qquad (73)$$

Proof 3:

I show that the trading cost is a function of the investor's expected local risk aversion and the value of profits and approximately the variance in the futures market. If the futures drift is not too great, then the futures price is a good predictor of its value in the future $F_0 \approx E(F_t)$.

$$E_0 \rho_t \left(V_{t,i}^{(1)} (F_t - F_0)^2\right) \approx E_0 \left(\rho_t V_{t,i}^{(1)} (F_t - E_0 F_t)^2\right) \qquad (74)$$

Using the definition of the expected covariance, the equation (75) becomes:

$$\begin{aligned} & E_0 \rho_t \left(V_{t,i}^{(1)} (F_t - F_0)^2\right) \approx \\ & E_0 (\rho_t V_{t,i}^{(1)}) E_0 \left((F_t - E_0 F_t)^2\right) + Cov_0 \left(\rho_t V_{t,i}^{(1)}, \left((F_t - E_0 F_t)^2\right)\right) \end{aligned} \qquad (76)$$

Assuming away the conditional covariance on the left side, equation (77) further simplifies to:

$$E_0 \rho_t \left(V_{t,i}^{(1)} (F_t - F_0)^2\right) \approx E_0 (\rho_t V_{t,i}^{(1)}) \times Var_0 (F_t) \qquad (78)$$

Therefore, the trading cost is a function for the investor's expected local risk aversion and the value of profits and approximately the variance in the futures market.

What does the simplification on the covariance imply?

I define the shock in the futures variance ε_t using the definition of expectations by the following equation $Var_0(F_t) = E_0\left((F_t - E_0F_t)^2\right) = (F_t - E_0F_t)^2 + \varepsilon_t$. Using this definition in equation (79) it becomes:

$$E_0\rho_t\left(V_{t,i}^{(1)}(F_t - F_0)^2\right) \approx E_0(\rho_t V_{t,i}^{(1)})E_0\left((F_t - E_0F_t)^2\right) + Cov_0\left(\rho_t V_{t,i}^{(1)}, \varepsilon_t\right) \qquad (77)$$

Hence removing the covariance is equivalent to assuming that shocks to the variance of futures prices have no impact on risk aversion and marginal utility and consequently on the futures price.

ILLIQUIDITY AND THE GRAPH OF THE IMPLIED VOLATILITY FUNCTIONS

In the previous paper, we have shown that the wealth effect in the futures market is strengthened in the presence of illiquidity. In the following paper, we consider the impact of illiquidity on European put options. Illiquidity in the put option's market can potentially jointly explain the empirical puzzles concerning the graph of the implied volatility as a function of moneyness and as a function of the volume or open interest. Illiquidity proves to be closely related to volatility in the underlying spot price for options close to or at the money. It can therefore be partially controlled by hedging the Vega risk of a change in volatility.

ILLIQUIDITY AND THE GRAPH OF THE IMPLIED VOLATILITY FUNCTION

ABSTRACT

Illiquidity in the put option's market can potentially jointly explain the empirical puzzles concerning the graph of the implied volatility as a function of moneyness and as a function of the volume or open interest. Illiquidity proves to be closely related to volatility in the underlying spot price for options close to or at the money. It can therefore be partially controlled by hedging the Vega risk of a change in volatility. This could explain why Bollen and Whaley (2002) found that abnormal returns from selling put options disappear when controlling for Vega risk. Put options, as well as futures contracts, are priced using the technique of changes in the subjective probability measure normally reserved for risk neutral pricing. In addition, we explore how changes in the hedging period, size of the cash good being hedged and profits change the demand for put options and futures contracts.

INTRODUCTION

The underlying asset's volatility is assumed constant in the Black and Scholes option pricing models and therefore, the implied volatility of put options on the same underlying asset should be the same. However, empirical research has found that this is not the case and that implied volatilities of put options with different exercise prices on the same underlying asset are different. This phenomenon is a puzzle that academics are trying to explain. In addition, Judd and Leisen (2002) report that fixed maturity plots of a call option's open interest across strike prices peaks for the at-the-money contract. The third puzzle, is that time variation in the implied volatility depends also on net buying pressure (Bollen and Whaley (2002)).

Illiquidity in the put option's market can potentially jointly explain these three empirical puzzles concerning the graph of the implied volatility as a function of moneyness and as a function of the volume or open interest. Illiquidity, derived as an endogenous trading cost, proves to be closely related to volatility in the underlying spot price for options close to or at the money. It can therefore be partially controlled by hedging the Vega risk of a change in volatility. This could explain why Bollen and Whaley (2002) found that abnormal returns from selling put options disappear when controlling for Vega risk. Put options, as well as futures contracts, are priced using the technique of changes in the subjective probability measure normally reserved for risk neutral pricing. In addition, we explore how changes in the hedging period, size of the cash good being hedged and profits change the demand for put options and futures contracts. The demands for put options and futures contracts are shown to increase with the investor's wealth and are therefore normal goods. In addition, the demand for put options and futures contracts increase with the number of goods produced and the hedging period.

The paper is organized as follows. In section 1, previous work on explaining the implied volatility functions is presented: In section 2, the model is presented. In section 3, the illiquidity trading costs are derived first for each investor and in the aggregate equilibrium using the market clearing condition that the aggregate supply and demand are equal. In section 4, illiquidity is shown to potentially explain puzzles in option pricing related to the implied volatility surface as a function of volume and moneyness. In section 5, equilibrium-pricing solutions are derived using changes in the probability measure using a variant of the equivalent martingale measure. In section 6, the determinants of the demand for futures and put options are derived.

1. PREVIOUS WORK ON THE IMPLIED VOLATILITY FUNCTIONS

Models of derivatives prices by previous researchers generally assume that markets are complete so that each derivative has a perfect substitute to replace or replicate it. The demand and supply for the assets are therefore perfectly elastic so that buying or selling does not create a pressure on the markets. Markets are defined as liquid by Friend and Blume (1975), as anyone can buy or sell any quantity at no cost. However, in practice, markets are incomplete as there are frictions (credit risk, borrowing limits…) limiting the ability to arbitrage[19]. The derivatives prices are then determined in equilibrium by equating their supply and demand. Trading becomes risky and derivatives markets illiquid.

The study of illiquidity trading costs is part of the derivatives microstructure literature which seeks to explain known limitations of current pricing models. These are summarized in the graph of the implied volatility, which should be a constant function of the strike price and independent of volume. However, the graph of the implied volatility decreases monotonically as a function of the strike price or moneyness. Judd and Leisen (2002) report that fixed maturity plots of a call option's open interest across strike prices peaks for the at-the-money contract. Bollen and Whaley (2002) find that the time variation (term structure) of the implied volatility of an

[19] See Constantinides and Perrakis (2002) for an example of the impact of exogenous transaction costs on the ability to arbitrage. Illiquidity will imply an endogenous transaction cost in this paper.

option series is a function of net buying pressure from public order flow using S&P500 index put options and twenty individual stocks. Net buying pressure is defined as the number of contracts traded at or above the prevailing bid/ask midpoint less the number of contracts traded below the prevailing midpoint. Finally, Bates (2001) studies a market where negative jumps (crashes) can occur. Less crash-averse investors insure the more risk-averse through the options market. Bates finds that heterogeneity amplifies the impact of jumps (crashes). Option prices then overpredict volatility and jump risk due to the wealth effect.

This paper shows that net buying pressure creates illiquidity in the derivatives market in the form of an endogenous illiquidity trading cost that amplifies the wealth effect. This can potentially explain empirical puzzles related to the graph of the implied volatility functions as a function of moneyness, volume or open interest and time structure.

2. MODEL

There are G producers hedging in the derivatives markets and N speculators willing to take on their risk in these markets. The price of the producers' good, the number of producers and hedgers are determined exogenously. It takes t periods before a producer sells his good, while his production costs are incurred at the beginning, time 0. This leaves him with the risk that the price of his good might fall between the time he starts production and the time he sells it. He can hedge this price using futures or insure against a fall in the price of his good using put options. To avoid introducing illiquidity as a risk of reversing a given position and thereby complicating the notion of illiquidity, the option is assumed to have the same maturity as the hedge and the underlying good is assumed liquid. The futures contract margins are ignored and considered an additional source of friction behind the assumption of market incompleteness limiting the ability to replicate assets without risk. Both put options and futures will be in demand as neither instrument can fix in advance the price of the production, due to frictions behind the assumption of market incompleteness, such as the aforementioned marking to market or limits to borrowing.

To hedge the price of the good he produces, the producer can short f_0 futures contracts at the price of F_0. To insure against a fall in the price of his production good, the producer can purchase n_0 long put options at a price of P_0 giving him the right to sell at the strike price K a unit of his good at the maturity t of the option. The producers' profits π_t at time t come from selling y_0 of his goods at a price of S_t at a cost of production of $c(y_0)$. In addition, the producer pays initially the premium $n_0 P_0$ for the put option with a payoff of $n_0 Max(K - S_t, 0)$ at maturity. Hedging with futures is initially cost free for a payoff of $f_0(F_0 - F_t)$ at maturity. The profits to the producer are summarized in equation 1

$$\pi_t = S_t y_0 - c(y_0) + f_0(F_0 - F_t) + n_0 Max(K - S_t, 0), \;\; \pi_0 = -n_0 P_0 \qquad (1)$$

Speculators, defined as having no position in the production good, are willing to be the counterpart of the hedgers in the derivatives markets for a price. Their profits are given by equation (1) with the production part removed. We will refer to producers as hedgers in the text, while the term "investors" stands for speculators and hedgers in the derivatives markets.

Speculators and producers maximize the utility $V(\pi_{t+j}), j = 1,...,\infty$ they derive from their profits. The utility function is assumed increasing and concave as a function of profits and separable preferences. I will use the notation $V_t^{(i)}$ for the ith derivative with respect to profits of the utility function $V(\pi_{t+j}), j = 1,...,\infty$ at time t.

The investor can choose the amount f_0 of futures contracts so as to maximize his utility $V(\pi_{t+j}), j = 1,...,\infty$. Differentiating $V(\pi_{t+j}), j = 1,...,\infty$ with respect to f_0 under the constraint of equation (1) gives the futures pricing equation. Further dividing by the marginal utility at time 0 $V_t^{(0)}$ yields equation (2).

$$E_0 M_t (F_0 - F_t) = 0 \tag{2}$$

The price of the futures contract F_0 is the expected value of its payoff in terms of the marginal rate of substitution $M_t = V_t^{(1)} / V_0^{(1)}$. It represents the investor's risky discount rate and would equal the risk free discount rate in complete markets after a risk neutral transformation. Futures contracts are valuable as they reduce the producer's risk or enhance the speculator's risk, that is they change his risky discount rate.

The futures demand and supply f_0, derived in a previous paper[20] and thereafter referred to as the futures demand, is found by differentiating equation (2) as a function of the futures price F_0 at time 0. Hedgers offer futures contracts to speculators who have the same utility maximization problem, but are long in the future contract and have no position in the underlying asset. The resulting futures demand is given by the following equation (3):

$$f_0 = E_0 V_t^{(1)} / E_0 V_t^{(2)} (F_t - F_0) = E_0 M_t / E_0 \rho_t M_t (F_0 - F_t) \tag{3}$$

The demand for futures contracts by the hedger depends on the expected value of profits at maturity $E_0 M_t$ as well as the expected value of the futures contract on a risk adjusted basis $E_0 \rho_t M_t (F_t - F_0)$, where $\rho_t = -V_t^{(2)} / V_t^{(1)}$ is the degree of risk aversion. Note that the demand for futures contracts depends on the options contract through the impact of profits on the utility and therefore on the marginal rate of substitution $M_t = V_t^{(1)} / V_0^{(1)}$ as well as on the degree of risk aversion $\rho_t = -V_t^{(2)} / V_t^{(1)}$.

[20] Galy (2002)

The investor can choose the amount n_0 of put options so as to maximize his utility $V(\pi_{t+j}), j=1,...,\infty$. Differentiating $V(\pi_{t+j}), j=1,...,\infty$ with respect to n_0 under the constraint of equation (1) yields equation (4).

$$P_0 = E_0 M_t \, Max(K - S_t, 0) \tag{4}$$

The price of the put option P_0 is the expected value of its payoff discounted by the marginal rate of substitution[21] $M_t = V_t^{(1)} \big/ V_0^{(1)}$. Put options are valuable for hedgers as they reduce their risk while speculators enhance their risk.

The demand for put options n_0 is found by differentiating equation (4) as a function of the put options price P_0 at time 0 yielding equation (5).

$$n_0^{-1} = \rho_0 E_0 M_t \, Max(K - S_t, 0) = \rho_0 P_0 \tag{5}$$

The demand for put options is an inverse function of the investor's current degree of risk aversion[22] ρ_0 and the expected value of the payoff. Note that the demand for options contracts depends on the futures contracts through the impact of profits on the utility and therefore on the marginal rate of substitution $M_t = V_t^{(1)} \big/ V_0^{(1)}$ as well as on the current degree of risk aversion $\rho_0 = -V_0^{(2)} \big/ V_0^{(1)}$.

3. ILLIQUIDITY AS AN ENDOGENOUS TRADING COST

In this section, we relax the assumption that investors can buy or sell any quantity of derivatives without additional cost[23]. Equivalently these markets are imperfectly liquid. This

[21] Under the assumption of market completeness used in the arbitrage approach, after a risk neutral transformation, M_t would become the risk free discount rate.

[22] An increase in the degree of risk aversion increases the discounting of the payoff through the marginal rate of substitution, so that its net effect is unclear.

[23] This did not correspond to the case of complete markets, as frictions are still assumed to prevent

changes the investors' problem, as the prices of the derivatives are no longer independent of the size of the trade. In addition, producers must initiate a trade thereby creating a pressure on the derivatives market that will attract speculators, defined here as having no position in the underlying asset. This creates an illiquidity trading cost that was derived for futures contracts in a previous paper[24] and is derived here for put options. The trading cost for futures contracts will change in this model only in that the profit function now has options contracts in addition to the futures contracts. The equations are otherwise not modified.

Now that the futures and put options prices are no longer independent of the quantities that are traded in these markets, the investor's optimization problem endogenously creates an additional illiquidity trading cost. Therefore, the futures and put options demand are given respectively by equations (6) and (7):

$$E_0 V_t^{(1)}(F_t - F_0) + f_0 E_0 V_t^{(1)}\left(\frac{\partial (F_t - F_0)}{\partial f_0}\right) = 0 \tag{6}$$

$$P_0 + n_0\, \partial P_0 / \partial n_0 = E_0 M_t\, (K - S_t, 0) \tag{7}$$

Hedgers pay the illiquidity trading cost in futures contracts $TC_F = f_0 E V_t^{(1)}\left(\partial (F_t - F_0)/\partial f_0\right)$ and in put options contracts $TC_P = n_0\, \partial P_0 / \partial n_0$ as they need to attract speculators to shift the risk of selling their production at time t. These costs would disappear only if the supply was perfectly elastic or the trade was negligible in size.

The slope of the futures contract $\partial (F_t - F_0)/\partial f_0$ and put options $\partial P_0/\partial n_0$ demands in equations (6) and (7) can be determined from equations (2) and (4) where trades were assumed to be too small to matter. The slope of the futures and put options demand are therefore given respectively by,

$$E_0 V_t^{(1)} \frac{\partial (F_t - F_0)}{\partial f_0} = -E_0 \rho_t V_t^{(1)} (F_t - F_0)^2 = E_0 V_t^{(2)} (F_t - F_0)^2 \tag{8}$$

arbitrage.

$$\frac{\partial P_0}{\partial n_0} = -E_0 \rho_t M_t \, Max(K - S_t, 0)^2 - P_0 \,/ n_0 \tag{9}$$

See Proof 1 for more details

The futures and put options prices in the presence of illiquidity are therefore, combining equations (6) with (8) and (7) with (9):

$$E_0 V_t^{(1)} (F_t - F_0) - I_i f_0 \, E_0 \rho_t V_t^{(1)} (F_t - F_0)^2 = 0 \tag{10}$$

$$\begin{aligned} &P_0 \; - I_i n_{0,i} E_0 M_{t,i} \left[\rho_{t,i} Max(K - S_t, 0)^2 + P_0 \rho_{0,i} Max(K - S_t, 0) \right] \\ &= E_0 M_{t,i} Max(K - S_t, 0) \end{aligned} \tag{11}$$

$I_i = 1$ if i is a producer (hedger) and 0 otherwise (speculator)

Now that we have derived the pricing equation for individual investors with hedgers paying an illiquidity trading cost to speculators, we must find the equilibrium price for the futures and put options markets by imposing the market clearing conditions for these markets $\sum_{i=1}^{N+G} f_{0,i} = 0$ and $\sum_{i=1}^{N+G} n_{0,i} = 0$, where G is the number of hedgers and N the number of speculators.

The prices of the futures and put options contracts are therefore given by equations (10) and (11) in equilibrium. The futures value is pushed down by the illiquidity trading cost, as hedgers must short at a lower price in an illiquid futures market to attract speculators. The put option's price increases when that market is illiquid as speculators ask hedgers for a higher premium to cover the risk of a fall in the price of the underlying good.

$$E_0 \left(\sum_{i=1}^{N+G} V_{t,i}^{(1)} \right) (F_t - F_0) - E_0 \sum_{i=1}^{N+G} f_{0,i} E_0 \rho_{t,i} \left(V_{t,i}^{(1)} (F_t - F_0)^2 \right) = 0 \tag{12}$$

$$P_0 \; - \sum_{i=1}^{G} n_{0,i} E_0 \frac{M_{t,i}}{N} \left[\rho_{t,i} Max(K - S_t, 0)^2 \right] = \sum_{i=1}^{N+G} E_0 \frac{M_{t,i}}{N} Max(K - S_t, 0) \tag{13}$$

The condition that speculators have no exposure to the underlying spot good is not imposed directly in the equation (13). Under a risk neutral transformation (assuming market completeness)

[24] Galy (2002)

one would transform the left side of equations (12) and (13) into a discrete version of the Black and Scholes formula for futures and put options. However, these prices would ignore the illiquidity trading costs. For futures contracts, the illiquidity trading cost diminishes the futures value linearly with the futures demand at approximately the square of the futures price variance. For put options contracts, the illiquidity trading cost increases the option's price linearly with the put options demand at approximately the square of the moneyness measured by $(1 - S_t / K)$. Hence, far in the money puts will be much more expensive[25] than in the absence of an illiquidity trading cost in incomplete markets $P_0 \ = \sum_{i=1}^{N+G} E_0 \frac{M_{t,i}}{N} Max(K - S_t, 0)$ or under a discrete version of the Black and Scholes formula in complete markets $P_0 \ = 1/(1 + r_t)\, E_0^Q Max(K - S_t, 0)$. In proof 2, we show that the risk neutral probability Q is used to transform the risky discount of the marginal rate of substitution M_t into a risk free one $1/(1 + r_t)$.

4. ILLIQUIDITY AS AN EXPLANATION OF SOME PRICING PUZZLES

The underlying asset's volatility is assumed constant in the Black and Scholes option pricing models and therefore, the implied volatility of put options on the same underlying asset should be the same. However, empirical research has found that this is not the case and that implied volatilities of put options with different exercise prices on the same underlying asset are different. This phenomenon is a puzzle that academics are trying to explain. In addition, Judd and Leisen (2002) report that fixed maturity plots of a call option's open interest across strike prices peaks for the at-the-money contract. The third puzzle, is that time variation in the implied volatility depends also on net buying pressure (Bollen and Whaley (2002)).

[25] Investors will use derivatives as long as the benefit from using them exceeds their cost (proof 9).

Illiquidity in the put option's market can potentially jointly explain these three empirical puzzles concerning the graph of the implied volatility as a function of moneyness and as a function of the volume or open interest. Illiquidity, derived as an endogenous trading cost, proves to be closely related to volatility in the underlying spot price for options close to or at the money (see proof 8). It can therefore be partially controlled by hedging the Vega risk of a change in volatility. This could explain why Bollen and Whaley (2002) found that abnormal returns from selling put options disappear when controlling for Vega risk. While these results are generally derived for indexes, the model's concept of hedgers covering a future price risk applies also to indexes, as these are also used to manage the risk of an uncertainty in the price of an underlying asset.

The first puzzle, concerning the graph of the implied volatility as a function of moneyness, can potentially be solved in this model. Illiquidity may explain why the graph of the implied volatility is related to net buying pressure. Since the implied volatility is determined using observed put options prices from markets that are not liquid, it will include an illiquidity premium. This implies higher prices than under Black and Scholes and therefore a Black-Scholes implied volatility that increases approximately at the square of the moneyness $\sum_{i=1}^{G} n_{0,i} E_0 \left(M_{t,i} / N \right) \left[\rho_{t,i} Max(K - S_t, 0)^2 \right]$ in equation (13) of our model. Hence, illiquidity would create the nonlinear graph of the implied volatility observed in the data.

The second puzzle concerning the graph of the implied volatility as a function of the put options demand can also potentially be solved in this model. First, let us show that a plot of call option open interest across strikes would peak for the at-the-money contract. If the option is close to or at the money $S_0 \approx K$, then the illiquidity trading cost in equation (13) $TC_p = \sum_{i=1}^{G} n_{0,i} E_0 \left(M_{t,i} / N \right) \left[\rho_{t,i} Max(\underbrace{K - S_0}_{\approx 0}, 0)^2 \right] \approx 0$ will be close to zero, replacing $E_0 S_t$ with S_0. Therefore, the option is the most liquid when the option is close to or at the money as the trading cost approximately disappears, making it more attractive to investors. Hence, the demand for put options should peak close to or at the money, as put options are the less costly in terms of illiquidity trading costs.

Second, let us show that the graph of the implied volatility, in particular its time variation, is directly related to net buying pressure from hedgers. The higher the net buying pressure from hedgers $\sum_{i=1}^{G} n_{0,i}$, the higher the illiquidity trading cost. If most put options are close to being at the money $S_0 \approx K$, then the illiquidity trading cost is approximately related to the variance in the underlying asset at the maturity of the option, assuming again that the underlying asset follows a random walk $E_0 S_t = S_0$, $TC_P = \sum_{i=1}^{G} n_{0,i} E_0(M_{t,i} / N) \rho_{t,i} Var_0(S_t)$, where the conditional variance of the underlying asset price is $Var_0(S_t)$ (see proof 8). Hence, the expected variance in the spot market $Var_0(S_t)$ increases in importance for pricing, the larger the net buying pressure from hedgers. This may potentially explain the result of Bollen and Whaley (2002) that hedging pressure impacts future volatility. This suggests also that controlling for changes in the volatility of the underlying asset, called Vega risk, is a way to control[26] illiquidity for put options close to or at the money.

5. PRICING DERIVATIVES IN THE PRESENCE OF ILLIQUIDITY

The equations (12) and (13) which price futures and put options contracts in the presence of illiquidity depend on the choice of an unobserved utility function limiting their value for practical applications. In this section, we show how equilibrium-pricing solutions for the derivatives are derived using a variant of the equivalent martingale measure and can be more

[26] Control in the sense of bringing prices closer to their arbitrage or risk neutral values.

easily applied. In practice, we would need only to specify the path of the underlying asset with the appropriate mean depending on the change of probability measure.

5.1 PRICING FUTURES IN THE PRESENCE OF ILLIQUIDITY

The equation (12) which prices futures contracts in the presence of illiquidity can be transformed using changes of the subjective probability measure $P(s)$. We will define the probabilities $Q(s)$ and $W(s)$ as transformations of the investor's subjective probability $P(s)$ of the event s.

Dividing equation (12) by the marginal value of a dollar at time 0, we find that equation (12) can be rewritten to introduce the marginal rate of substitution through time.

$$E_0 \sum_{i=1}^{N+G} M_{t,i}(F_t - F_0) - \sum_{i=1}^{G} f_{0,i} E_0 M_{t,i} \rho_t (F_t - F_0)^2 = 0 \qquad (14)$$

The probability $Q(s) = P(s) M_t(s) / E_0 M_t$ is called the risk neutral probability as payoffs are discounted at the risk free rate under this probability (see proof 7). Under this risk neutral transformation, the future's expected payoff is discounted at the risk free rate. This transformation is independent of the utility function that is used as the marginal rate of substitution disappears from the pricing equation. In incomplete markets, this can only be considered an approximation. However, this has the advantage of having the illiquidity trading cost as a premium to the risk neutral or arbitrage price whose computation is well understood.

The utility function can be removed entirely from equation (12) which prices the futures contract by changing the subjective probability $P(s)$ of the event s by the probability $W(s) = P(s)\rho_{,it}(s) M_{t,i}(s)(F_t - F_0) / E_0 \rho_{,it} M_{t,i}(F_t - F_0)$. Under this transformation, the illiquidity trading cost equals the present value of the futures contract (see proof 2). This can be verified by multiplying and dividing by the inverse of the futures demand $f_0 = -E_0 M_t / E_0 \rho_t M_t (F_0 - F_t)$ into the illiquidity trading cost of the equation (12) which prices futures contracts and using the marginal condition $1/1+r_t = E_0 M_{it}$ (see proof 6). Under this double change of probability measure, the futures pricing equation (9) becomes independent of the choice of utility.

$$F_0 = (1/N)\left[\sum_{i=1}^{N+G} E_0^Q(F_t) + \sum_{i=1}^{G} E_0^W(F_t)\right] \tag{15}$$

See Proof 2 in the Annex

Where Q and W are two probability measures defined in proof 2, N the number of speculators and G the number of hedgers. The futures price is a weighted average of the expected futures prices. The illiquidity trading cost takes the form here of a downward pressure, under the probability measure W, on the expected futures price as hedgers seek to shift their risk by selling futures contracts. The equation could price futures in practice by assuming a path for the stock price at maturity and finding the mean corresponding to each probability measure change.

5.2 PRICING PUT OPTIONS IN THE PRESENCE OF ILLIQUIDITY

The equation (13) which prices put options contracts in the presence of illiquidity can be transformed using a change of probability measure. Let $P(s)$ be the subjective probability of the event s used by investors, while $Q(s)$ and $X(s)$ are other probabilities of the same event[27].

The probability $Q(s) = P(s)M_t(s)/E_0 M_t$ is the classic risk neutral probability as can be easily verified (see proof 7). Under this risk neutral transformation, the option's expected payoff is discounted at the risk free rate and this irrespective of the agent's utility function. In incomplete markets, this can only be considered an approximation. However, this has the advantage of having illiquidity as a premium over the risk neutral or arbitrage price whose computation is well understood.

[27] These probabilities will be chosen as transformations of the subjective probability $P(s)$.

The utility function can be removed entirely from the option's pricing equation (13), which prices the option's contract by changing the subjective probability $P(s)$ to the probability $X(s) = P(s)\rho_{0,i}(s)M_{t,i}(s)Max(K - S_T, 0) \big/ E_0\rho_{0,i}M_{t,i}Max(K - S_T, 0)$. Under this double change of probability measure, the option's pricing equation (12) becomes

$$P_0 = (1/N)\left[\sum_{i=1}^{G} E_0^X\left[(\rho_{t,i}/\rho_{0,i})Max(K - S_t, 0)\right] + (1/(1 + r_t))\sum_{i=1}^{N+G} E_0^Q Max(K - S_t, 0)\right] \quad (16)$$

See Proof 3 in the Annex.

The option's expected payoff under the probability measure $Q(s)$ is discounted at the risk free rate, while the illiquidity trading cost equals the option's expected payoff under the probability $X(s)$ multiplied by the change in the degree of risk aversion. Intuitively, the more fearful hedgers are of a crash (measured by the degree of risk aversion at maturity ρ_t), the more illiquid and expensive put options will be and this beyond the wealth effect[28] assumed away by making a risk neutral transformation of the payoff. Speculators are willing to bear the hedger's risks only if they receive an additional compensation over the option's expected payoff.

The option's pricing equation (16) could be further simplified and solved using the decomposition of an expected option payoff, assuming a diffusion process for the spot good, and making an assumption on the joint distribution of degree of risk aversion $(\rho_{t,i}/\rho_{0,i})$ with the option's payoff $Max(K - S_t, 0)$ (see Garcia and Renault (1998) for an example of this approach).

[28] See Galy (2002) for a discussion of the wealth effect and its relation with illiquidity in the futures market

6. DETERMINANTS OF THE DEMAND FOR FUTURES AND PUT OPTIONS

In this section, we look at the determinants of the demand for futures and put options. We find that futures and put options are normal goods, whose demand increases with production as long as producing is profitable. The impact of the hedging period increases the demand for put options if profits are expected to increase.

Proposition 1 shows that futures contracts are normal goods whose demand increases with production and the hedging period as long as it is profitable.

Proposition 1: Determinants of the demand for futures contracts

1. $sign(\partial f_0/\partial t) \; = -sign(F_t - F_0)sign\left(\partial\pi/\partial t\right)$

2. $\dfrac{\partial f_0}{\partial y} \geq 0 \;\; if \; F_t - F_0 \leq 0 \; and \; \pi_y \leq 0$

   $\dfrac{\partial f_0}{\partial y} \leq 0 \;\; if \; F_t - F_0 \leq 0 \; and \; \pi_y \geq 0$

3. $sign\dfrac{\partial f_0}{\partial \pi_t} \; = -sign(F_t - F_0)$

Hedgers use futures the longer the hedging period as long as profits are expected to increase. In addition, the use of futures contracts is an increasing function of the production of the cash good being hedged if it is an economic liability $(\pi_y \leq 0)$. It is a decreasing function when the cash good is an economic asset $(\pi_y \geq 0)$. These results hold only when the futures prices are declining, that is in contango. Finally, futures contracts are a normal good as their demand increases and decreases with the investors' wealth as measured by profits.

Proposition 2 shows that put options are also normal goods under some weak conditions. They are more in demand as production increases as long as it is profitable. Finally, the demand for put options increases with the hedging period.

Proposition 2: Determinants of the demand for put options

1. $\partial n_0 / \partial t \geq 0$ if proposition 2.3 holds
2. $\partial n_0 / \partial y_0 \geq 0$ if proposition 2.3 holds and $p \geq C_y$
3. $\dfrac{\partial n_0}{\partial \pi_0} \geq 0$ if $0 \leq \partial \rho_0 / \partial \pi_0 \leq \rho_0 (\rho_t - \rho_0)$ and $V_t^{(3)} \leq 0$

The put options demand changes with the hedging period, the investor's profits and the size of the cash good being hedged. The option's demand increases as the investor's hedging horizon or the option's time to expiration increases.

In addition, the put options demand increases with the investor's wealth as measured by his profits if he is prudent $V_t^{(3)} \leq 0$. The degree of prudence $p_{t,i} = -V_{t,i}^{(3)} / V_{t,i}^{(2)}$, sometimes called precaution, measures the investor's willingness to bear risk as his profits or wealth changes (see proof 1). Therefore, put options are a normal good for the investor. Equivalently, the investor's degree of risk aversion increases with his wealth (or equivalently his profits). The increase in his degree of risk aversion must however be bounded by a function of his current degree of risk aversion and his expected degree of risk aversion at the maturity of the option's contract. Should the investor expect a crash, he will demand more put options, as his expected degree of risk aversion in the future will grow with his expected loss in wealth[29]. This same ratio of degree of risk aversion to current degree of risk aversion is present in the liquidity cost of equation (15).

[29] The demand for put options was shown to increase (decrease) with increases (decreases) in the investor's wealth (not future wealth)

The higher the demand to hedge against a crash, the more put options will be asked for and the more expensive they will become as the put options market becomes illiquid.

Finally, the demand for put options increases as long as profits increase and the good produced is still profitable.

CONCLUSION

Illiquidity can potentially explain the graph of the implied volatility as a function of moneyness and as a function of volume or open interest. Illiquidity is closely related to volatility in the underlying spot price for options close to or at the money. It can therefore be partially controlled by hedging the Vega risk of a change in volatility. This could explain why Bollen and Whaley (2002) found that abnormal returns from selling put options disappear when controlling for Vega risk. Futures contracts and put options can be priced in equilibrium in illiquid markets using transformations of the subjective probability measures. In addition, the demand for futures contracts and put options increases as the hedging period, profits and size of production increases.

ANNEX

Proposition 1: Determinants of the demand for futures contracts

1. $sign(\partial f_0/\partial t) = -sign(F_t - F_0)sign(\partial\pi/\partial t)$

2. $\dfrac{\partial f_0}{\partial y} \geq 0 \;\; if \;\; F_t - F_0 \leq 0 \; and \; \pi_y \leq 0 \qquad \dfrac{\partial f_0}{\partial y} \leq 0 \;\; if \;\; F_t - F_0 \leq 0 \; and \; \pi_y \geq 0$

3. $sign\dfrac{\partial f_0}{\partial \pi_t} = -sign(F_t - F_0)$

Proposition 1.1: Hedging period

$sign(\partial f_0/\partial t) = -sign(F_t - F_0)sign(\partial\pi/\partial t)$

Proof of proposition 1.1

Differentiating equation (3) as a function of π_t:

$$\frac{\partial f_0}{\partial t} = \frac{\partial f_0}{\partial \pi_t}\frac{\partial \pi_t}{\partial t} \tag{17}$$

From proposition 1.3, therefore proposition 1.1 holds.

Proof of proposition 1.2

Differentiating equation (3) as a function of y and using equation (3):

$$\frac{\partial f_0}{\partial y_t} = \frac{E_0(V_t^{(2)}\pi_{y_t}) - f_0 E_0\left[V_t^{(3)}\pi_{y_t}(F_t - F_0)\right]}{E_0 V_t^{(2)}(F_t - F_0)} \tag{18}$$

Noting from proof 5 that $sign\{F_t - F_0\} = sign(V_t^{(3)})$, it follows that proposition 1.2 holds.

Proof of proposition 1.3

Differentiating equation (3) as a function of π_t:

$$\frac{\partial f_0}{\partial \pi_t} = \frac{E_0 V_t^{(2)} E_0 V_t^{(2)} (F_t - F_0) - E_0 V_t^{(1)} E_0 V_t^{(3)} (F_t - F_0)}{\left(E_0 V_t^{(2)} (F_t - F_0)\right)^2} \tag{19}$$

Using equation (3) to simplify

$$\frac{\partial f_0}{\partial \pi_t} = \frac{E_0 (V_t^{(2)}) - f_0 E_0 \left[V_t^{(3)} (F_t - F_0)\right]}{E_0 V_t^{(2)} (F_t - F_0)} \tag{20}$$

Using equation (34), the expression further simplifies to:

$$\frac{\partial f_0}{\partial \pi_t} = \frac{E_0 (V_t^{(2)}) - 2 E_0 (V_t^{(2)})}{E_0 V_t^{(2)} (F_t - F_0)} = - \frac{E_0 V_t^{(2)}}{E_0 V_t^{(2)} (F_t - F_0)} \tag{21}$$

The utility being concave, it follows that proposition 1.3 holds.

Proposition 2: Determinants of the demand for put options

1. $\partial n_0 / \partial t \geq 0$ if proposition 2.3 holds

2. $\partial n_0 / \partial y_0 \geq 0$ if proposition 2.3 holds and $p \geq C_y$

3. $\frac{\partial n_0}{\partial \pi_0} \geq 0$ if $0 \leq \partial \rho_0 / \partial \pi_0 \leq \rho_0 (\rho_t - \rho_0)$ and $V_t^{(3)} \leq 0$

Proof of proposition 2.1:

$$\frac{\partial n_0}{\partial t} = \frac{\partial n_0}{\partial \pi_t} \frac{\partial \pi_t}{\partial t} \tag{22}$$

From proposition 2.3, this implies that the $sign(\partial n_0 / \partial t) =$ Proposition 2.3 holds and $sign(\partial \pi / \partial t)$. If profits are increasing and proposition 1 holds then the demand for put options increases with the hedging period. Therefore proposition 2.1 holds.

Proof of proposition 2.2:

Using the compound rule for a derivative: $\frac{\partial n_0}{\partial y_0} = \frac{\partial n_0}{\partial \pi_0}\frac{\partial \pi_0}{\partial y_0} = \frac{\partial n_0}{\partial \pi_0}(p - C_y)$

Hence, the demand for derivatives increases as a function of the size of the good being hedged as long as it is marginally profitable and proposition 2.3 holds.

Proof of proposition 2.3:

Differentiating the option's demand (5) as a function of the profits,

$$\frac{\partial n_0}{\partial \pi_o} = -\frac{\partial\left[E_0 M_T (K - S_T)^+ \rho_0\right]\big/\partial \pi_0}{\left[E_o M_T (K - S_T)^+ \rho_0\right]^2} \tag{23}$$

Let us derive the numerator separately:

$$\partial\left[E_0 M_T (K - S_T)^+ \rho_0\right]\big/\partial \pi_0 = \partial\left[E_0 \frac{-V_0^{(2)} V_T^{(1)}}{V_0^{(1)2}}(K - S_T)^+\right]\Bigg/\partial \pi_0 \tag{24}$$

$$\partial\left[E_0 M_T (K - S_T)^+ \rho_0\right]\big/\partial \pi_0 = E_0\left(\partial\left[\frac{-V_0^{(2)} V_T^{(1)}}{V_0^{(1)2}}\right]\Bigg/\partial \pi_0\right)(K - S_T)^+ \tag{25}$$

$$\partial\left[E_0 M_T (K - S_T)^+ \rho_0\right]\big/\partial \pi_0 = \\ -E_0\left(\frac{(V_0^{(3)} V_T^{(1)} + V_0^{(2)} V_T^{(2)}) V_0^{(1)2} - 2V_0^{(2)} V_0^{(1)} (V_0^{(2)} V_T^{(1)})}{V_0^{(1)4}}\right)(K - S_T)^+ \tag{26}$$

$$\partial\left[E_0 M_T (K - S_T)^+ \rho_0\right]\big/\partial \pi_0 = -E_0\left(\frac{V_0^{(3)} V_T^{(1)} + V_0^{(2)} V_T^{(2)}}{V_0^{(1)2}} - 2\frac{V_0^{(2)2} V_T^{(1)})}{V_0^{(1)2} V_0^{(1)}}\right)(K - S_T)^+ \tag{27}$$

$$\partial\left[E_0 M_T (K - S_T)^+ \rho_0\right]\big/\partial \pi_0 = -E_t\left(\frac{V_0^{(3)} V_T^{(1)}}{V_0^{(1)2}} + \rho_0 \rho_T M_T - 2\rho_0^2 M_{t,T}\right)(K - S_T)^+ \tag{28}$$

$$\partial\left[E_0 M_T (K - S_T)^+ \rho_0\right]\big/\partial \pi_0 = \\ -E_t\left(\frac{-V_0^{(3)}}{V_0^{(2)}}\frac{-V_0^{(2)}}{V_0^{(1)}}\frac{V_T^{(1)}}{V_0^{(1)}} + \rho_0 \rho_T M_T - 2\rho_0^2 M_T\right)(K - S_T)^+ \tag{29}$$

$$\partial\left[E_0 M_T (K-S_T)^+ \rho_0\right]\big/\partial\pi_0 = -E_t\left(\upsilon_0 \rho_{0t} M_T + \rho_0 \rho_T M_T - 2\rho_0^2 M_{t,T}\right)(K-S_T)^+ \tag{30}$$

$$\partial\left[E_0 M_T (K-S_T)^+ \rho_0\right]\big/\partial\pi_0 = -E_0 M_T (K-S_T)^+ \rho_0 \left(\upsilon_0 + \rho_T - 2\rho_0\right) \tag{31}$$

Where, $\upsilon_0 = -V_0^{(3)} \big/ V_0^{(2)}$ represents the prudence which controls how risk aversion changes with wealth (see equation (51) below). Introducing the result of equation (49) into equation (41), we find that:

$$\frac{\partial n_0}{\partial \pi_0} = \frac{E_0 M_T (K-S_T)^+ \rho_0 (\upsilon_0 + \rho_T - 2\rho_0)}{\left[E_0 M_T (K-S_T)^+ \rho_0\right]^2} \tag{32}$$

The elements of equation (50) are all positive with the potential exception of $(\upsilon_0 + \rho_T - 2\rho_0)$ whose sign must be determined.

$\upsilon_0 + \rho_T - 2\rho_0 \geq 0$ (true if $V_0^{(3)}$ is sufficiently negative)

if $\upsilon_0 - \rho_0 \geq \rho_0 - \rho_T$

$\rho_0(\upsilon_0 - \rho_0) \geq \rho_0(\rho_0 - \rho_T)$

$\rho_0(-\upsilon_0 + \rho_0) \leq -\rho_0(\rho_0 - \rho_T)$

$\partial\rho_0 \big/ \partial\pi_0 \leq \rho_0(\rho_T - \rho_0)$

where, $\partial\rho_0 \big/ \partial\pi_0 = \rho_0(-\upsilon_0 + \rho_0) = -\dfrac{V_0^{(3)}}{V_0^{(1)}} + \rho_0^2 \geq 0$

$0 \leq \partial\rho_0 \big/ \partial\pi_0 \leq \rho_0(\rho_T - \rho_0)$ if $V_0^{(3)} \leq 0$

Hence, $\dfrac{\partial n_0}{\partial \pi_0} \geq 0$ if $0 \leq \partial\rho_0 \big/ \partial\pi_0 \leq \rho_0(\rho_T - \rho_0)$ and $V_0^{(3)} \leq 0$

The demand for put options increases when the investor is risk intolerant $V_0^{(3)} \leq 0$ in that his degree of risk aversion increases with his profits:

$$\partial\rho_t \big/ \partial\pi_t = -V_t^{(3)} \big/ V_t^{(1)} + \rho_t^2 \geq 0 \quad if \quad V_t^{(3)} \leq 0 \tag{33}$$

This increase is relatively slow for reasonable values of the degree of risk aversion, assuming no major crash. A major crash would push the investor's degree of risk aversion at the maturity of the option very high.

Proof 1: Illiquidity trading cost for put options

Differentiating equation (4) as a function of the number of put options traded, we find that:

$$\frac{\partial P_0}{\partial n_0} = E_0 \frac{V_t^{(2)} Max(K - S_T, 0) V_0^{(1)} - V_0^{(2)} (-P_0) V_t^{(1)}}{V_0^{(1)2}} Max(K - S_T, 0) \tag{34}$$

$$\frac{\partial P_0}{\partial n_0} = E_0 \frac{V_t^{(2)}}{V_t^{(1)}} \frac{V_t^{(1)}}{V_0^{(1)}} Max(K - S_T, 0)^2 + P_0 \frac{V_0^{(2)}}{V_0^{(1)}} E_0 \frac{V_t^{(1)}}{V_0^{(1)}} Max(K - S_T, 0) \tag{35}$$

Using the definitions of the marginal rate of substitution $M_t = V_t^{(1)} \big/ V_0^{(1)}$ and degree of risk aversion $\rho_t = - V_t^{(2)} \big/ V_t^{(1)}$, we further simplify the equation to:

$$\frac{\partial P_0}{\partial n_0} = -E_0 \rho_t M_t \, Max(K - S_T, 0)^2 - P_0 \, \rho_0 E_0 M_t \, Max(K - S_T, 0) \tag{36}$$

Using the option's demand of equation (5), the expression further simplifies to:

$$\frac{\partial P_0}{\partial n_0} = -E_0 \rho_t M_t \, Max(K - S_T, 0)^2 - P_0 \big/ n_0 \tag{9}$$

The illiquidity trading cost is therefore:

$$TC_P = n_0 \, \partial P_0 \big/ \partial n_0 = -n_0 E_0 \rho_t M_t \, Max(K - S_T, 0)^2 - P_0 \tag{37}$$

Proof 2: Pricing Futures Contracts in the Presence of Illiquidity

Multiplying and dividing the illiquidity trading cost in equation (12) by the futures demand of equation (3) $f_0 = - E_0 M_t \big/ E_0 \rho_t M_t \, (F_t - F_0)$, and multiplying and dividing the expected value of the futures contract by the expected marginal rate of substitution, equation (12) which prices futures contracts in the presence of illiquidity becomes:

$$\begin{aligned} &\sum_{i=1}^{N+G} (E_0 M_{t,i}) E_0 \frac{M_{t,i}}{E_0 M_{t,i}} (F_t - F_0) \\ &+ \sum_{i=1}^{G} f_{0,i} \frac{E_0 M_{t,i}}{f_{0,i}} E_0 \frac{M_{t,i} \rho_{t,i} (F_t - F_0)}{E_0 M_{t,i} \rho_{t,i} (F_t - F_0)} (F_t - F_0) = 0 \end{aligned} \tag{38}$$

$Q(s) = P(s)M_{t,i}(s) \big/ E_0 M_{t,i}(s)$ is the classical risk neutral probability measure used to change risky discounts into risk free ones, while $W(s) = P(s)\rho_{,it}M_{t,i}(F_t - F_0) \big/ E_0\rho_{,it}M_{t,i}(F_t - F_0)$ is a probability measure, as:

$$\sum_{s\in\Omega} W(s) = \sum_{s\in\Omega} \frac{P(s)\rho_{,it}M_{t,i}(F_t - F_0)}{E_0\rho_{,it}M_{t,i}(F_t - F_0)} = \frac{\sum\limits_{s\in\Omega} P(s)\rho_{,it}M_{t,i}(F_t - F_0)}{E_0\rho_{,it}M_{t,i}(F_t - F_0)}$$

$$= \frac{E_0\rho_{,it}M_{t,i}(F_t - F_0)}{E_0\rho_{,it}M_{t,i}(F_t - F_0)} = 1$$

where possible events are noted by s and belong to the set of all possible events Ω.

$W(s) = \dfrac{P(s)\rho_{,it}M_{t,i}(F_t - F_0)}{E_0\rho_{,it}M_{t,i}(F_t - F_0)} \geq 0$ as the marginal rate of substitution, degree of risk aversion and the payoff are positive. Hence, $W(s)$ can be defined as a probability.

The equation (22) can be further simplified using the usual marginal condition $E_0 M_{t,i} = 1/(1+r_t)$ (see proof 6) so that equation (22) becomes:

$$\sum_{i=1}^{N+G} \frac{1}{1+r_t} E_0^Q (F_t - F_0) + \sum_{i=1}^{G} \frac{1}{1+r_t} E_0^W (F_t - F_0) = 0 \tag{39}$$

Equation (23) can be further simplified as:

$$\sum_{i=1}^{N+G} (E_0^Q (F_t) - F_0) + \sum_{i=1}^{G} (E_0^W (F_t) - F_0) = 0 \tag{40}$$

Therefore, the futures price in the presence of illiquidity is given by the following equation:

$$F_0 = (1/N)\left[\sum_{i=1}^{N+G} E_0^Q (F_t) + \sum_{i=1}^{G} E_0^W (F_t)\right] \tag{15}$$

Proof 3: Pricing Put options in the Presence of Illiquidity

Multiplying and dividing by the demand $n_0^{-1} = E_0 \rho_0 M_t \, Max(K - S_T, 0)$ the illiquidity trading cost in equation (13), and multiplying and dividing the expected payoff by the expected marginal rate of substitution $E_0 M_{t,i}$, we find that:

$$P_0 \; - (1/N) \sum_{i=1}^{G} n_{0,i} n_{0,i}^{-1} E_0 \, \frac{\rho_{0,i} M_{t,i} Max(K - S_t, 0)}{E_0 \rho_{0,i} M_{t,i} Max(K - S_t, 0)} \left[\frac{\rho_{t,i}}{\rho_{0,i}} Max(K - S_t, 0) \right] =$$
$$(1/N) \sum_{i=1}^{N+G} E_0 M_{t,i} E_0 \, \frac{M_{t,i}}{E_0 M_{t,i}} Max(K - S_t, 0) \qquad (41)$$

This equation (25) can be further simplified using the equation $E_0 M_{t,i} = 1/(1 + r_t)$ (see proof 6) so that the option's pricing equation becomes:

$$P_0 \; = (1/N) \sum_{i=1}^{G} E_0^X \left[\frac{\rho_{t,i}}{\rho_{0,i}} (K - S_t)^+ \right] + (1/N) \sum_{i=1}^{N+G} (1 + r_t \,)^{-1} E_0^Q (K - S_t)^+ \qquad (42)$$

Is $X(s) = \dfrac{P(s) \rho_{0,i} M_{t,i} (K - S_T)^+}{E_0 \rho_{0,i} M_{t,i} (K - S_T)^+}$ a probability?

Summing the probability as a function of all the events s possible in the set of all possible events Ω, we find that:

$$\sum_{s \in \Omega} X(s) = \sum_{s \in \Omega} \frac{P(s) \rho_{0,i} M_{t,i} (K - S_T)^+}{E_0 \rho_{0,i} M_{t,i} (K - S_T)^+} = \frac{\sum_{s \in \Omega} P(s) \rho_{0,i} M_{t,i} (K - S_T)^+}{E_0 \rho_{0,i} M_{t,i} (K - S_T)^+}$$
$$= \frac{E_0 \rho_{0,i} M_{t,i} (K - S_T)^+}{E_0 \rho_{0,i} M_{t,i} (K - S_T)^+} = 1$$

$X(s) = \dfrac{P(s) \rho_{0,i} M_{t,i} (K - S_T)^+}{E_0 \rho_{0,i} M_{t,i} (K - S_T)^+} \geq 0$ as the marginal rate of substitution, degree of risk aversion and the payoff are positive. Hence, $X(s)$ can be defined as a probability.

Proof 4: Degree of risk aversion is an increasing function of profits if $V^{(3)} \leq 0$.

$$\frac{\partial \rho_t}{\partial \pi_t} = -\frac{V_t^{(3)}V_t^{(1)} - V_t^{(2)2}}{V_t^{(1)2}} = -\frac{V_t^{(3)}}{V_t^{(1)}} + \left[-\frac{V_t^{(2)}}{V_t^{(1)}}\right]^2 = -\frac{V_t^{(3)}}{V_t^{(1)}} + \rho_t^2 \neq 0 \tag{43}$$

Proof 5:

$F_t - F_0 \leq 0 \Rightarrow V_t^{(3)} \leq 0$ and $f_0 \geq 0$

$F_t - F_0 \geq 0 \Rightarrow V_t^{(3)} \geq 0$ and $f_0 \leq 0$

Differentiate equation (2) as a function of F_t twice to obtain equation (44) and then (45):

$$f_0 EV_t^{(2)}(F_t - F_0) - EV_t^{(1)} = 0 \tag{44}$$

$$f_0 EV_t^{(2)} + f_0 EV_t^{(3)}(-f_0)(F_t - F_0) + EV_t^{(2)} f_0 = 0 \tag{45}$$

Equation (45) further simplifies to

$$E_0 V_t^{(3)}(F_t - F_0) = \frac{2E_0 V_t^{(2)}}{f_0} \tag{46}$$

$$sign\{E_0 V_t^{(3)}(F_t - F_0)\} = -\,sign\{f_0\}$$

From equation (3), we know that sign $\{f_0\} = -sign\{F_t - F_0\}$

Combining these two results to deduce the relation between $V_t^{(3)}$ and $(F_t - F_0)$, we have:

$$sign\{E_0 V_t^{(3)}(F_t - F_0)\} = sign\{F_t - F_0\} = -\,sign\{f_0\}$$

Proof 5 therefore holds.

Proof 6: $E_0 M_{t,i} = 1/(1 + r_t)$

Assume that the investor borrows the quantity B_t of a one period risk free bond with a payoff of X, and a return of $(1 + r_t)$. Then the profit function (1) then becomes:

$$\pi_t = S_t y_0 - c(y_0) + f_0(F_0 - F_t) + n_0 Max(K - S_t, 0) + B_t X - B_{t-1} X(1 + r_t),\ \ \pi_0 = -n_0 P_0$$

Investors chooses the number of bonds so as to maximize his utility $V(\pi_{t+j}), j = 1,...,\infty$. This yields the “classic” result that the marginal rate of substitution equals the risk free discount rate.

$$E_0 M_{t,i} = 1/(1+r_t) \tag{47}$$

Proof 7: $P_0 \; = (1/(1+r_t))E_0^Q Max(K - S_t, 0)$

Equation (4) can be modified as follows:

$$P_0 \; = E_0 M_t \, Max(K - S_t, 0) = \sum_{s \in \Omega} P(s) M_t \, Max(K - S_t, 0) \tag{48}$$

Dividing and multiplying by the expected marginal rate of substitution $E_0 M_t$, equation (48) becomes:

$$P_0 \; = E_0 M_t \sum_{s \in \Omega} \frac{P(s) M_t}{E_0 M_t} Max(K - S_t, 0) \tag{49}$$

$Q(s) = \dfrac{P(s) M_t}{E_0 M_t}$ is the so called probability as the option under that probability $Q(s)$ is discounted at the risk free rate. Using the definition of the risk neutral probability and proof 6:

$$P_0 \; = (1/(1+r_t))E_0^Q Max(K - S_t, 0) \tag{50}$$

Proof 8: $TC_P = (1/N)\sum_{i=1}^{G} n_{0,i} E_0 M_{t,i} \rho_{t,i} Var_0(S_t)$

The trading cost for the put option, when the option is close to or at the money is by definition such that $S_0 \approx K$. Assuming the price of the underlying asset follows a random walk $E_0 S_t = S_0$, we find that $K \approx S_0 = E_0 S_t$, so that the trading cost

$$TC_p = \sum_{i=1}^{G} n_{0,i} E_0 \left(M_{t,i} / N \right) \left[\rho_{t,i} Max(K - S_0, 0)^2 \right] \tag{51}$$

becomes by replacing K by $E_0 S_t$,

$$TC_p \approx \sum_{i=1}^{G} n_{0,i} E_0 \left(M_{t,i} / N \right) \left[\rho_{t,i} Max((E_0 S_t - S_0)^2, 0) \right] \tag{52}$$

which further simplifies to $TC_P \approx Var_0(S_t)(1/N) \sum_{i=1}^{G} n_{0,i} E_0 M_{t,i} \rho_{t,i}$, where $Var_0(S_t)$ is the conditional variance of the underlying asset price.

Proof 9: $P_0 - \frac{1}{1+r_t} E_0 Max(K - S_t, 0) = Cov_0 \left(M_t, Max(K - S_t, 0) \right)$

Using the definition of conditional covariance, the price of the put option in equation (4) can be decomposed as follows:

$$P_0 = E_0 M_t \, E_0 Max(K - S_t, 0) + Cov_0 \left(M_t, Max(K - S_t, 0) \right) \tag{53}$$

Using equation (47) pricing a bond, we find therefore that:

$$P_0 - \frac{1}{1+r_t} E_0 Max(K - S_t, 0) = Cov_0 \left(M_t, Max(K - S_t, 0) \right) \tag{54}$$

The marginal rate of substitution is the risky rate at which the investor discounts payoffs. Hence, an investor is willing to pay a price P_0 above the arbitrage price $\left(1/(1+r_t) \right) E_0 Max(K - S_t, 0)$ if he can reduce with the option's payoff his risk, as the risky discount rate depends on the underlying asset price S.

CONCLUSION

Illiquidity is introduced in the optimisation or trading problem of the investor as an inability to trade and share risk without changing the market price. This creates an endogenous transaction cost and bid-ask spread without the need for informational asymmetries, inventory or order processing costs. This trading cost is a linear function of the variance of the spot market multiplied by the volume generated by hedging pressures, making the futures price dynamics richer. Illiquidity will exist in equilibrium even with very small trades if there is no price mechanism to force market clearing at a single futures price confirming that risk sharing is a source of illiquidity.

When the investor's degree of risk aversion is allowed to change with wealth, risk sharing creates an illiquidity trading cost that strengthens the wealth effect. This in turn increases the fatness of the left tail and skewness of the distribution of futures prices beyond that created by the wealth effect. Risk sharing becomes increasingly difficult as investors find themselves at risk, creating a pressure on the futures prices for speculators to accept the risk unloaded by hedgers. In the presence of non-separable preferences, this mechanism is again strengthened as investors worry about how uncertainty will resolve itself.

Illiquidity can potentially explain the graph of the implied volatility as a function of moneyness and as a function of volume or open interest. Illiquidity is closely related to volatility in the underlying spot price for options close to or at the money. It can therefore be partially controlled by hedging the Vega risk of a change in volatility. This could explain why Bollen and Whaley (2002) found that abnormal returns from selling put options disappear when controlling for Vega risk. Futures contracts and put options can be priced in equilibrium in illiquid markets using transformations of the subjective probability measures. In addition, the demand for futures contracts and put options increases as the hedging period, profits and size of production increases.

This thesis leaves open several questions for future research. First, what role do rational expectations play in the liquidity of financial markets? Markets may become illiquid if investors believe that a crash is going to occur. This belief may be self-fulfilling, thus precipitating the actual crash. Second, what is the impact of illiquidity on the Value at Risk (VAR) measures used to determine the amount of cash that banks should hold in reserve against a run on their deposits? The VAR depends on the shape of the left tail of the price distribution of their portfolio of assets and liabilities, which will change in a crash under selling pressures. There is therefore a clear incentive to find an improvement to VAR measures, which are currently used.

SUMMARY

Derivatives markets can quickly become illiquid in periods of high uncertainty. Neither the source of this illiquidity nor it implications are well understood. First, this thesis shows that hedgers' trades have an adverse impact on the futures price creating effectively an endogenous trading cost increasing in times of uncertainty and acting as the source of illiquidity in these markets. Second, illiquidity is shown to strengthen the wealth effect, which has been proven to be too weak empirically to explain the behavior of prices. The wealth effect is the mechanism through which changes in investors' wealth impact their attitude towards risk. As investors lose wealth, they become more risk averse and ask for a higher compensation to hold a risky asset thereby decreasing its price. Third, illiquidity is shown to potentially explain the shape of the implied volatility function not only as a function of moneyness but also of the options' volume or open interest. These results are derived from models, where producers maximize their expected utility derived from their profits. They seek to hedge the uncertain price at which they will sell their product in the future. They can use futures or put options to reduce this price risk but must pay speculators, defined as having no position in the underlying asset, a premium. This premium disappears normally as trades are assumed to be too small to matter and the risk of trading perfectly shared. Both of these assumptions are relaxed to derive the illiquidity trading costs and its implications.

REFERENCES

Acharya V., and L. Pedersen, 2002, Asset Pricing with Liquidity Risk, New York University, Stern School Finance Department Working Paper Series.

Amihud Y., 2002, Illiquidity and Stock Returns: Cross-Section and Time-Series Effects, *Journal of Financial Markets*, 5, 31-56.

Amihud Y., and H. Mendelson, 1986, Asset Pricing and the Bid-Ask Spread, *Journal of Financial Economics*, 17, 223-249.

Bangia A., Diebold F., Schuermann, T. and J. Stroughair, 1998, Modeling Liquidity Risk With Implications for Traditional Market Risk Measurement and Management, New York University, Stern School Finance Department Working Paper Series.

Bates D., 2001, The Market for Crash Risk, National Bureau of Economic, Research Working Paper 8557.

Bollen N. and R. Whaley, 2002, Does Net Buying Pressure Affect the Shape of Implied Volatility Functions?, Fuqua School of Business, Working Paper.

Brown D., and J. Jackwerth, 2001, The Pricing Kernel Puzzle: Reconciling Index Option Data and Economic Theory, University of Wisconsin at Madison, Working Paper.

Chueh H., and S. Yen, 2002, Decomposition of Bid-Ask Spreads in the Stock Index Futures Market, National Chengchi University, Working Paper.

Constantinides G., 1986, Capital Market Equilibrium with Transaction Costs, *The Journal of Political Economy*, 94(5), 842-862.

Constantinides, G., and D. Duffie, 1996, Asset Pricing with Heterogeneous Consumers, *The Journal of Political Economy*, 104, 219–240.

Constantinides G. and S. Perrakis, 2002, Stochastic Dominance Bounds on Derivative Prices in a Multiperiod Economy with Proportional Transaction Costs, NBER Working Paper W8867.

Demsetz H., 1968, The Cost of Transacting, *Quarterly Journal of Economics*, 82, 33-53.

Detemple J., and S. Murphy, 1994, Intertemporal Asset Pricing with Heterogeneous Beliefs, *Journal of Economic Theory*, 62, 294-320.

Duffie D., and C. Huang, 1985, Implementing Arrow-Debreu Equilibria by Continuous Trading of Few Short-Lived Securities, *Econometrica*, 53, 1337-1356.

Dumas B., 1989, Two-Person Dynamic Equilibrium in the Capital Market, *Review of Financial Studies*, 2, 157-188.

Ericsson J., and O. Renault, 2001, Liquidity and Credit Risk, FAME-International Center for Financial Asset Management and Engineering, Research Paper 42.

Franke G., Stapleton R., and M. Subrahmanyam, 1998, Who Buys and who Sells Options: The Role of Options in an Economy with Background Risk, *Journal of Economic Theory*, 82, 89-109.

Friend I., and M. Blume, 1975, The Demand for Risky Assets, *American Economic Review*, 65(5), 900-922.

Galy S., 2002, Endogenous Illiquidity Trading Costs from Risk Sharing, Working Paper, John Molson School of Business.

Galy S., 2002, Illiquidity and the Wealth Effect, Working Paper, John Molson School of Business.

Garman M., 1976, Market Microstructure, *Journal of Financial Economics*, 3, 257-275.

Glosten L. and P. Milgrom, 1985, Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders, *Journal of Financial Economics*, 14, 71-100.

Grossman S., and Z. Zhou, 1996, Equilibrium Analysis of Portfolio Insurance, *Journal of Finance*, 51, 1379–1403.

Harrison J., and S. Pliska, 1983, A stochastic Calculus Model of Continuous Trading: Complete Markets, *Stochastic Processes and their Applications*, 15, 313-316.

Hasbrouck J., and D. Seppi, 2001, Common Factors in Prices, Order Flows and Liquidity *Journal of Financial Economics*, 59 (2), 383-411.

Hasbrouck J., 2002, Liquidity in the Futures Pits: Inferring Market Dynamics from Incomplete Data, New York University, Stern School Finance Department Working Paper Series.

Hirshleifer D., 1988, Residual Risk, Trading Costs, and Commodity Futures Risk Premia, *Review of Financial Studies*, 1, 173-193.

Hirshleifer D., 1989, Determinants of Hedging and Risk Premia in Commodity Futures Markets, *Journal of Financial and Quantitative Analysis*, 24, 313-331.

Jacoby G., D. Fowler and A. Gottesman, 2000, The Capital Asset Pricing Model and the Liquidity Effect: A Theoretical Approach, *Journal of Financial Markets*, 3(1), 69-81.

Jacoby G., D. Fowler and A. Gottesman, 2002, On Asset Pricing Model and the Bid-Ask Spread, University of Manitoba, Working Paper.

Judd K. and D. Leisen, 2002, Equilibrium Open Interest, Working Paper, McGill University.

Kim J., K. Ko, and K. Noh, 2000, Time-varying Bid-Ask Components of Nikkei 225 Index Futures on SIMEX, Korea's Security Research Institute, Working Paper.

Kyle A., 1984, Market Structure, Information, Futures markets, and Price Formation, in International Agricultural Trade: Advanced Readings in Price Formation, Market Structure, and Price Instability, edited by Gary G. Storey et al. Boulder: Westview, 45-65.

Kyle A., 1985, Continuous auctions and insider trading, Econometrica, 53, 1315-1335.

Lucas R., 1973, Some International Evidence on Output-Inflation Trade-Offs, *American Economic Review*, 63(3), 326-334.

Lucas R., 1976, Econometric Policy Evaluation: A Critique, in K. Brunner and A. Meltzer (eds.) The Phillips Curve and the Labor Markets, Vol. 1 of Carnegie-Rochester Conference Series on Public Policy, Amsterdam, North Holland, 19-46.

Mayers D., 1973, Non-Marketable Assets and Capital Market Equilibrium Under Uncertainty, *Journal of Business*, 46, 258-267.

Mayers D., 1976, Nonmarketable Assets, Market Segmentation, and the Level of Asset Prices, *Journal of Financial and Quantitative Analysis*, 11, 1-12.

Muth J., 1961, Rational Expectations and the Theory of Price Movements, *Econometrica*, 29(3), 315-335.

O'Hara M., 1994, Market Microstructure Theory, Oxford: Blackwell Publishing.

Jackwerth J., 2000, Recovering Risk Aversion from Option Prices and Realized Returns, *Review of Financial Studies*, 13, 433-451.

Keynes J., 1930, A Treatise on Money, Vol. II (McMillan, London).

Leisen D., 2002, Current Option Pricing Models are Inconsistent with Trade, Working Paper, McGill University.

Magill M., and M. Quinzii, 1996, Theory of incomplete markets, London: MIT Press.

Roll R., 1984, A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market, *Journal of Finance*, 39, 1127-40.

Scholes M., 1972, The Market for Securities: Substitution Versus Price Pressure and the Effects of Information on Share Prices, *Journal of Business*, 45(2), 179-211.

Tien D., 2002, Hedging Demand and Foreign Exchange Risk Premia, University of California at Berkeley, Working Paper.

Varian H., 1992, Intermediate Microeconomic, A Modern Approach, Norton.

Wang J., 1994, A Model of Competitive Stock Trading Volume, *Journal of Political Economy*, 102, 127-168.