Markets for Goods, Markets for Risk*

Baiyang Han[†], Junyi Hu[‡], Anthony Lee Zhang[§]

July 2025

Abstract

Spot markets facilitate allocative efficiency, reallocating goods to maximize society's money-equivalent wealth. Financial markets facilitate risk-sharing, redistributing optimized wealth according to consumers' risk preferences. When financial markets are incomplete, the wealth generated by efficient spot markets is not optimally shared across agents. Financial market incompleteness is not fully resolved by realistic financial securities, such as futures contracts. Market incompleteness implies that policies which distort spot market outcomes, but improve risk-sharing, can be welfare-improving. In our model, a widely criticized policy tool – commodity price controls – can be Pareto-improving, if it improves risk-sharing across agents sufficiently.

*We thank Doug Diamond, Arpit Gupta, David Hirshleifer, Baiyun Jing, Ralph Koijen, Hrishikesh Relekar, and Yijing Ren for helpful comments.

[†]Stanford University Graduate School of Business; byhan@stanford.edu

[‡]USC Marshall School of Business; junyihu@marshall.usc.edu

[§]University of Chicago, Booth School of Business; anthony.zhang@chicagobooth.edu.

1 Introduction

What are the distinct roles of markets for *goods*, and markets for *risk*?

This paper develops a tractable model to answer this question. Money and commodities are traded in *spot markets*, which achieve *allocative efficiency* by redistributing the commodity across agents to maximize its money-equivalent value. Spot markets allocate commodities efficiently, but do not generally distribute wealth efficiently across uncertain states of the world. The role of idealized *financial markets* is to transform the wealth generated by efficient spot markets into state-contingent payoffs that optimally share risks across agents.

When financial markets are absent or incomplete, market outcomes are allocatively efficient, but risks generated by spot markets are not optimally shared across agents. In such settings, interventions in spot markets that reduce allocative efficiency can be welfare-enhancing, if they sufficiently improve risk sharing. We show that *price controls* in commodity markets can be Pareto-improving, in expected-utility terms, through their benefits for risk-sharing.

Our model has two goods: "money", and a real commodity, such as oil, wheat, or steel. Agents have a concave *production technology* which converts the commodity into money-equivalents: in other words, conditional on shock realization, agents' utility is quasilinear over the commodity and money. Ex-ante, agents are risk-averse over money, with CARA utility with potentially different risk aversions. The only source of uncertainty in the economy is agents' random endowments of the commodity. This setting is very general, but can be thought of as modelling trade and monetary risk sharing in any real factor of production.

What is the social first-best outcome? The social planner, in this setting, essentially solves a two-stage problem. Conditional on any realization of shocks, market outcomes should be *allocatively efficient*: commodities should be in the hands of those agents who have the most efficient technologies to convert them into money-equivalent consumption, in order to maximize the consumption available to society in each state of the world. *Risk sharing* should be optimal: shocks to aggregate inventory, optimally filtered through agents' production technologies, imply that society faces risky aggregate money-equivalent consumption; aggregate consumption shocks should be divided proportionally across agents according to their risk aversions, following Borch (1962) and Wilson (1968).

There is a simple and classical Arrow-Debreu implementation of the first-best outcome, which can be thought of as a backward induction process. Allocative efficiency is achieved through *spot markets*, markets which open after inventory shocks are realized, in which money and commodities are traded for each other. Agents' preferences over goods are quasilinear, so Walrasian equilibrium in spot markets is unique and maximizes society's aggregated money-metric utility, as in the social planner's problem. Risk sharing is achieved through *financial markets*: when agents can trade contingent claims on states of the world – the entire vector of inventory shocks – then financial markets decentralize the social planner's first-best solution.

The core departure point of our paper, as in the classic literature on incomplete markets, is that financial markets are likely incomplete in practice, so the planner's first-best is likely unattainable. We thus consider two more realistic cases: no financial markets, and financial markets limited to simple *commodity derivative contracts*.

When there are no financial markets, spot markets are allocatively efficient, but the distributions of wealth induced by Walrasian equilibria in spot markets fail to efficiently distribute risk among agents. A simple way to see why this must be the case is that spot market equilibria are functions only of realized inventory shocks and production technologies; they do not depend on risk aversions, and thus spot market equilibria cannot possibly share monetary risk according to risk aversions.

What is the nature of these distortions? In the socially optimal division of wealth between agents, all agents' monetary wealths scale linearly with aggregate inventory shocks, with a quadratic "aggregate concavity" term. The equilibrium outcome features a number of distortions. "Market makers" are paid *positively* for volatility exposure, and correspondingly "liquidity takers" overpay for noise in aggregate inventory. There is a *pecuniary risk externality*: agents who tend to be net buyers or sellers suffer risk because they buy or sell at noisy prices.

The inefficiency of spot markets in our model implies that *price controls* can be Pareto-improving. Price controls unambiguously decrease allocative efficiency, as they induce rationing and deadweight loss. However, they can improve risk-sharing, since they redistribute wealth towards agents with extreme inventory shocks, who have high marginal utility of wealth. When these insurance benefits are larger than the allocative efficiency losses, all agents achieve higher expected utility under price controls relative to free spot markets.

We next analyze the case of realistically incomplete financial markets. Suppose agents can trade *cash-settled futures contracts*, whose payouts are proportional to realized spot prices. We analytically solve for equilibria with spot markets and futures contracts, for arbitrary sets of agents' primitives. We can then evaluate the extent to which futures contracts can approximate the first-best outcome of complete markets and perfect risk sharing.

Futures contracts allow agents to trade *directional price risk*. Agents who expect to be buyers suffer from price increases, and agents who expect to be sellers suffer from price decreases; ex-ante futures trade allows buyers and sellers to cross-insure these price risk exposures. Futures markets are Pareto-improving relative to spot markets alone. However, they generally cannot implement first-best outcomes, since they span only a one-dimensional subspace of the N-dimensional space of inventory shocks.

What drives the *pricing* of futures contracts? A classic idea, going back to Keynes (1930), is that futures risk premia reflect *hedging demand*: if most futures activity is driven by commodity *buyers* hedging, their long positions push futures prices up, creating a premium for liquidity providers on the short side. Hirshleifer (1990) shows that this idea does not survive in general equilibrium. We extend this result to incomplete markets, and offer a new intuition for it. In our model, individual agents' futures demand is driven by hedging pressure: optimal futures positions move one-for-one with expected spot market purchases. But spot markets must clear, so expected spot market purchases always sum to zero: what buyers expect to buy, sellers expect to sell. This implies that buy-side and sell-side hedging demands always exactly cancel. Hedging pressure thus drives *trade volume* in futures market, but does not influence futures risk premia.

Rather than demand pressure, futures risk premia in our model – just as in the complete-markets model of Hirshleifer (1990) – derive from *consumption correlations*. Futures contracts pay more when spot prices are high; when the commodity has positive value, spot prices are high in states where aggregate inventory is low, society's wealth is low, and marginal utility is high. Since futures are valuable insurance against consumption risk, agents are willing to accept low expected returns on futures contracts, causing risk premia to be negative.

In our model, futures markets do not eliminate the gains from price controls. There are settings where futures markets are not used in equilibrium, but price controls are Pareto improving. There are also settings where price controls increase welfare when imposed alongside functioning futures markets.

Our primary contribution is a tractable incomplete-markets general equilibrium model, which distinguishes the roles of spot markets and financial markets in achieving efficient outcomes under uncertainty, and shows where realistic financial securities fall short of idealized Arrow securities. While straightforward, our model is, to our knowledge, new to the literature. Technically, we borrow some elements from Zhang (2022), who studies the setting of derivative market manipulation.

We contribute to a classic literature on general equilibrium in incomplete markets, surveyed in Geanakoplos (1990) and Magill and Quinzii (2002). Our model imposes two important restrictions relative to this literature.

First, we assume agents' preferences between goods and money are *quasilinear*. This creates a clean analytical separation between the roles of spot and financial markets: spot markets can be thought of as "converting" goods into wealth, and financial markets have the simple role of optimally dividing this "produced" aggregate wealth across agents. Quasilinearity removes income effects, thus avoiding many of the pathologies in the classic GEI literature, such as the possibility of equilibrium indeterminacy (Hart, 1975) and constrained inefficiency (Stiglitz, 1982; Geanakoplos and Polemarchakis, 1986). Equilibrium outcomes in our setting are constrained-efficient in the sense of Geanakoplos and Polemarchakis, since financial trades do not affect spot prices and thus do not generate pecuniary externalities. We explore a different kind of intervention: we show that direct distortions to *spot markets*, which cause outcomes to deviate from competitive-equilibrium allocations, can be Pareto improving if they sufficiently improve risk-sharing.

Second, we further assume inventory shocks are normally distributed, the production technology is quadratic, and agents have CARA utility over wealth. These strong assumptions allow us to solve our model analytically.

We contribute to a classic literature on futures contracts, by analyzing futures in a GEI setting. A closely related paper is Hirshleifer (1990), who analyzes futures risk premia assuming complete markets. By imposing stronger functional-form restrictions on utility and the distributions of inventory shocks, we can solve our model in the case of incomplete markets; this allows us to explicitly analyze the gap between futures equilibrium outcomes and first-best outcomes, and how policies such as price controls can complement futures markets in improving welfare.

In decomposing total welfare into allocative-efficiency and risk-sharing components, we also relate to Dávila and Schaab (2022) and Dávila and Schaab (2023), who propose a general framework to decompose welfare effects in heterogeneous-agent models; our "allocative efficiency" and "risk-sharing" notions correspond to the "aggregate efficiency" and "risk-sharing" components of the general setting of Dávila and Schaab (2022).

We are loosely related to a literature on incomplete markets and credit cycles. Krishnamurthy (2003) and Lorenzoni (2008) analyze the role of market incompleteness in amplifying credit cycles, and Dávila and Korinek (2018) characterize externalities and optimal corrective policy in such settings. Our model ignores borrowing and capital accumulation, focusing instead on the pecuniary externalities generated by spot markets.

The paper proceeds as follows. Section 2 introduces the model. Section 3 characterizes first-best outcomes. Section 4 analyzes outcomes in spot markets, together with complete Arrow securities, and then in autarky in the absence of financial markets. Section 5 analyzes price controls. Section 6 analyzes outcomes with futures markets. We discuss our results and conclude in Section 7.

2 Model

Notationally, we will use bold symbols to represent vectors, writing for example $\boldsymbol{x}$ to mean the vector $(x_1 \ldots x_N)$.

There are N "types" of consumers indexed by i, with a representative consumer of each type who behaves competitively, ignoring price impact.[1] For expositional simplicity, we will refer to the representative consumer of type i as simply "consumer i". Consumers have CARA utility over monetary wealth, with possibly different risk aversions α_i:

$$U_i(W_i) = -e^{-\alpha_i W_i} \tag{1}$$

There are two goods: money, and a single commodity. Consumer i is endowed with an initial constant amount m_i of money, and all consumers are allowed to hold infinitely large positive or negative positions in goods and money. We think of each consumer i as having a quadratic "production technology", which converts any positive or negative quantity z_i of goods into wealth:

$$W_i = m_i + \underbrace{\psi z_i - \frac{z_i^2}{2\kappa_i}}_{Production\ Technology} \tag{2}$$

Initial constant money endowments m_i have no effect on behavior under CARA utility, since m_i simply scales $U_i(W_i)$ in (1) by a constant factor. Thus, we proceed to set $m_i = 0$ for all i, so we can write W_i simply as a function of z_i:

$$W_i(z_i) = \psi z_i - \frac{z_i^2}{2\kappa_i} \tag{3}$$

Wealth $W_i(z_i)$ consists of a linear component ψz_i, which pays the consumer ψ per unit of the commodity; and a quadratic "inventory cost" component $\frac{z_i^2}{2\kappa_i}$, which implies that the marginal monetary value of the good is decreasing in the amount of the good held. Consumers with higher κ_i have lower inventory costs, and thus more elastic demand for the good. We will allow the edge case of $\kappa_i = 0$: we interpret a consumer with $\kappa_i = 0$ as having perfectly inelastic demand for exactly $z_i = 0$ units of the commodity, attaining $-\infty$ wealth if she finishes with any other value of z_i.

In the baseline model, we assume the only source of uncertainty is that consumers receive *inventory shocks*. Consumer i begins with a random endowment x_i of the commodity. We assume the x_i are independent normal random variables, but the mean μ_i and variance σ_i^2 of

[1] This is equivalent to assuming there is a unit measure of identical atomistic consumers of each type, who behave competitively because their trades are too small to move prices.

inventory shocks may vary across consumers. If i receives inventory shock x_i, and purchases q_i of the commodity at price p per unit, her final wealth is:

$$W_i = \psi\left(x_i + q_i\right) - \frac{\left(x_i + q_i\right)^2}{2\kappa_i} - pq_i \tag{4}$$

We call $W_i\left(z_i\right)$ a "production technology" because it is intuitive to think of z_i being literally transformed into units of consumable wealth. After "transformation" of z_i, the economy reduces to a single-good problem: each consumer has some amount of produced wealth, which can be redistributed across consumers arbitrarily, since money is tradable and consumers have deep pockets. Of course, it is isomorphic to think of $W_i\left(z_i\right)$ as a preference function for z_i rather than a production technology; in these terms, consumer i gets utility equivalent to having $W_i\left(z_i\right)$ extra dollars from having z_i units of the commodity.

There are two periods. The first period is a market for *risk*: consumers may trade *financial securities* which alter their endowments of goods or money in future states of the world. In the following sections, we will analyze three financial market structures: complete financial markets with Arrow securities; no financial markets; and commodity futures contracts, which we will show constitute an incomplete financial market. We ignore consumption in the first period, so all asset trades in the first period transfer consumption across future states of the world.

In the second period, consumers' inventory shocks x_i are realized. The second period is a market for *goods*, or in traditional terms, a *spot market*: conditional on financial securities trades made in period 1, and inventory shock realizations $x_1 \ldots x_N$, consumers trade money for the commodity.

2.1 Discussion of Model Assumptions

Formally, our setting is a simple case of *general equilibrium with incomplete markets*, sometimes referred to as GEI. Our key simplification is that we assume consumers' preferences over goods and money are *quasilinear*. In (3), consumers' wealth is the sum of money and a concave function of commodity holdings. Thus, in spot markets, we have a *partial-equilibrium* supply-demand model, in the spirit of Marshall (1920). Equilibria in spot markets are unique; equilibrium prices and quantities are unaffected by the distribution of "money" across agents; and equilibria increase society's money-metric welfare by an amount equal to the classic consumer and producer surplus triangle areas.

Our model can be thought of as a generalization of partial equilibrium analysis to settings with *risk*. We do this simply by assuming that consumers have concave utility over the money-metric wealth generated by spot market allocations. This drives the clean separation in our model between the roles of markets for goods and markets for risk. Spot markets facilitate *allocative efficiency*, maximizing the sum of money-metric wealth $W_i(z_i)$ across consumers. Financial markets facilitate *risk-sharing*, redistributing this wealth depending on consumers' risk preferences. This clean decomposition is possible because of quasilinearity: in the general case with income effects, spot markets and financial markets are entangled in ways that make it difficult to isolate the respective roles they play.

Quasilinearity also eliminates many of the known pathologies in the GEI literature. In general GEI models, there may be multiple equilibria, and equilibria need not be efficient or even constrained-efficient (Magill and Quinzii, 2002). These pathologies are mainly driven by *income effects*, which link financial market trades to spot market outcomes. Since our model has no income effects, equilibria are unique and are always constrained-efficient in the sense of Geanakoplos and Polemarchakis (1986), though we will discuss a different sense in which equilibria are inefficient.

In summary, our setting can be thought of as a *partial equilibrium incomplete-market model*. Its main benefits are that it is intuitive and tractable relative to general GEI models. Like classic partial-equilibrium models, our setting is less appropriate for studying goods that are "large" enough that income effects on demand are important.

Normal Shocks, Quadratic Technologies, CARA Utility. In addition to quasilinearity, we assume that inventory shocks are normally distributed, production technologies are quadratic, and utility is CARA. These assumptions yield analytic expressions for expected utility, and allow us to explicitly solve for futures market outcomes in Section 6. Relaxing any subset of these assumptions, while retaining quasilinearity, would preserve existence, uniqueness, and constrained optimality of equilibria. The model would still admit a clean separation between allocative efficiency and risk-sharing, and could in principle be solved numerically.

Differences from CARA-Normal Models. There are subtle, but very important, differences between our model and the standard "CARA-normal" setting. In many models, z_i is thought of as a *financial asset*, which has risky direct monetary payoffs ψz_i, where ψ is a random variable. This generates a risk-sharing motive for trade in financial markets: prior to the realization of ψ, holding z_i is risky, so consumers trade the financial asset according to their risk aversions. However, there is no spot market in these models: after ψ is realized, the financial asset is equivalent to money, and there is no further motive for trade.

Our model is different because z_i represents not a financial asset, but a generalized *commodity*, which is distinct from money even after uncertainty is realized. The good is *converted* to money (or money-equivalents), through the quadratic agent-dependent production technology in (2). After inventory shocks x_i are realized, even though there is no residual uncertainty, consumers trade to improve *allocative efficiency*, moving goods into the hands of those who are most effective at converting goods into money. In reality, our model could be thought of as representing commodities such as oil or wheat being sold from producers and intermediaries to consumers. For simplicity, we assume ψ is constant, so all price risk is generated by inventory shocks x_i.

Our model also abstracts entirely from *common values*, *information frictions*, and other forces analyzed in the *rational expectations equilibrium* literature following Grossman and Stiglitz (1980). We assume futures are traded before agents receive any signals, so there is no scope for adverse selection. The only imperfection is market incompleteness: Arrow-style state-contingent contracts exist, but fail to span the full state space.

3 The First-Best Outcome

Conditional on any vector of inventory shocks $\boldsymbol{x}$, the social planner can freely reallocate commodities across consumers; that is, the social planner chooses functions $z_1(\boldsymbol{x})$ to $z_N(\boldsymbol{x})$, satisfying, pointwise in $\boldsymbol{x}$, the aggregate resource constraint:

$$\sum_{i=1}^{N} z_i(\boldsymbol{x}) = \sum_{i=1}^{N} x_i \tag{5}$$

It is of course equivalent to assume that the social planner chooses the net trade amounts $q_i(\boldsymbol{x})$ rather than the final inventories. Second, the social planner can freely reallocate wealth across agents, pointwise in $\boldsymbol{x}$. Conditional on the planner's choice of final inventories, society's aggregate wealth is:

$$W(\boldsymbol{z}(\boldsymbol{x})) \equiv \sum_{i=1}^{N} W_i(z_i) = \sum_{i=1}^{N} \psi z_i(\boldsymbol{x}) - \frac{(z_i(\boldsymbol{x}))^2}{2\kappa_i} \tag{6}$$

The social planner thus chooses final monetary wealths of agents, which we will call $G_i(\boldsymbol{x})$, subject to the constraint, pointwise in $\boldsymbol{x}$, that:

$$\sum_{i=1}^{N} G_i(\boldsymbol{x}) \leq W(\boldsymbol{z}(\boldsymbol{x})) \tag{7}$$

Thus, in sum, the social planner chooses commodity allocations $z_i(\boldsymbol{x})$ and money allocations $G_i(\boldsymbol{x})$, satisfying (5) and (7). An allocation is *Pareto efficient* if it is not expected-utility dominated by some other allocation; formally, under our assumption of CARA utility, $\tilde{G}_i(\boldsymbol{x})$ Pareto-dominates $G_i(\boldsymbol{x})$ if:

$$E\left[-e^{-\alpha_i \tilde{G}_i(\boldsymbol{x})}\right] \geq E\left[-e^{-\alpha_i G_i(\boldsymbol{x})}\right] \quad \forall i, \quad \text{and } E\left[-e^{-\alpha_i \tilde{G}_i(\boldsymbol{x})}\right] > E\left[-e^{-\alpha_i G_i(\boldsymbol{x})}\right] \quad \text{for some } i \tag{8}$$

To handle probability-zero edge cases, we will additionally strengthen this definition by saying that $\tilde{G}_i(\boldsymbol{x})$ Pareto-dominates $G_i(\boldsymbol{x})$ if $\tilde{G}_i(\boldsymbol{x}) \geq G_i(\boldsymbol{x})$ for all i and all realizations of $\boldsymbol{x}$, and $\tilde{G}_i(\boldsymbol{x}) > G_i(\boldsymbol{x})$ for some i and $\boldsymbol{x}$, even if the set of $\boldsymbol{x}$ values on which the inequality is strict has measure zero.

Notice that, while commodity allocations $z_i(\boldsymbol{x})$ do not explicitly enter into (8), they matter because they constrain money allocations $G_i(\boldsymbol{x})$ through the wealth constraint (7).

Proposition 1. *Pareto-efficient commodity allocations $z_i^*(\boldsymbol{x})$ and money allocations $G_i^*(\boldsymbol{x})$ are characterized by two conditions:* spot market allocative efficiency*, and* optimal risk-sharing*. Spot market allocative efficiency requires that commodity allocations $z_i^*(\boldsymbol{x})$ satisfy:*

$$z_i^*(\boldsymbol{x}) = \frac{\kappa_i}{\sum_{j=1}^N \kappa_j} \sum_{j=1}^N x_j \tag{9}$$

In any efficient spot market allocation, society's aggregate wealth is:

$$W^*(\boldsymbol{x}) \equiv W(\boldsymbol{z}^*(\boldsymbol{x})) = \psi \sum_{i=1}^N x_i - \frac{\left(\sum_{i=1}^N x_i\right)^2}{2\sum_{i=1}^N \kappa_i} \tag{10}$$

Optimal risk-sharing requires that wealth is shared as:

$$G_i^*(\boldsymbol{x}) = C_i + \frac{\alpha_i^{-1}}{\sum_{j=1}^N \alpha_j^{-1}} W^*(\boldsymbol{x}), \tag{11}$$

where $\sum_{i=1}^N C_i = 0$.

The intuition behind Proposition 1 is straightforward. In any Pareto-efficient allocation, spot market commodity allocations $z_i^*(\boldsymbol{x})$ must be efficient, in the sense that commodities are distributed in a way which optimally produces money, given consumers' heterogeneous production technologies $W_i(z_i)$. If this were not the case for any realization $\boldsymbol{x}$, it would be possible to simply generate more wealth for society, and redistribute this wealth in a way that increases consumers' utility. Expression (9) is intuitive: since all consumers have quadratic inventory costs, the aggregate endowment $\sum_{j=1}^{N} x_j$ is simply divided among consumers proportional to their inventory capacities κ_i; higher-κ_i consumers have more elastic demand, suffering lower costs for absorbing inventory, and thus absorb a larger fraction of aggregate inventory shocks in equilibrium.

Through the optimal spot market allocations, society simply transforms commodities $\boldsymbol{x}$ into some total monetary wealth $W^*(\boldsymbol{x})$, characterized by (10). This expression has a simple interpretation: when spot markets function optimally, the N consumers' wealth is equivalent to a single representative consumer with inventory capacity:

$$K \equiv \sum_{i=1}^{N} \kappa_i$$

Conditional on spot market optimal allocations, society then faces a simple one-good risk-sharing problem: there is some random total monetary wealth $W^*(\boldsymbol{x})$ which is to be divided amongst risk-averse consumers. Then, Pareto efficiency requires the equalization of the ratio of marginal utility across states. Under our assumption of CARA utility, the classic results of Borch (1962) and Wilson (1968) imply that any Pareto-efficient allocations redistribute risks in wealth, driven by uncertainty in $\boldsymbol{x}$, affinely according to consumers' risk aversions, as in (11).

Proposition 1 shows that Pareto efficiency is a very restrictive criterion in our model: consumers' spot market outcomes are fully pinned down, and wealths are pinned down across states up to consumer-specific constants. Thus, with slight abuse of terminology, we will occasionally refer to the outcomes described in Proposition 1 as "the first-best outcome" in singular form, implicitly ignoring the constant terms in (11).

We now formally prove Proposition 1.

Proof. After the realization of shocks, consumers' utility is quasilinear in money, implying that all Pareto-efficient outcomes must maximize the sum of consumers' monetary-equivalent values of goods; any non-maximizing allocation is Pareto-dominated with transfers. That is, efficient allocations must solve:

$$\max_{z_i} \sum_{i=1}^{N} W_i(z_i) = \max_{z_i} \sum_{i=1}^{N} \psi z_i - \frac{z_i^2}{2\kappa_i}$$

$$s.t. \sum_{i=1}^{N} z_i = \sum_{i=1}^{N} x_i$$

The Lagrangian is:

$$\Lambda = \max_{z_i} \left[\sum_{i=1}^{N} \psi z_i - \frac{z_i^2}{2\kappa_i} \right] - \lambda \left(\sum_{i=1}^{N} z_i - \sum_{i=1}^{N} x_i \right)$$

The first-order condition is:

$$0 = \frac{\partial \Lambda_i}{\partial z_i} = \psi - \frac{z_i}{\kappa_i} - \lambda$$

Implying simply that consumers' marginal rate of substitution between wealth and goods, $\frac{\partial W_i}{\partial z_i} = \psi - \frac{z_i}{\kappa_i}$, must be equated:

$$\frac{z_i}{\kappa_i} = C$$

Combining this with the resource constraint (5), we get (9), which uniquely characterizes the allocations $z_i^* (\boldsymbol{x})$ which maximize aggregate wealth, conditional on any inventory shock realization $\boldsymbol{x}$. Plugging (9) into consumers' production technology (3) and summing, we then get (10).

To show that (9) is necessary for Pareto efficiency, suppose $z_i (\boldsymbol{x})$ does not satisfy (9) for some i and $\boldsymbol{x}$. For any realization $\boldsymbol{x}$ where (9) is violated, replacing $z_i (\boldsymbol{x})$ with $z_i^* (\boldsymbol{x})$ increases total social wealth $W (\boldsymbol{z} (\boldsymbol{x}))$ and loosens the constraint (7). We can thus increase $G_i (\boldsymbol{x})$ for all i, leading to a Pareto improvement. We can also conclude from the above analysis that (7) must be binding.

Suppose now that $z_i (\boldsymbol{x})$ does satisfy (9). The optimal risk-sharing condition is simply the result of Borch (1962) in our setting. Pareto efficiency requires agents to equate the ratio of their marginal utilities across all states:

$$\frac{w_i}{w_j} = \frac{U_i' (G_i (\boldsymbol{x}))}{U_j' (G_j (\boldsymbol{x}))} = \frac{\alpha_i e^{-\alpha_i G_i(\boldsymbol{x})}}{\alpha_j e^{-\alpha_j G_j(\boldsymbol{x})}},$$

where w_i is the weight for i's utility. Combining this with the binding constraint (7), we get (11), following Wilson (1968). Hence, the wealth allocation is uniquely characterized up to agent-specific, state-independent constants. □

4 Spot Market Equilibrium

We next solve for equilibrium in spot markets, and illustrate why spot markets fail to implement the first-best outcome.

After the realization of inventory shocks x_i, there is no residual uncertainty; the problem reduces to a simple quasilinear-utility Walrasian equilibrium, where the only goods are money and the commodity. From (4), consumer i's wealth, as a function of the quantity q_i of the commodity she purchases, is:

$$W_i = C_i + \psi\left(x_i + q_i\right) - \frac{\left(x_i + q_i\right)^2}{2\kappa_i} - pq_i \tag{12}$$

where C_i is some money endowment that i may have attained, through pre-spot market trade. Consumers' marginal rate of substitution between q_i and wealth can be derived by differentiating (12) with respect to q_i:

$$\frac{\partial W_i}{\partial q_i} = \psi - \frac{x_i + q_i}{\kappa_i} - p \tag{13}$$

The marginal wealth value of q_i thus depends on x_i and q_i, but not C_i: preference quasilinearity implies that there are no income effects, so wealth transfers do not affect spot market demand, and thus spot market equilibrium prices and quantities.

W_i is concave in q_i, so consumers simply purchase until the marginal wealth value of purchasing becomes negative. Setting (13) to zero and solving for q_i, we obtain consumers' demand for the commodity, as a function of the spot price p of commodities in terms of money:

$$q_i\left(p\right) = -x_i - \kappa_i\left(p - \psi\right) \tag{14}$$

Hence, the inventory shock x_i determines the intercept of the demand curve, and inventory capacity κ_i determines the slope.

Spot market equilibrium is characterized by a scalar price p which leads spot markets to clear. Summing over consumers' demand, we need:

$$\sum_{i=1}^{N} q_i\left(p\right) = 0$$

implying:

$$\sum_{i=1}^{N}\left[x_i + \kappa_i\left(p - \psi\right)\right] = 0$$

The spot market clearing price is thus simply a function of consumers' inventory shocks:

$$p^{SpotEqm}\left(\boldsymbol{x}\right) - \psi = -\frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} \kappa_i} \tag{15}$$

Intuitively, the equilibrium price deviation from ψ is simply the aggregate inventory shock

$\sum_{i=1}^{N} x_i$, divided by the aggregate "inventory capacity", or alternatively the slope of aggregate demand, $\sum_{i=1}^{N} \kappa_i$. As in standard partial-equilibrium settings, $p^{SpotEqm}(x_1 \dots x_N)$ can be interpreted as society's marginal monetary value of having an additional unit of the commodity, which depends only on the aggregate amount of the commodity society is endowed with, $\sum_{i=1}^{N} x_i$.

Plugging (15) into consumer demand (14), we can calculate consumers' equilibrium inventories:

$$x_i + q_i^{SpotEqm}(x_1 \dots x_N) = \frac{\kappa_i}{\sum_i \kappa_i} \sum_i x_i \tag{16}$$

That is, consumer i ends up holding a fraction $\frac{\kappa_i}{\sum_i \kappa_i}$ of the aggregate inventory shock $\sum_i x_i$, implementing the first-best outcome (9). Intuitively, conditional on the realization of inventory shocks, the two-good money-and-commodities market is trivially complete, and the welfare theorems hold. Spot market competitive equilibria are *allocatively efficient*, in the sense of always allocating commodities in a way which maximizes society's aggregate monetary wealth.

Let W_i^0 represent i's welfare in autarky, from consuming her endowment x_i:

$$W_i^0 = C_i + \psi x_i - \frac{x_i^2}{2\kappa_i} \tag{17}$$

Taking the difference between (12) and (17), plugging in (14), and simplifying, i's money-metric welfare gains from trade are simply:

$$W_i - W_i^0 = \frac{q_i^2}{2\kappa_i} \tag{18}$$

Since preferences are quadratic, expression (18) is just i's *consumer surplus triangle*: it is half the product of her trade quantity, q_i, and her marginal WTP for the good when trading nothing, $\frac{q_i}{\kappa_i}$. Quasilinear preferences in second-stage markets imply that compensating and equivalent variation are equal to each other, and to the integral of Marshallian demand over prices, which is (18). Society's total monetary welfare gains from trade are simply the sum of surplus triangles over all consumers.

We can also calculate the *wealth distribution* induced by competitive equilibria in spot markets, by plugging equilibrium quantities (16) and prices (15) into consumers' wealth, (12).

In Appendix A.1, we show that this simplifies to:

$$W_i^{SpotEqm} = m_i + \underbrace{\psi x_i}_{Wealth\ Shocks} + \underbrace{\frac{\kappa_i}{\sum_{j=1}^N \kappa_j} \frac{\left(\sum_{j=1}^N x_j\right)^2}{2\sum_{j=1}^N \kappa_j}}_{Inventory\ Absorption} - \underbrace{x_i \frac{\sum_{j=1}^N x_j}{\sum_{j=1}^N \kappa_j}}_{Price\ Exposure} \tag{19}$$

4.1 Arrow-Debreu Securities, and Implementation of First-Best

Suppose agents can trade *Arrow securities*, denominated in units of wealth, which fully span the state space. Markets are then complete, the welfare theorems hold, and market equilibrium implements the first-best outcome.

Since our state space $\boldsymbol{x}$ is continuous, Arrow security prices constitute a *state price density* (Duffie, 2010, ch. 2), which we will refer to as $\pi(\boldsymbol{x})$.[2] Let $\theta_i(\boldsymbol{x})$ denote the *security demand function* of consumer i; that is, consumer i purchases securities paying her a net amount $\theta_i(\boldsymbol{x})$ in state $\boldsymbol{x}$. Since there is no first-stage consumption, agents trade money across states of the world by buying Arrow securities in some states and selling them in other states. Agents' budget constraint is that their total expenditures must integrate to 0 across states:

$$\int \pi(\boldsymbol{x})\,\theta_i(\boldsymbol{x})\,d\boldsymbol{x} = 0\ \forall i \tag{20}$$

Note that, as is traditional in the literature, we absorb the physical probability density $dF(\boldsymbol{x})$ into the definition of $\pi(\boldsymbol{x})$.

Agents purchase Arrow securities to maximize expected utility subject to (20). We will require *market clearing* pointwise in $\boldsymbol{x}$; since Arrow securities are financial assets in zero net supply, asset demands must sum to zero across agents:

$$\sum_{i=1}^N \theta_i(\boldsymbol{x}) = 0\ \forall \boldsymbol{x} \tag{21}$$

Equilibrium is described by a state price density $\pi(\boldsymbol{x})$ and security demands $\boldsymbol{\theta}(\boldsymbol{x})$, such that all consumers are maximizing utility, and markets for Arrow securities clear.

Proposition 2. *When Arrow securities are available, the unique equilibrium state price*

[2] A subtle difference between our model and the canonical setting is that, since we assume there is no first-period consumption, there is no natural numeraire in our setting. Thus, we leave $\pi(\boldsymbol{x})$ defined only up to scale. An equivalent alternative approach would be to choose some arbitrary value of $\boldsymbol{x}$ as the numeraire good.

density is:

$$\pi(\boldsymbol{x}) = C \cdot \exp\left(-\frac{W^*(\boldsymbol{x})}{\sum_{j=1}^{N} \alpha_j^{-1}}\right) \cdot dF(\boldsymbol{x}) \tag{22}$$

where C is an arbitrary positive constant. Agents' asset demands are:

$$\theta_i(\boldsymbol{x}) = W_i^*(\boldsymbol{x}) - W_i^{SpotEqm}(\boldsymbol{x}) \tag{23}$$

where:

$$W_i^*(\boldsymbol{x}) = \frac{\mathbb{E}\left[\exp\left(-\frac{W^*}{\sum_{j=1}^{N}\alpha_j^{-1}}\right) \cdot \left(W_i^{SpotEqm} - \frac{\alpha_i^{-1}}{\sum_{j=1}^{N}\alpha_j^{-1}} W^*\right)\right]}{\mathbb{E}\left[\exp\left(-\frac{W^*}{\sum_{j=1}^{N}\alpha_j^{-1}}\right)\right]} + \frac{\alpha_i^{-1}}{\sum_{j=1}^{N}\alpha_j^{-1}} W^*(\boldsymbol{x}) \tag{24}$$

The equilibrium with Arrow securities is Pareto-efficient.

Proof. Spot market equilibrium endows agent i with $W_i^{SpotEqm}(\boldsymbol{x})$ wealth in state $\boldsymbol{x}$. When agents can trade Arrow securities, markets are trivially complete, so the first welfare theorem implies that equilibrium allocations are Pareto-efficient. We use $W_i^*(\boldsymbol{x})$ to denote i's total equilibrium wealth in state $\boldsymbol{x}$: this is the sum of spot wealth $W_i^{SpotEqm}(\boldsymbol{x})$ and any Arrow security payoffs $\theta_i(\boldsymbol{x})$. Agents' FOC for optimal security demand implies that the state price density is determined by agents' marginal utilities at $W_i^*(\boldsymbol{x})$:

$$\frac{\pi(\boldsymbol{x})}{\pi(\boldsymbol{x}')} = \frac{m(\boldsymbol{x}) \cdot dF(\boldsymbol{x})}{m(\boldsymbol{x}') \cdot dF(\boldsymbol{x}')} = \frac{U_i'(W_i^*(\boldsymbol{x})) \cdot dF(\boldsymbol{x})}{U_i'(W_i^*(\boldsymbol{x}')) \cdot dF(\boldsymbol{x}')}.$$

Using the representation of first-best wealth allocations in (11) of Proposition 1, we have:

$$\frac{U_i'(W_i^*(\boldsymbol{x})) \cdot dF(\boldsymbol{x})}{U_i'(W_i^*(\boldsymbol{x}')) \cdot dF(\boldsymbol{x}')} = \frac{\exp\left(-\frac{W^*(\boldsymbol{x})}{\sum_{j=1}^{N}\alpha_j^{-1}}\right) \cdot dF(\boldsymbol{x})}{\exp\left(-\frac{W^*(\boldsymbol{x}')}{\sum_{j=1}^{N}\alpha_j^{-1}}\right) \cdot dF(\boldsymbol{x}')} \tag{25}$$

which gives (22). Notice that, under CARA utility, all Pareto efficient allocations imply the same state-price density: the constant terms in (11) fall out of the ratio in (25).

To calculate equilibrium Arrow security demands, note that spot market equilibrium endows i with wealth $W_i^{SpotEqm}(\boldsymbol{x})$ in state $\boldsymbol{x}$, and Pareto-efficient wealth allocations $W_i^*(\boldsymbol{x})$ have the form in (11) of Proposition 1. In order for $\theta_i(\boldsymbol{x})$ to induce Pareto-efficient wealth allocations, we must have, for each i:

$$\theta_i(\boldsymbol{x}) = W_i^*(\boldsymbol{x}) - W_i^{SpotEqm}(\boldsymbol{x}) = C_i + \frac{\alpha_i^{-1}}{\sum_{j=1}^N \alpha_j^{-1}} W^*(\boldsymbol{x}) - W_i^{SpotEqm}(\boldsymbol{x}). \tag{26}$$

for some C_i. We can find C_i using the budget constraint (20), substituting (22) and (26):

$$C \int \left(C_i + \frac{\alpha_i^{-1}}{\sum_{j=1}^N \alpha_j^{-1}} W^*(\boldsymbol{x}) - W_i^{SpotEqm}(\boldsymbol{x}) \right) \cdot \exp\left(-\frac{W^*(\boldsymbol{x})}{\sum_{j=1}^N \alpha_j^{-1}} \right) \cdot dF(\boldsymbol{x}) = 0$$

Solving, we have:

$$C_i = \frac{\mathbb{E}\left[\exp\left(-\frac{W^*}{\sum_{j=1}^N \alpha_j^{-1}} \right) \cdot \left(W_i^{SpotEqm} - \frac{\alpha_i^{-1}}{\sum_{j=1}^N \alpha_j^{-1}} W^* \right) \right]}{\mathbb{E}\left[\exp\left(-\frac{W^*}{\sum_{j=1}^N \alpha_j^{-1}} \right) \right]}$$

This gives (24). □

The proof of Proposition 2 illustrates cleanly the separate roles of *markets for goods* and *markets for risk* in our model. Markets for goods – spot market equilibrium – optimally convert commodities into wealth: the two-good problem reduces to a one-good problem, in which each consumer is endowed with $W_i^{SpotEqm}(\boldsymbol{x})$ dollars in state $\boldsymbol{x}$. This optimizes total social wealth – the sum of $W_i^{SpotEqm}(\boldsymbol{x})$ is equal to $W^*(\boldsymbol{x})$ – but does not efficiently distribute this wealth across agents. Markets for risk then simply allow agents to trade from their spot-equilibrium wealth $W_i^{SpotEqm}(\boldsymbol{x})$ to their first-best wealths $W_i^*(\boldsymbol{x})$.

Markets for risk are very generally needed for efficiency because, from comparing (19) and (11), $W_i^{SpotEqm}(\boldsymbol{x})$ and $W_i^*(\boldsymbol{x})$ in general have very different expressions. We discuss these distortions in detail below; one simple observation is that risk aversion α appears in first-best wealth $W_i^*(\boldsymbol{x})$, but not in spot equilibrium wealth $W_i^{SpotEqm}(\boldsymbol{x})$. Spot markets occur *conditional* on the realization of all uncertainty. Risk aversion over wealth is thus not relevant in spot markets, and cannot influence $W_i^{SpotEqm}(\boldsymbol{x})$, simply because there is no residual risk.[3] Clearly, there must be some form of financial market trade, in order to allow consumers' risk preferences to influence their final allocations of wealth.

When financial markets are incomplete, wealth will be created efficiently, but distributed

[3] Another way to see this is that, in a two-good spot market after $\boldsymbol{x}$ is realized, utility is ordinal rather than cardinal: a consumer's preferences are fully described by indifference curves between money and goods, which are traced out by (12). Expression (19) for $W_i^{SpotEqm}(\boldsymbol{x})$ is thus valid *regardless* of what consumers' preferences over wealth are: (19) holds for any choice of CARA-utility risk aversions, or indeed any other classes of risk-averse preferences over wealth.

inefficiently. We demonstrate this first in the simple case where there are no financial markets, so consumers' final wealth is simply their spot market equilibrium wealth.

4.2 Wealth Distortions induced by Spot Markets

What are the deviations between the spot market-induced wealth outcomes, $W_i^{SpotEqm}$ in (19), and the first-best wealth outcomes? We illustrate the nature of these distortions within two stylized examples, then discuss the general case.

4.2.1 Market Makers and Liquidity Takers

First suppose we have two types of agents, the liquidity taker (T) and the market maker (M). T has risk aversion α_T, receives an inventory shock $x_T \sim N(0, \sigma_T)$, and has no capacity to absorb inventory, $\kappa_T = 0$. M has risk aversion α_M, receives no inventory shock, $x_M = 0$, and has positive capacity $\kappa_M > 0$.

In equilibrium, given that the liquidity taker has completely inelastic demand, the market maker simply buys the entire inventory shock from the liquidity taker:

$$q_T = -x_T, \; q_M = x_T$$

From (11), the efficient wealths are:

$$W_T^* = C_T + \psi \frac{\alpha_T}{\alpha_T + \alpha_M} x_T - \frac{1}{2} \frac{\alpha_T}{\alpha_T + \alpha_M} \frac{x_T^2}{\kappa_M} \tag{27}$$

$$W_M^* = C_M + \psi \frac{\alpha_M}{\alpha_M + \alpha_M} x_T - \frac{1}{2} \frac{\alpha_M}{\alpha_M + \alpha_M} \frac{x_T^2}{\kappa_M} \tag{28}$$

From (19), equilibrium wealths are:

$$W_T^{eq} = x_T \psi - \frac{x_T^2}{\kappa_M} \tag{29}$$

$$W_M^{eq} = \frac{1}{2} \frac{x_T^2}{\kappa_M} \tag{30}$$

This example thus illustrates, in the setting of a totally canonical market maker-noise trader model, the *wealth distribution* induced in spot market competitive-equilibrium.

Wealth Shock Sharing. In our model, the commodity is valuable on average – ignoring the concave term, each unit of the commodity is worth ψ dollars – so inventory shocks serve as *wealth shocks*. In the first-best outcome, society is wealthier when the endowment x_T is higher, and this shock is split between T and M according to their risk aversions: this is the linear term in (27) and (28). In contrast, in equilibrium, the entire linear term $x_T\psi$ is kept by T in (29): M has no linear exposure to the asset.

A simple way to see this is that, in the limit as $\kappa_M \to \infty$ – a well-defined limit within our model – T's wealth is linear in inventory, so there is no concavity and a unit of the commodity is equivalent to ψ units of wealth. Inventory shocks to T are thus simply *idiosyncratic wealth shocks*: a shock of x_T is equivalent to ψx_T units of money. Clearly, spot markets – which are trivial in this limit – do not allow T to trade risk associated with these idiosyncratic wealth shocks.

Volatility Exposures. In the social optimum, all agents have concave exposure to aggregate inventory shocks, because agents receive constant proportions of society's wealth in the first-best outcome, and society's aggregate wealth is a concave function of aggregate inventory: society has decreasing marginal value for the commodity. Interestingly, in spot market equilibrium, the sign of M's volatility exposure is wrong. Market makers, due to their ability to offer spot liquidity to the market, acquire *long-volatility* positions in spot market equilibrium: their expected wealth is *convex* in aggregate inventory shocks. Liquidity takers thus are overexposed to volatility. In equilibrium, their wealth is even more concave in aggregate inventory shocks than society's wealth. Effectively, liquidity takers can be thought of as paying both due to society's decreasing marginal value of the commodity, and also due to a "liquidity tax" paid to market makers.

4.2.2 Buyers and Sellers

Next, we consider an example with T, M and two additional consumers, called buyer (B) and seller (S). B and S have nonrandom inventories, $x_B = -1$, and $x_S = 1$, and have no ability to absorb capacity, $\kappa_B = \kappa_S = 0$. Intuitively, B and S simply want to inelastically buy and sell a unit of the commodity, and their net demands cancel out.

The efficient allocations are:

$$q_T = -x_T,\ q_M = x_T,\ q_B = 1,\ q_S = -1$$

This allocation generates identical societal wealth to the last example, so in the first-best, this wealth is divided between agents according to their risk aversions, for example:

$$W_S^* = C_S + \psi \frac{\alpha_S}{\alpha_T + \alpha_M + \alpha_B + \alpha_S} x_T - \frac{1}{2} \frac{\alpha_S}{\alpha_T + \alpha_M + \alpha_B + \alpha_S} \frac{x_T^2}{\kappa_M} \tag{31}$$

and likewise for B, T, M. Prices are exactly the same as in the previous example, so the wealths of M and T are unchanged from (29) and (30). Equilibrium wealths for the buyer and seller are now:

$$W_B^{eq} = -\frac{x_T}{\kappa_M} \tag{32}$$

$$W_S^{eq} = \frac{x_T}{\kappa_M} \tag{33}$$

What is happening is that, regardless of the behavior of M and T, S can be thought of as simply selling a unit of the commodity to B. However, spot market equilibrium implies that the price at which this S, B trade happens depends on x_T, since the spot market price is $\frac{x_T}{\kappa_M}$. When x_T is random, B and S thus become exposed to commodity price risk in spot markets, lowering their ex-ante expected utility, despite the fact that their endowments are completely nonrandom. Their price risks exactly offset each other, and do not on net contribute to absorbing any component of aggregate endowment risk, as in the first-best outcome (31): the linear component of aggregate endowment risk is still completely borne by T.

4.2.3 General case

The two examples above illustrate, in a simplified way, most of the wealth distortions generated by spot markets within our system. By rearranging (19), we can decompose wealth further by separating consumer i's wealth exposures to her own shock x_i, and her exposure to the aggregate shock across all other agents $\sum_{j \neq i} x_j$; we show in Appendix A.2 that we can write:

$$W_i^{SpotEqm} = m_i + \underbrace{\psi x_i}_{Wealth\ Shock} - \underbrace{\frac{x_i^2}{\sum_{j=1}^N \kappa_j}\left(1 - \frac{\kappa_i}{2\sum_{j=1}^N \kappa_j}\right)}_{Liquidity\ Taking} + \underbrace{\frac{1}{2}\frac{\kappa_i}{\sum_{j=1}^N \kappa_j}\frac{\left(\sum_{-i} x_j\right)^2}{2\sum_{j=1}^N \kappa_j}}_{Market\ Making} - \underbrace{\frac{x_i \sum_{-i} x_j}{\sum_{j=1}^N \kappa_j}\left(1 - \frac{\kappa_i}{\sum_{j=1}^N \kappa_j}\right)}_{Product\ Risk} \tag{34}$$

The "liquidity taking" term is negative and quadratic in x_i; this can be thought of as a convex cost that i pays to other agents for absorbing her inventory shock, generalizing the exposure of T in the examples above. This cost is larger when x_i^2 is larger, when aggregate liquidity $\sum_{j=1}^N \kappa_j$ is low, and when i's inventory capacity κ_i is small relative to aggregate liquidity. The "market making" term is a positive exposure to the squared aggregate shock of others, $\left(\sum_{-i} x_j\right)^2$, generalizing the exposure of M in the examples; i makes more profits from market making when i's inventory capacity κ_i is large relative to aggregate inventory. The "product risk" term depends on the relative signs of x_i and $\sum_{-i} x_j$: if i buys when others are also buying, she pays more for what she is buying, generalizing the price risks borne by B and S in our examples.

5 Price Controls

When financial markets are incomplete, *price controls* in spot markets can be Pareto-improving, because the benefits they induce for risk-sharing can outweigh their costs for allocative efficiency.

We assume a policymaker can set a price ceiling p_{ceil}, which is constant and does not depend on the realization of $\boldsymbol{x}$. The price ceiling imposes an upper bound on market prices, which induces costless rationing when it is binding. Formally, suppose the unconstrained market clearing price exceeds p_{ceil}. Let $\mathcal{A}_B$ be the set of agents who purchase at p_{ceil}, that is, $q_i(p_{ceil}) > 0$, and let $\mathcal{A}_S$ be those agents with $q_i(p_{ceil}) < 0$. We assume all trade occurs at p_{ceil}, total trade volume equals total supply at p_{ceil}, and trade volume is rationed across buyers according to their relative demands at p_{ceil}. Formally,

$$q_i^{ceil}(p_{ceil}) = q_i(p_{ceil}) \;\; \forall i \in \mathcal{A}_S \tag{35}$$

$$q_i^{ceil}(p_{ceil}) = q_i(p_{ceil}) \left(-\frac{\sum_{i \in \mathcal{A}_S} q_i(p_{ceil})}{\sum_{i \in \mathcal{A}_B} q_i(p_{ceil})} \right) \;\; \forall i \in \mathcal{A}_B \tag{36}$$

Analogously, a price floor p_{floor} sets a lower bound on market prices; when binding, total trade volume equals total demand at p_{floor}, and quantity is rationed across sellers according to relative supply amounts:

$$q_i^{floor}(p_{floor}) = q_i(p_{floor}) \left(-\frac{\sum_{i \in \mathcal{A}_B} q_i(p_{floor})}{\sum_{i \in \mathcal{A}_S} q_i(p_{floor})} \right) \;\; \forall i \in \mathcal{A}_S \tag{37}$$

Evaluating welfare under price controls is straightforward. Given any inventory shocks $\boldsymbol{x}$, we solve for equilibrium prices, imposing price controls if they bind. We then use (35) and (36), and their analogs for price floors, to calculate equilibrium quantities; we then plug prices and quantities into (12) to calculate agents' equilibrium wealth levels, and thus CARA-utility levels, for any realization of $\boldsymbol{x}$. Expected utility under the price control regime is then calculated by integrating over the distribution of shocks.

Numerical Example. Suppose there are two consumers, with $\psi = 0$, $\alpha = 2$, $\kappa = 1$, and symmetrically distributed inventory shocks $x_1, x_2 \sim N(0, \sigma^2)$ with $\sigma^2 = 0.45$. Consumers are fully symmetric, so their ex-ante expected utilities are always identical. Figure 1 plots consumers' expected utility, in unconstrained spot markets (blue) and under varying levels of symmetric price controls (red), where we set a price ceiling $p_{ceil} = \bar{p}$ and a price floor $p_{floor} = -\bar{p}$. Price controls can be Pareto-improving: both consumers' expected utility is higher, for any value of $\bar{p}$ greater than around 1.2, relative to free spot markets.

The intuition for this result is illustrated in Figure 2. Spot markets alone fail to achieve perfect risk-sharing: consumers' marginal utilities are not equalized across states. Panel A plots the normalized difference in spot-equilibrium marginal utility, as a function of inventory shocks $\boldsymbol{x}$:

$$\Delta MU\left(\boldsymbol{x}\right) = \frac{MU_1\left(\boldsymbol{x}\right) - MU_2\left(\boldsymbol{x}\right)}{MU_1\left(\boldsymbol{x}\right) + MU_2\left(\boldsymbol{x}\right)} \tag{38}$$

In unconstrained spot market equilibrium, consumers with more extreme inventory shocks end up with lower wealth and higher marginal utility: 1's MU is greater towards the right and left, and 2's is greater upwards and downwards.[4] Thus, risk-sharing could improve, and aggregate welfare could increase, if wealth could be transferred from 2 to 1 on the right and left sides of the figure, and from 1 to 2 towards the top and bottom.

Panel B plots the net *wealth transfer* induced by the price control policy, defined as:

$$WealthTransfer\left(\boldsymbol{x}\right) = \frac{\left[W_1^{PC}\left(\boldsymbol{x}\right) - W_1^{SpotEqm}\left(\boldsymbol{x}\right)\right] - \left[W_2^{PC}\left(\boldsymbol{x}\right) - W_2^{SpotEqm}\left(\boldsymbol{x}\right)\right]}{2} \tag{39}$$

In words, (39) is a double-difference, measuring whether price controls increase 1's wealth more than they increase 2's wealth. The transfers induced by price controls are directionally consistent with improved risk-sharing. 1 is the net transfer recipient towards the left and right of the plot, and 2 is the net recipient towards the top and bottom. Intuitively, when x_1 is large and positive and x_2 is near 0, 1 is a net seller and 2 is a net buyer. Thus, a price floor tends to increase 1's welfare at the expense of 2, at the cost of some deadweight loss. This does not induce a net transfer ex-ante, because the reverse transfer occurs when x_2 is high and x_1 is near zero. Analogously, price ceilings transfer welfare towards 1 when x_1 is very negative and x_2 is near 0, and towards 2 in the reverse case.

Panel C of Figure 2 plots the *deadweight loss* from price controls, defined simply as the

[4] Intuitively, spot market outcomes are qualitatively similar to no-trade outcomes in this case: if each consumer consumed their endowment, due to quadratic costs, consumers' marginal utilities of wealth would be decreasing in the magnitude of their inventory shocks.

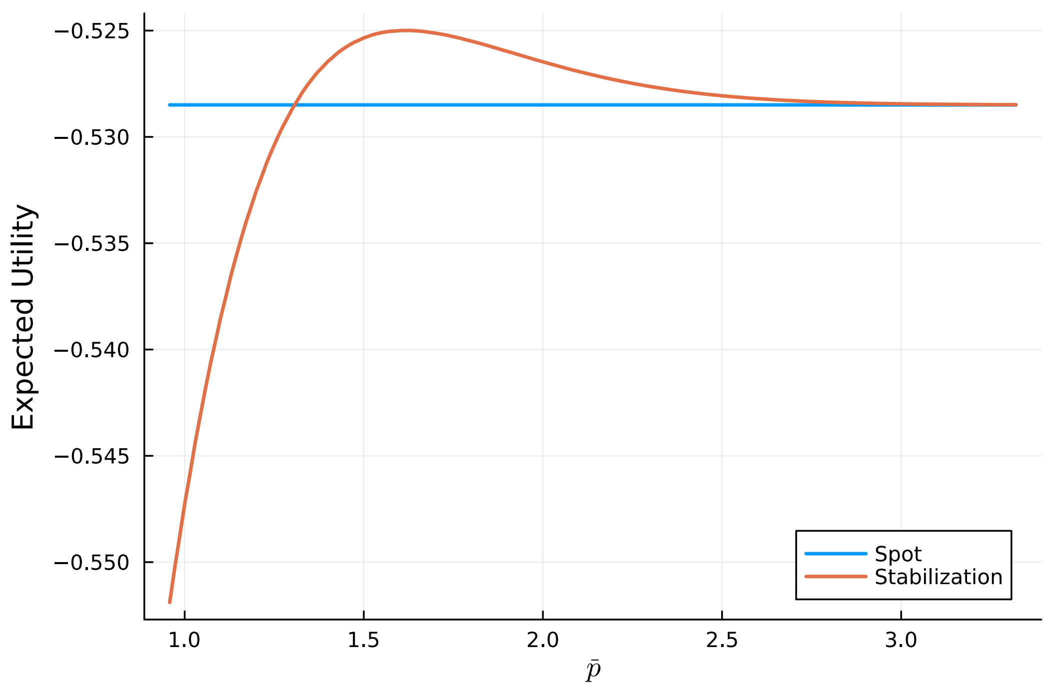
change in total social wealth:

$$DeadweightLoss\left(\boldsymbol{x}\right) = \frac{\left[W_1^{SpotEqm}\left(\boldsymbol{x}\right) + W_2^{SpotEqm}\left(\boldsymbol{x}\right)\right] - \left[W_1^{PC}\left(\boldsymbol{x}\right) + W_2^{PC}\left(\boldsymbol{x}\right)\right]}{2} \quad (40)$$

Deadweight loss is always positive, and tends to be greater when the net wealth transfer induced by price controls is larger. However, it is also in these regions that the risk-sharing benefits of price controls are greatest.

Figure 1: Price Controls and Expected Utility

This figure plots consumers' expected utility in unconstrained spot markets (blue) as well as under symmetric price floors and ceilings (red), where we set $p_{ceil} = -p_{floor} = \bar{p} \geq 0$. Higher values of $\bar{p}$, on the x-axis, thus correspond to looser price controls. For each $\bar{p}$ we calculate utility for a grid of $\boldsymbol{x}$-values, and numerically integrate to find expected utility.

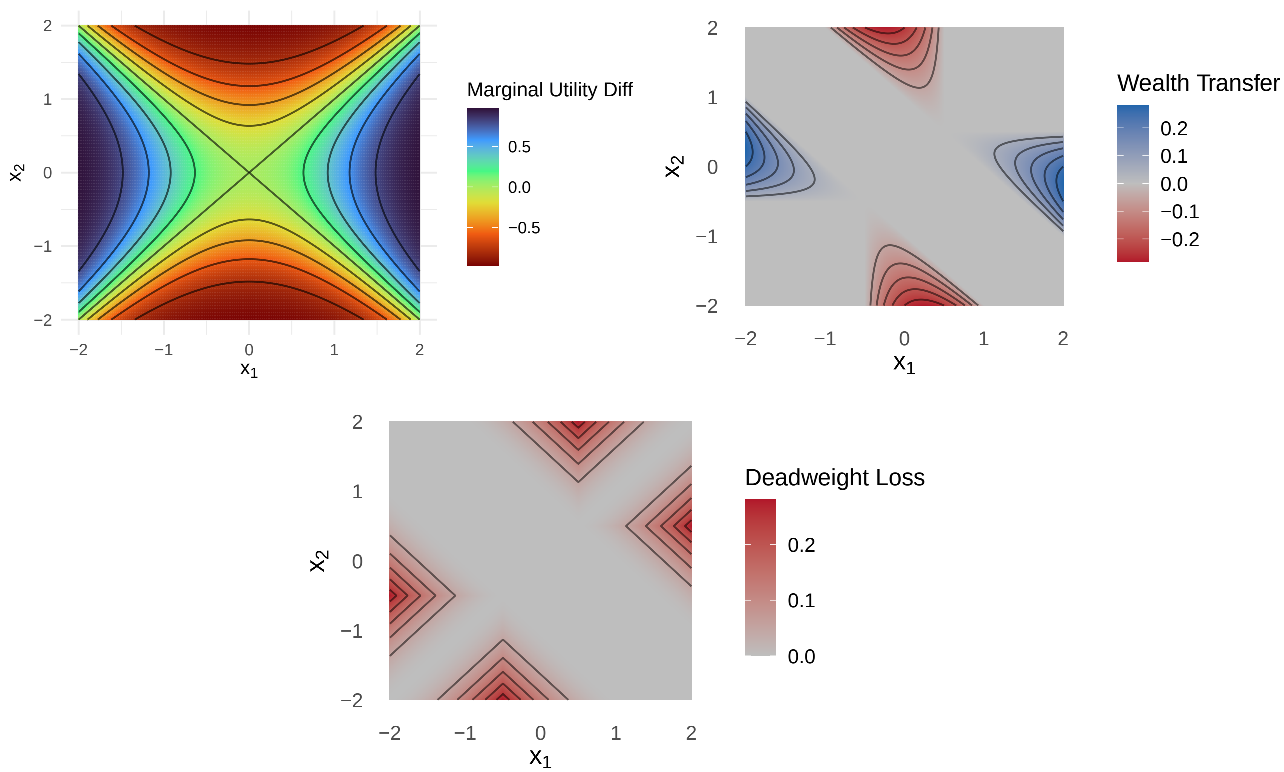


In models without uncertainty, price controls are *transfer* instruments: price ceilings transfer surplus to buyers, and price floors to sellers. Our model introduces a distinct *risk-sharing* role for price controls. In our stylized example, both agents are symmetric, and neither is a buyer or seller on average: price controls induce exactly offsetting transfers across uncertain states of the world, which can have the effect of improving both agents' ex-ante welfare. This is true even though price controls are harmful for allocative efficiency: social aggregate wealth unambiguously decreases whenever price controls are binding.

Market incompleteness is crucial for this result. Complete financial markets perfectly

Figure 2: Price Controls: Mechanisms

In each panel, consumers' inventory shocks x_1 and x_2 are shown on the x and y axes respectively. Panel A shows (38), the normalized difference between 1 and 2's marginal utility of wealth, induced by unconstrained spot market equilibrium outcomes. Panel B plots (39), the net wealth transfer from 2 to 1 induced by price controls, defined as the difference between 1's wealth gain under price controls relative to unconstrained spot market equilibrium, and 2's wealth gain. Price controls can improve risk sharing because they tend to transfer wealth in the direction of MU differences: wealth is transferred to 1 towards the right and left, where 1's marginal utility is higher, and to 2 towards the top and bottom, where 2's marginal utility is higher. Panel C plots (40), price control-induced deadweight loss, defined as total social wealth in spot market equilibrium minus total wealth under price controls. DWL is always positive, and is higher when price controls are more binding and induce larger transfers between agents. In both panels B and C, we consider a symmetric price control $p_{ceil} = -p_{floor} = 0.5$.



equalize agents' marginal utilities across states, leaving no further room for improvements in risk sharing. In incomplete financial markets, agents' marginal utilities may differ across states, leaving room for price stabilization policies to be welfare-improving.

Our results can thus be thought of as an instance of the "theorem of the second best": price controls are unambiguously welfare-reducing in the frictionless complete-market benchmark, but can be welfare-improving in the more realistic setting of incomplete financial markets.

6 Futures Markets

The simplest possible financial security we can introduce in our setting is a *commodity futures contract*. For simplicity, we will analyze a single *cash-settled futures contract* on the commodity. In our model, cash-settled futures attain equivalent outcomes to "physical delivery" contracts; we focus on the cash-settlement case because contracts which transfer wealth across states of the world are more intuitive in our setting.[5]

A futures contract in our setting is a *contract for differences*: A consumer holding a long contract position promises to pay some fixed p^c in the second period, and in exchange receives the uncertain commodity spot price $p^{SpotEqm}(\boldsymbol{x})$. The net monetary payoff to i of buying c_i net contracts at p^c is thus:

$$\left(p^{SpotEqm}(\boldsymbol{x}) - p^c\right) c_i$$

Intuitively, this is valuable because the contract pays out exactly enough money to purchase c units of the commodity, regardless of $p^{SpotEqm}(\boldsymbol{x})$. We call the fixed payment, p^c, the *contract price*. Markets clear through p^c adjusting until aggregate demand for long and short contract positions is equal.

As in the case of Arrow securities, futures contracts are straightforward to analyze in our model because spot market equilibrium prices and quantities are unaffected by any monetary transfers across consumers. This implies that i's monetary payoff, from purchasing c_i contracts, is simply her anticipated spot market equilibrium wealth $W_i^{SpotEqm}$ plus her

[5] Kyle (2007) shows that cash-settled and physical-delivery derivatives are exactly equivalent whenever a set of "microstructure fungibility" conditions hold; our model satisfies these conditions. Essentially, since spot market outcomes are modelled simply through competitive equilibria, it is equivalent to consumer i's budget set whether she receives x units of the commodity through a physical-delivery contract, or xp dollars through a cash-settled contract. Trade quantities vary – if the consumer receives an endowment of x units of the commodity, she must adjust her net trade quantity in spot markets by x – but market clearing in zero-net-supply futures markets imply that these net quantity adjustments cancel out across consumers, leaving net wealths and spot market prices unchanged. This is shown formally in Section 5.4 of Zhang (2022), in a related model to ours. The expositional advantage of using cash-settled contracts in our setting is that, since they only transfer wealth across states of the world, they do not influence consumers' spot market bids, so spot market outcomes are unchanged from the previous section.

contract payoff:

$$W_i^{SpotEqm} + \left(p^{SpotEqm}(\boldsymbol{x}) - p^c\right) c_i \tag{41}$$

Alternatively, using the expression for $W_i^{SpotEqm}(\boldsymbol{x})$ in (19), i's wealth, as a function of $\boldsymbol{x}$ and c_i, is:

$$\psi x_i + \frac{\kappa_i}{\sum_{j=1}^N \kappa_j} \frac{\left(\sum_{j=1}^N x_j\right)^2}{2\sum_{j=1}^N \kappa_j} - x_i \frac{\sum_{j=1}^N x_j}{\sum_{j=1}^N \kappa_j} + \left(p^{SpotEqm}(\boldsymbol{x}) - p^c\right) c_i \tag{42}$$

In the first period, consumers have no information about inventory shocks $\boldsymbol{x}$, so each consumer i simply chooses a scalar quantity c_i of contracts to purchase. Facing futures price p^c, i chooses c_i to maximize her expected utility, over second-period uncertainty in $\boldsymbol{x}$:

$$\max_{c_i} -E\left[\exp\left(-\alpha_i \left(W_i^{SpotEqm}(\boldsymbol{x}) + \left(p^{SpotEqm}(\boldsymbol{x}) - p^c\right) c_i\right)\right)\right] \tag{43}$$

Each consumer i's optimization problem defines a contract demand curve $c_i(p^c)$. Futures contracts are in net zero supply: each long contractholder is paid by a short contractholder. Market clearing thus requires contract demand to sum to zero across consumers:

$$\sum c_i(p^c) = 0 \tag{44}$$

An equilibrium in contract markets is thus defined by a scalar contract price p^c under which (44) holds.

Recall that we used μ_i and σ_i^2 to refer to the mean and variance of i's inventory shock x_i respectively. We impose the additional normalization that the mean of the aggregate inventory shock is zero:

$$E\left[\sum_{i=1}^N x_i\right] = \sum_{i=1}^N \mu_i = 0 \tag{45}$$

Appendix B.2 shows that (45) is without loss of generality, because any nonzero average in inventory shocks can be absorbed into the definition of ψ. (45) is convenient because it implies that i's expected spot market purchase quantity is $-\mu_i$: agents with $\mu_i > 0$ expect to be sellers, and agents with $\mu_i < 0$ expect to be buyers.[6] In addition, (45) and (15) imply

[6] To see this, note from (16) that:

$$E\left[q_i^{SpotEqm}(\boldsymbol{x})\right] = -E[x_i] + \frac{\kappa_i}{\sum_i \kappa_i} \sum_{i=1}^N E[x_i] = -\mu_i$$

Since we have

$$\sum_{i=1}^N E[x_i] = \sum_{i=1}^N \mu_i = 0$$

that the expected equilibrium spot price is $E\left[p^{SpotEqm}\left(\boldsymbol{x}\right)\right]=\psi$.

Technically, our setting is tractable because, given normally distributed inventory shocks, spot equilibrium wealths $W_i^{SpotEqm}$ are second-order polynomials in independent normal random variables, and thus follow *noncentered chi-squared* distributions. CARA-expected utility is analytically solvable for such distribution, generalizing the standard CARA-normal wealth setting. We impose an additional regularity condition to ensure expected utility remains finite.

Assumption 1. *We assume:*

$$\kappa_i > \sum_{j=1}^{N}\kappa_j\left(2-\frac{\sum_{j=1}^{N}\kappa_j}{\alpha_i\sigma_i^2}\right)\ \forall i \tag{46}$$

Intuitively, when α_i and σ_i^2 are too large relative to κ_i, the integral underlying (43) is not finite. (46) imposes a lower bound on κ_i, to avoid this case; this lower bound is negative and thus trivial when α_i,σ_i^2 are small, and becomes more binding as α_i,σ_i^2 increase.

The following proposition shows that equilibrium outcomes in futures markets can be solved analytically, for arbitrary combinations of κ_i,σ_i,μ_i. The equilibrium expressions are complex in general case; we will illustrate outcomes through various simpler examples.

Proposition 3. *There exists a unique equilibrium in spot and futures markets. Consumer i's futures demand is an affine function of the contract price p^c, i's ex-ante mean inventory shock μ_i, and the parameter ψ:*

$$c_i\left(p^c\right)=\underbrace{\beta_i\left(\psi-p^c\right)}_{Price\ \ Sensitivity}-\underbrace{\mu_i}_{Hedging\ \ Pressure}+\underbrace{\zeta_i\psi}_{Risk\ \ Premium} \tag{47}$$

The equilibrium futures price is:

$$p^c=\left(1+\frac{\sum_i\zeta_i}{\sum_i\beta_i}\right)\psi \tag{48}$$

The parameters β_i and ζ_i are positive constants which depend on the primitives $\alpha_j,\kappa_j,\sigma_j$:

$$\beta_i\equiv\frac{\Sigma_2}{\alpha_iA_2^2+\frac{\Sigma_2}{\Sigma_1}\frac{\alpha_i}{\left(\sum_{j=1}^{N}\kappa_j\right)^2}}>0,\ \zeta_i\equiv\frac{\left(\frac{\alpha_i}{\sum_{j=1}^{N}\kappa_j}\right)\left(\frac{\alpha_i}{\sum_{j=1}^{N}\kappa_j}+\frac{1}{\sum_{j\neq i}\sigma_j^2}\right)\frac{1}{\Sigma_1}}{\alpha_iA_2^2+\frac{\Sigma_2}{\Sigma_1}\frac{\alpha_i}{\left(\sum_{j=1}^{N}\kappa_j\right)^2}}>0 \tag{49}$$

$$\Sigma_1\equiv 1/\sigma_i^2+\frac{\alpha_i\kappa_i}{\sum_{j=1}^{N}\kappa_j}\left(\frac{1}{\sum_{j=1}^{N}\kappa_j}-\frac{2}{\kappa_i}\right)>0 \tag{50}$$

$$\Sigma_2 \equiv \frac{1}{\sum_{j\neq i}\sigma_j^2} + \frac{\alpha_i\kappa_i}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{\alpha_i^2\left(\frac{\kappa_i}{\left(\sum_{j=1}^N\kappa_j\right)^2} - \frac{1}{\sum_{j=1}^N\kappa_j}\right)^2}{1/\sigma_i^2 + \frac{\alpha_i\kappa_i}{\sum_{j=1}^N\kappa_j}\left(\frac{1}{\sum_{j=1}^N\kappa_j} - \frac{2}{\kappa_i}\right)} > 0 \tag{51}$$

$$A_2 \equiv \frac{\frac{\alpha_i}{\sum_{j=1}^N\kappa_j}\left(\frac{\kappa_i}{\left(\sum_{j=1}^N\kappa_j\right)^2} - \frac{1}{\sum_{j=1}^N\kappa_j}\right)}{\Sigma_1} - \frac{1}{\sum_{j=1}^N\kappa_j} \tag{52}$$

The introduction of futures contracts is Pareto-improving, weakly increasing all consumers' expected utilities, relative to spot markets alone.

6.1 Symmetric Consumers

Suppose consumers have identical values of $\sigma_i, \alpha_i, \kappa_i$, but differ in their mean inventory shock μ_i. In the absence of futures markets, substituting $x_i = \mu_i + \epsilon_i$ into consumers' spot market wealths (19), we get:

$$W_i^{SpotEqm} = \psi\left(\mu_i + \epsilon_i\right) + \frac{\left(\sum_{j=1}^N \mu_i + \epsilon_i\right)^2}{2N^2\kappa} - x_i\frac{\sum_{j=1}^N \mu_i + \epsilon_i}{N\kappa}$$

Ignoring the constant $\psi\mu_i$ term, and using that $\sum_j \mu_j = 0$, we have:

$$W_i = \underbrace{\psi\epsilon_i}_{Wealth\ Shock} + \underbrace{\frac{\left(\sum_{j=1}^N\epsilon_j\right)^2}{2N^2\kappa}}_{Market\ Making} - \underbrace{\epsilon_i\frac{\sum_{j=1}^N\epsilon_j}{N\kappa}}_{Unexpected\ Product\ Risk} - \underbrace{\mu_i\frac{\sum_{j=1}^N\epsilon_j}{N\kappa}}_{Expected\ Product\ Risk} \tag{53}$$

The first term can be thought of as i's idiosyncratic "wealth shock" from her endowment, and the second positive term represents i's profits from absorbing part of the aggregate inventory shock. The third and fourth terms depend on the relative signs of ϵ_i and μ_i, and the aggregate inventory shock $\sum_{j=1}^N \epsilon_j$. Intuitively, wealth is higher if i's inventory shock happens to have opposite sign to others' shocks; it is more expensive to sell $(\epsilon_i > 0, \mu_i > 0)$ when others are also selling $(\sum_{j=1}^N \epsilon_j > 0)$, since from (15), the equilibrium price depends linearly on the aggregate inventory shock. This effect divides into a μ_i term, which reflects the expected component of i's purchases, and an ϵ_i term reflecting the unexpected component.

Suppose then that futures contracts are available. From Proposition 3, consumers' equilibrium futures demands are simply:

$$c_i(p^c) = \beta(\psi - p^c) - \mu_i + \zeta\psi \tag{54}$$

where β, ζ are constant across consumers. Plugging the equilibrium futures price (48) into contract demand, i's equilibrium contract purchase quantity is simply:

$$c_i(p^c) = -\mu_i$$

That is, each consumer exactly purchases contracts equal to their expected inventory shocks μ_i. Consumers' equilibrium wealth with futures is:

$$W_i^{SpotEqm} + \left(p^{SpotEqm}(\boldsymbol{x}) - p^c\right)c$$

Expression (15) for $p^{SpotEqm}(\boldsymbol{x})$ implies:

$$p^{SpotEqm}(\boldsymbol{x}) - \psi = -\frac{\sum_{i=1}^N x_i}{\sum_{i=1}^N \kappa_i} = -\frac{\sum_{i=1}^N \epsilon_i}{N\kappa} \tag{55}$$

Substituting for $p^{SpotEqm}(\boldsymbol{x})$, and for p^c using (48), and rearranging slightly, we have:

$$W_i^{SpotEqm} + \left(p^{SpotEqm}(\boldsymbol{x}) - p^c\right)c =$$
$$\psi\epsilon_i + \frac{\left(\sum_{j=1}^N \epsilon_j\right)^2}{2N^2\kappa} - \epsilon_i\frac{\sum_{j=1}^N \epsilon_j}{N\kappa} - \mu_i\frac{\sum_{j=1}^N \epsilon_j}{N\kappa} + (-\mu_i)\left(-\frac{\sum_{i=1}^N \epsilon_i}{N\kappa}\right)p^{SpotEqm}(\boldsymbol{x}) + \mu_i\frac{\zeta}{\beta}\psi$$

$$W_i^{SpotEqm} + \left(p^{SpotEqm}(\boldsymbol{x}) - p^c\right)c = \psi\epsilon_i + \frac{\left(\sum_{j=1}^N \epsilon_j\right)^2}{2N^2\kappa} - \epsilon_i\frac{\sum_{j=1}^N \epsilon_j}{N\kappa} + \mu_i\frac{\zeta}{\beta}\psi \tag{56}$$

Comparing (56) to (53), futures contracts simply eliminate the "expected product risk" term from all agents' equilibrium wealths.

There is a simple semi-technical intuition for this effect, related to our example in Subsection 4.2.2 above. In (53), the "expected product risk" term for agent i can be written as:

$$\mu_i\frac{\sum_{j=1}^N \epsilon_j}{N\kappa} = (-\mu_i)\left(p^{SpotEqm}(\boldsymbol{x}) - \psi\right)$$

Like B and S in our earlier example, i faces wealth risk due to *directional exposure to prices*. i expects to buy $-\mu_i$ units of the commodity, at the noisy price $p^{SpotEqm}(x_1 \ldots x_N)$; if $-\mu_i > 0$, i is worse off if prices tend to be higher. But this price risk can be perfectly hedged by buying $-\mu_i$ units of a contract, which pays $p^{SpotEqm}(\boldsymbol{x})$. Moreover, contract markets naturally clear if all agents follow this strategy, since $\sum_j \mu_j = 0$ – the total expected trade quantity across agents sums to zero. The role of futures in this setting is thus to allow agents to mutually self-insure "expected product risk": expected buyers take long positions, expected sellers take short positions, and thus all agents enter spot markets with zero directional exposure to $p^{SpotEqm}(\boldsymbol{x})$.

Futures contracts cannot implement the first-best outcome: agents' equilibrium wealths in (56) still contain wealth shocks, market-making profits, and unanticipated product risk, all terms which do not show up in agents' first-best wealths (11). This is because they do not span the entire state space. We showed in Subsection 4.1 that the first-best outcome requires financial contracts which nontrivially span the entire state space. Futures contracts span the space of equilibrium spot prices; from (55), $p^{SpotEqm}(\boldsymbol{x})$ is a function of the *aggregate* inventory shock $\sum_{j=1}^{N} \epsilon_j$.

The fact that futures are Pareto-improving relative to spot markets alone seems intuitive; however, it relies strongly on our assumption of quasilinearity. Since agents have no income effects in our setting, wealth redistribution through futures contracts does not influence spot market outcomes. Thus, consumer i can guarantee at least $W_i^{SpotEqm}$, pointwise in $\boldsymbol{x}$, simply by not participating in futures markets, regardless of what other agents do in financial markets. As a result, each agent i is made weakly better off, in expected utility terms, from the introduction of futures markets. In the general case, adding incomplete financial markets need not lead to a Pareto improvement relative to pure spot markets. The key issue is income effects: if financial trades reallocate wealth, spot market outcomes can change in ways that make some agents worse off than they would be in the absence of financial markets.

6.2 The Futures Risk Premium

We now depart from our symmetric example and return to the general model. We can define the *futures risk premium* as the difference between the futures contract p^c and the expected spot price, $E\left[p^{SpotEqm}(\boldsymbol{x})\right] = \psi$. Rearranging (48), we have:

$$p^c - \psi = \frac{\sum_i \zeta_i}{\sum_i \beta_i}\psi. \tag{57}$$

Hedging Demand and Risk Premia. An idea dating to Keynes (1930) is that risk premia in futures markets are generated by hedging pressure: if producers tend to short futures to hedge risk, demand pressure causes futures prices to fall below expected spot prices. Hirshleifer (1990) shows that this idea does not work in general equilibrium: in a complete-markets model where futures span the entire state space, hedging pressure alone does not generate futures risk premia.

This insight also holds in our incomplete-markets setting, and we illustrate a new intuition driving it. As we discussed in the symmetric example, hedging pressure shows up in *individual traders' demand*. This generalizes to the full model: expression (54) implies that, regardless of parameters, traders' contract demand increases one-for-one with expected spot purchases μ_i. Thus, contract *trade volumes* are very generally driven by hedging demand: commodity buyers tend to take long futures positions, and commodity sellers tend to take short positions. In the simple symmetric example of Subsection 6.1, hedging pressures are the *only* drivers of contract trade.

But hedging pressures do not create risk premia in equilibrium because *spot market clearing* implies that hedging pressure always cancels out across agents. Intuitively, if i anticipates purchasing $-\mu_i$ units of the commodity in spot markets, (54) shows that she hedges by entering a long contract position of the same size. But, since $\sum_i \mu_i = 0$,[7] the sum of these hedging trades is always zero: for every agent that expects to buy, there is another that expects to sell. Thus, when we calculate aggregate contract demand by summing agents' $c_i(p^c)$ functions, the μ_i terms always add to zero, and vanish from the sum. This holds both in our symmetric example and in the general model.

As Hirshleifer observes, futures risk premia can emerge from hedging pressure when futures market participation is restricted. Many models derive premia simply by imposing participation constraints (Hirshleifer, 1988*a*,*b*; Gorton, Hayashi and Rouwenhorst, 2013; Acharya, Lochstoer and Ramadorai, 2013; Goldstein, Li and Yang, 2014; Goldstein and Yang, 2022), often excluding commodity buyers from participation in futures markets; in such models, risk premia tend to reflect producers' hedging demands and financial speculators' risk absorption capacity, possibly in addition to other forces which are absent from our model, such as informational frictions.

Risk Premia and Consumption Betas. Somewhat surprisingly, (57) implies that the sign of the futures risk premium always matches the sign of ψ. Intuitively, risk premia are driven by *consumption betas*. When $\psi > 0$, the commodity has positive average value. Commodity spot prices are high when the aggregate inventory shock $\sum_{i=1}^{N} \epsilon_i$ is low, aggregate wealth is low, and thus marginal utility is high. Futures contracts, which have payoffs linked to spot prices, are thus valuable insurance against consumption risk, leading to a negative risk premium, $p^c > \psi$. Hirshleifer (1990) shows that this holds in a complete-market model

[7] Once again, while we assumed $\sum_i \mu_i = 0$ in (45), this assumption is purely a normalization. Appendix B.2 shows that any model in which aggregate inventory shocks are not mean-zero can be written as a model with mean-zero inventory shocks, where we simply redefine ψ to absorb the average inventory shock.

of futures; this logic extends quite cleanly to our incomplete-market model.

6.3 Price Controls and Futures Contracts

Futures contracts do not eliminate the potential welfare gains from price controls. A simple way to see this is that, in the example of Section 5 where price controls improve welfare, futures contracts would not be traded even if they were available, since agents are ex-ante identical, and thus markets can only clear if all agents' contract demand is identically 0.

In our setting, equilibria with only futures contracts are constrained-efficient, in the sense of Geanakoplos and Polemarchakis (1986): no Pareto improvements are possible through first-stage reallocations of money and futures contracts alone, maintaining competitive equilibrium outcomes in spot markets. In incomplete-markets models, equilibria are generally constrained-inefficient (Geanakoplos and Polemarchakis, 1986), because reallocating financial assets in first-stage markets generally affects later-stage spot prices; these price changes induce income transfers, which the planner can exploit to improve social welfare.[8] This mechanism is absent in our model because we assume quasilinear second-stage utility, so spot prices are unaffected by financial market activity. Instead, we show that the planner can directly intervene in *spot markets*: interventions which distort spot market allocations relative to competitive-equilibrium outcomes, but improve risk-sharing, can be Pareto-improving.

7 Conclusion

The technical contribution of this paper is a tractable model of general equilibrium with incomplete markets. By assuming quasilinear preferences between wealth and goods, we eliminate income effects; equilibria in spot markets are thus unique and invariant to outcomes in financial markets. Under the additional assumptions that shocks are normally distributed, preferences are quadratic, and agents have CARA utility, equilibrium outcomes in both spot and financial markets can be solved fully analytically.

Using the model, we show that markets for goods and markets for risk play distinct roles. Markets for goods efficiently reallocate scarce resources; markets for risk distribute the resultant wealth optimally across agents. When markets for risk are incomplete, society generates wealth optimally, but redistributes it suboptimally. Policy interventions in spot markets that would reduce welfare in the absence of uncertainty, such as price controls, can be Pareto-improving in our setting, since they force goods markets to perform some of the

[8] See also Magill and Quinzii (2002, pp. 263–274) for a discussion of generic constrained inefficiency.

risk-sharing functions that financial markets fail to deliver.

Markets are powerful forces for good, and our goal in this paper is not to argue that they are not generally better than realistic alternatives. We aim simply to highlight an underappreciated feature of classic general equilibrium theory: incompleteness in financial markets, even in an otherwise textbook environment, can rationalize interventions in spot markets that depart sharply from the laissez-faire benchmark.

References

Acharya, Viral V., Lars A. Lochstoer, and Tarun Ramadorai. 2013. “Limits to arbitrage and hedging: Evidence from commodity markets.” *Journal of Financial Economics*, 109(2): 441–465.

Borch, Karl. 1962. “Equilibrium in a Reinsurance Market.” *Econometrica: Journal of the Econometric Society*, 30(3): 424–444.

Dávila, Eduardo, and Andreas Schaab. 2022. “Welfare assessments with heterogeneous individuals.” National Bureau of Economic Research.

Dávila, Eduardo, and Andreas Schaab. 2023. “Welfare Accounting.” National Bureau of Economic Research.

Dávila, Eduardo, and Anton Korinek. 2018. “Pecuniary externalities in economies with financial frictions.” *The Review of Economic Studies*, 85(1): 352–395.

Duffie, Darrell. 2010. *Dynamic asset pricing theory.* Princeton University Press.

Geanakoplos, John. 1990. “An introduction to general equilibrium with incomplete asset markets.” *Journal of Mathematical Economics*, 19(1-2): 1–38.

Geanakoplos, John, and Heraklis M. Polemarchakis. 1986. “Existence, regularity, and constrained suboptimality of competitive allocations.”

Goldstein, Itay, and Liyan Yang. 2022. “Commodity financialization and information transmission.” *The Journal of Finance*, 77(5): 2613–2667.

Goldstein, Itay, Yan Li, and Liyan Yang. 2014. “Speculation and hedging in segmented markets.” *The Review of Financial Studies*, 27(3): 881–922.

Gorton, Gary B., Fumio Hayashi, and K. Geert Rouwenhorst. 2013. “The fundamentals of commodity futures returns.” *Review of Finance*, 17(1): 35–105.

Grossman, Sanford J., and Joseph E. Stiglitz. 1980. “On the impossibility of informationally efficient markets.” *The American Economic Review*, 70(3): 393–408.

Hart, Oliver D. 1975. “On the optimality of equilibrium when the market structure is incomplete.” *Journal of Economic Theory*, 11(3): 418–443.

Hirshleifer, David. 1988*a*. “Residual risk, trading costs, and commodity futures risk premia.” *The Review of Financial Studies*, 1(2): 173–193.

Hirshleifer, David. 1988*b*. "Risk, futures pricing, and the organization of production in commodity markets." *Journal of Political Economy*, 96(6): 1206–1220.

Hirshleifer, David. 1990. "Hedging pressure and futures price movements in a general equilibrium model." *Econometrica: Journal of the Econometric Society*, 411–428.

Keynes, John Maynard. 1930. *A Treatise on Money.* Macmillan & Company.

Krishnamurthy, Arvind. 2003. "Collateral constraints and the amplification mechanism." *Journal of Economic Theory*, 111(2): 277–292.

Kyle, Albert S. 2007. "Cash settlement, price manipulation, and the Modigliani-Miller Theorem."

Lorenzoni, Guido. 2008. "Inefficient credit booms." *The Review of Economic Studies*, 75(3): 809–833.

Magill, Michael, and Martine Quinzii. 2002. *Theory of incomplete markets.* Vol. 1, MIT press.

Marshall, Alfred. 1920. *Principles of Economics.* . 8th ed., Macmillan and Co.

Stiglitz, Joseph E. 1982. "The inefficiency of the stock market equilibrium." *The Review of Economic Studies*, 49(2): 241–261.

Wilson, Robert. 1968. "The theory of syndicates." *Econometrica: Journal of the Econometric Society*, 119–132.

Zhang, Anthony Lee. 2022. "Competition and manipulation in derivative contract markets." *Journal of Financial Economics*, 144(2): 396–413.

Internet Appendix

A Proofs and Supplementary Material for Section 4

A.1 Derivation of Spot Market Equilibrium Wealth (19)

Copying (12), consumers' wealth is:

$$W_i = \psi\left(x_i + q_i\right) - \frac{\left(x_i + q_i\right)^2}{2\kappa_i} - pq_i \tag{58}$$

For convenience, we define:

$$X \equiv \sum_{j=1}^{N} x_j,\ \ K \equiv \sum_{j=1}^{N} \kappa_j$$

We can then define equilibrium quantities (16) and prices (15) as:

$$p^{SpotEqm}\left(x_1 \dots x_N\right) - \psi = -\frac{X}{K} \tag{59}$$

$$x_i + q_i^{SpotEqm}\left(x_1 \dots x_N\right) = \frac{\kappa_i}{K} X \tag{60}$$

Rearranging (58) slightly,

$$W_i = \psi x_i - \frac{\left(x_i + q_i\right)^2}{2\kappa_i} - \left(p - \psi\right) q_i$$

Substituting (59) and (60), we have:

$$W_i = \psi x_i - \frac{\left(\frac{\kappa_i}{K} X\right)^2}{2\kappa_i} - \left(-\frac{X}{K}\right)\left(\frac{\kappa_i}{K} X - x_i\right)$$

Simplifying, we have:

$$W_i = \psi x_i + \frac{\kappa_i}{K}\frac{X^2}{2K} - \frac{X}{K} x_i$$

Substituting the definitions of X and K, we have (19).

A.2 Rearranging Spot Market Equilibrium Wealth Into (34)

We can write (19) as:

$$W_i = m_i + \psi x_i + \frac{\kappa_i}{\sum_{j=1}^{N} \kappa_i} \frac{\left(x_i + \sum_{j \neq i} x_j\right)^2}{2 \sum_{j=1}^{N} \kappa_i} - x_i \frac{\left(x_i + \sum_{j \neq i} x_j\right)}{\sum_{j=1}^{N} \kappa_i}$$

$$W_i = m_i + \psi x_i + \frac{\kappa_i}{\sum_{j=1}^{N} \kappa_i} \frac{x_i^2 + 2x_i \sum_{j \neq i} x_j + \left(\sum_{j \neq i} x_j\right)^2}{2 \sum_{j=1}^{N} \kappa_i} - \frac{x_i^2 + x_i \sum_{j \neq i} x_j}{\sum_{j=1}^{N} \kappa_i}$$

Grouping terms, we get (34).

B Proofs and Supplementary Material for Section 6

B.1 Proof of Proposition 3

Copying (61), given a futures price p^c, agents choose contract quantity c_i to solve:

$$\max_{c_i} -E\left[\exp\left(-\alpha_i\left(W_i^{SpotEqm}(\boldsymbol{x}) + \left(p^{SpotEqm}(\boldsymbol{x}) - p^c\right)c_i\right)\right)\right], \tag{61}$$

We will use $E[U_i]$ as shorthand for consumers' expected utility, (61). Plugging in (15) and (19) for the spot market equilibrium price and wealth $p^{SpotEqm}(\boldsymbol{x})$ and $W_i^{SpotEqm}(\boldsymbol{x})$, we have:

$$E[U_i] = -E\left[\exp\left(-\alpha_i\left(m_i + \psi x_i + \frac{\kappa_i}{\sum_{j=1}^N \kappa_j}\frac{\left(\sum_{j=1}^N x_j\right)^2}{2\sum_{j=1}^N \kappa_j} - x_i\frac{\sum_{j=1}^N x_j}{\sum_{j=1}^N \kappa_j} + \left(\psi - \frac{\sum_{j=1}^N x_j}{\sum_{j=1}^N \kappa_j} - p^c\right)c_i\right)\right)\right]. \tag{62}$$

We assumed that $x_i = \mu_i + \epsilon_i$, with the normalization from (45) that $\sum_{i=1}^N \mu_i = 0$. Expected utility can thus be rearranged to:

$$E[U_i] = -E\left[\exp\left(-\alpha_i\left(m_i + \psi(\mu_i + \epsilon_i) + \frac{\kappa_i}{2}\left(\frac{\sum_{j=1}^N \epsilon_j}{\sum_{j=1}^N \kappa_j}\right)^2 - (\mu_i + \epsilon_i)\frac{\sum_{j=1}^N \epsilon_j}{\sum_{j=1}^N \kappa_j} + \left(\psi - \frac{\sum_{j=1}^N \epsilon_j}{\sum_{j=1}^N \kappa_j} - p^c\right)c_i\right)\right)\right]. \tag{63}$$

We assumed that the ϵ_i variables are independent mean-zero normal random variables, with possibly different variances σ_i^2. Let $\epsilon_{-i} \equiv \sum_{j\neq i}\epsilon_j$; we thus have that ϵ_{-i} is also normally distributed, $N(0, \sum_{j\neq i}\sigma_j^2)$, and is independent of ϵ_i. Thus, we can write (63) as an integral over the distributions of ϵ_i and ϵ_{-i}:

$$E[U_i] = -e^{-\alpha_i m_i}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp\left[-\alpha_i\left(\psi(\mu_i + \epsilon_i) + \frac{\kappa_i}{2}\left(\frac{\epsilon_i + \epsilon_{-i}}{\sum_{j=1}^N \kappa_j}\right)^2 - (\mu_i + \epsilon_i)\frac{\epsilon_i + \epsilon_{-i}}{\sum_{j=1}^N \kappa_j} + \left(\psi - \frac{\epsilon_i + \epsilon_{-i}}{\sum_{j=1}^N \kappa_j} - p^c\right)c_i\right)\right] dF(\epsilon_i)\, dF(\epsilon_{-i}), \tag{64}$$

where $F(\epsilon_i)$ is the cumulative distribution function of ϵ_i and $F(\epsilon_{-i})$ is the cumulative distribution function of ϵ_{-i}.

While (64) is complex, it is simply a double integral over a second-order polynomial in the independent normal random variables ϵ_i and ϵ_{-i}, and this is analytic in general. We first state a general lemma characterizing the solutions to such integrals; the lemma is proved in Appendix B.1.1 below.

Lemma 1. *Let $x \sim N(0, \sigma^2)$. If:*

$$1/\sigma^2 - 2D > 0 \tag{65}$$

Then:

$$\int_{-\infty}^{+\infty} (Ax + B) \exp\left[Dx^2 + Ex + G\right] dF(x) = \frac{\frac{AE}{1/\sigma^2 - 2D} + B}{\sigma\sqrt{1/\sigma^2 - 2D}} \exp\left[G + \frac{E^2}{2/\sigma^2 - 4D}\right] \tag{66}$$

Otherwise, if $1/\sigma^2 - 2D \leq 0$, the integral on the LHS of (66) diverges.

We characterize the optimal contract quantity $c_i(p^c)$ through consumers' first-order condition. Differentiating (64) and setting to 0, we have:

$$0 = \frac{\partial E[U_i]}{\partial c_i} = \alpha_i e^{-\alpha_i m_i} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left(\psi - \frac{\epsilon_i + \epsilon_{-i}}{\sum_{j=1}^N \kappa_j} - p^c\right)$$
$$\exp\left[-\alpha_i \left(\psi(\mu_i + \epsilon_i) + \frac{\kappa_i}{2}\left(\frac{\epsilon_i + \epsilon_{-i}}{\sum_{j=1}^N \kappa_j}\right)^2 - (\mu_i + \epsilon_i)\frac{\epsilon_i + \epsilon_{-i}}{\sum_{j=1}^N \kappa_j} + \left(\psi - \frac{\epsilon_i + \epsilon_{-i}}{\sum_{j=1}^N \kappa_j} - p^c\right) c_i\right)\right]$$
$$dF(\epsilon_i)\, dF(\epsilon_{-i}) \tag{67}$$

We then evaluate the double integral in (67) by applying Lemma 1 twice. Rearranging slightly, we can see that the inner integral takes the form in Lemma 1, with:

$$\sigma_1 = \sigma_i,\ A_1 = -\frac{1}{\sum_{j=1}^N \kappa_j},\ B_1 = \psi - p^c - \frac{\epsilon_{-i}}{\sum_{j=1}^N \kappa_j},\ D_1 = -\alpha_i\left(\frac{\kappa_i}{2}\left(\frac{1}{\sum_{j=1}^N \kappa_j}\right)^2 - \frac{1}{\sum_{j=1}^N \kappa_j}\right), \tag{68}$$

$$E_1 = -\alpha_i \left(\psi + \frac{\kappa_i \epsilon_{-i}}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{\mu_i + \epsilon_{-i}}{\sum_{j=1}^N \kappa_j} - \frac{c_i}{\sum_{j=1}^N \kappa_j} \right),$$

$$G_1 = -\alpha_i \left(\mu_i \psi + \frac{\kappa_i \epsilon_{-i}^2}{2\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{\mu_i \epsilon_{-i}}{\sum_{j=1}^N \kappa_j} + \left(\psi - p^c - \frac{\epsilon_{-i}}{\sum_{j=1}^N \kappa_j} \right) c_i \right). \quad (69)$$

Hence, applying Lemma 1 to evaluate the inner integral in (67), we have:

$$\frac{\partial E\left[U_i\right]}{\partial c_i} = \alpha_i e^{-\alpha_i m_i} \int_{-\infty}^{+\infty} \frac{\frac{A_1 E_1}{1/\sigma_i^2 - 2c_1} + B_1}{\sigma_i \sqrt{1/\sigma_i^2 - 2D_1}} \exp\left[G_1 + \frac{E_1^2}{2/\sigma_i^2 - 4D_1} \right] dF\left(\epsilon_{-i}\right), \quad (70)$$

And the integral converges if and only if:

$$\Sigma_1 \equiv 1/\sigma_i^2 - 2D_1 = 1/\sigma_i^2 + \frac{\alpha_i \kappa_i}{\sum_{j=1}^N \kappa_j} \left(\frac{1}{\sum_{j=1}^N \kappa_j} - \frac{2}{\kappa_i} \right) > 0. \quad (71)$$

Now, we can substitute the expressions in (68) and (69), to write (70) as:

$$\frac{\partial E\left[U_i\right]}{\partial c_i} =$$

$$\frac{\alpha_i e^{-\alpha_i m_i}}{\sigma_i \sqrt{\Sigma_1}} \int_{-\infty}^{+\infty} \left(\frac{\frac{\alpha_i}{\sum_{j=1}^N \kappa_j} \left(\psi + \frac{\kappa_i \epsilon_{-i}}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{\mu_i + \epsilon_{-i}}{\sum_{j=1}^N \kappa_j} - \frac{c_i}{\sum_{j=1}^N \kappa_j} \right)}{\Sigma_1} + \psi - p^c - \frac{\epsilon_{-i}}{\sum_{j=1}^N \kappa_j} \right)$$

$$\exp\left[-\alpha_i \left(\mu_i \psi + \frac{\kappa_i \epsilon_{-i}^2}{2\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{\mu_i \epsilon_{-i}}{\sum_{j=1}^N \kappa_j} + \left(\psi - p^c - \frac{\epsilon_{-i}}{\sum_{j=1}^N \kappa_j} \right) c_i \right) + \right.$$

$$\left. \frac{\alpha_i^2 \left(\psi + \frac{\kappa_i \epsilon_{-i}}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{\mu_i + \epsilon_{-i}}{\sum_{j=1}^N \kappa_j} - \frac{c_i}{\sum_{j=1}^N \kappa_j} \right)^2}{2\Sigma_1} \right] dF\left(\epsilon_{-i}\right). \quad (72)$$

Once again, (72) takes the form in Lemma 1, with:

$$\sigma_2^2 = \sum_{j\neq i} \sigma_j^2,\;\; A_2 = \frac{\frac{\alpha_i}{\sum_{j=1}^N \kappa_j}\left(\frac{\kappa_i}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{1}{\sum_{j=1}^N \kappa_j}\right)}{\Sigma_1} - \frac{1}{\sum_{j=1}^N \kappa_j},$$

$$B_2 = \frac{\frac{\alpha_i}{\sum_{j=1}^N \kappa_j}\left(\psi - \frac{\mu_i + c_i}{\sum_{j=1}^N \kappa_j}\right)}{\Sigma_1} + \psi - p^c, \quad (73)$$

$$D_2 = -\frac{\alpha_i \kappa_i}{2\left(\sum_{j=1}^N \kappa_j\right)^2} + \frac{\alpha_i^2\left(\frac{\kappa_i}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{1}{\sum_{j=1}^N \kappa_j}\right)^2}{2\Sigma_1}, \quad (74)$$

$$E_2 = \alpha_i\left(\frac{\mu_i + c_i}{\sum_{j=1}^N \kappa_j}\right) + \frac{\alpha_i^2\left(\psi - \frac{\mu_i + c_i}{\sum_{j=1}^N \kappa_j}\right)\left(\frac{\kappa_i}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{1}{\sum_{j=1}^N \kappa_j}\right)}{\Sigma_1}. \quad (75)$$

We will not write out the full expression for G_2 here: it is a constant in the exponent of (72), so it simply scales $\frac{\partial E[U_i]}{\partial c_i}$, and affects neither the sign nor the roots of $\frac{\partial E[U_i]}{\partial c_i}$, and thus will not affect the first-order condition.

We thus apply Lemma 1 to evaluate the outer integral, finding that:

$$\frac{\partial E\left[U_i\right]}{\partial c_i} = \frac{\alpha_i e^{-\alpha_i m_i}}{\sigma_i\sqrt{\Sigma_1}} \frac{\frac{A_2 E_2}{\frac{1}{\sum_{j\neq i}\sigma_j^2} - 2D_2} + B_2}{\sqrt{\sum_{j\neq i}\sigma_j^2}\sqrt{\frac{1}{\sum_{j\neq i}\sigma_j^2} - 2D_2}} \exp\left[G_2 + \frac{E_2^2}{\frac{2}{\sum_{j\neq i}\sigma_j^2} - 4D_2}\right], \quad (76)$$

and the integral converges if and only if:

$$\Sigma_2 \equiv \frac{1}{\sum_{j\neq i}\sigma_j^2} - 2D_2 = \frac{1}{\sum_{j\neq i}\sigma_j^2} + \frac{\alpha_i\kappa_i}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{\alpha_i^2\left(\frac{\kappa_i}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{1}{\sum_{j=1}^N \kappa_j}\right)^2}{1/\sigma_i^2 + \frac{\alpha_i\kappa_i}{\sum_{j=1}^N \kappa_j}\left(\frac{1}{\sum_{j=1}^N \kappa_j} - \frac{2}{\kappa_i}\right)} > 0. \quad (77)$$

Using (76), we can then solve for contract demand using the first-order condition. We can separate out components of (76) that are always positive:

$$K(c_i) = \frac{\alpha_i e^{-\alpha_i m_i}}{\sigma_i\sqrt{\Sigma_1}} \frac{\exp\left[G_2 + \frac{E_2^2}{2\Sigma_2}\right]}{\sqrt{\sum_{j\neq i}\sigma_j^2}\Sigma_2^{3/2}} > 0. \tag{78}$$

In order for the FOC $\frac{\partial E[U_i]}{\partial c_i} = 0$ to hold, we must then have:

$$\frac{A_2 E_2}{\frac{1}{\sum_{j\neq i}\sigma_j^2} - 2D_2} + B_2 = 0$$

or, using (77),

$$A_2 E_2 + B_2 \Sigma_2 = 0$$

Substituting the expressions in (73), (74), and (75), we can write:

$$A_2 E_2 + B_2 \Sigma_2 =$$

$$A_2 \left(\alpha_i \left(\frac{\mu_i + c_i}{\sum_{j=1}^N \kappa_j} \right) \left(1 - \frac{\alpha_i \left(\frac{\kappa_i}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{1}{\sum_{j=1}^N \kappa_j} \right)}{\Sigma_1} \right) + \frac{\alpha_i^2 \psi \left(\frac{\kappa_i}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{1}{\sum_{j=1}^N \kappa_j} \right)}{\Sigma_1} \right) +$$

$$\left(\frac{\frac{\alpha_i \psi}{\sum_{j=1}^N \kappa_j}}{\Sigma_1} + \psi - p^c - \frac{\alpha_i \frac{\mu_i + c_i}{\left(\sum_{j=1}^N \kappa_j\right)^2}}{\Sigma_1} \right) \Sigma_2$$

$$= \alpha_i \left(\frac{\mu_i + c_i}{\sum_{j=1}^N \kappa_j} \right) \left(A_2 \left(1 - \frac{\alpha_i \left(\frac{\kappa_i}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{1}{\sum_{j=1}^N \kappa_j} \right)}{\Sigma_1} \right) - \frac{1}{\Sigma_1} \frac{1}{\sum_{j=1}^N \kappa_j} \Sigma_2 \right) +$$

$$\frac{\psi}{\Sigma_1} \left(A_2 \alpha_i^2 \left(\frac{\kappa_i}{\left(\sum_{j=1}^N \kappa_j\right)^2} - \frac{1}{\sum_{j=1}^N \kappa_j} \right) + \left(\frac{\alpha_i}{\sum_{j=1}^N \kappa_j} \right) \left(\frac{1}{\sum_{j\neq i}\sigma_j^2} - 2D_2 \right) \right) +$$

$$\left(\frac{1}{\sum_{j\neq i}\sigma_j^2} - 2D_2 \right) (\psi - p^c)$$

$$= (\mu_i + c_i) \left(-\alpha_i A_2^2 - \frac{\Sigma_2}{\Sigma_1} \frac{\alpha_i}{\left(\sum_{j=1}^N \kappa_j\right)^2} \right) + \frac{\psi}{\Sigma_1} \left(\frac{\alpha_i}{\sum_{j=1}^N \kappa_j} \right) \left(\frac{\alpha_i}{\sum_{j=1}^N \kappa_j} + \frac{1}{\sum_{j \neq i} \sigma_j^2} \right) + \Sigma_2 \left(\psi - p^c\right). \tag{79}$$

Expression (79) is a *linear*, downwards-sloping function of c_i, since the coefficient:

$$-\alpha_i A_2^2 - \frac{\Sigma_2}{\Sigma_1} \frac{\alpha_i}{\left(\sum_{j=1}^N \kappa_j\right)^2}$$

is strictly negative. Thus, for any p^c, there is a unique value of c_i which sets (79) to zero:

$$c_i \left(p_c\right) = \frac{\Sigma_2 \left(\psi - p^c\right) + \left(\frac{\alpha_i}{\sum_{j=1}^N \kappa_j} \right) \left(\frac{\alpha_i}{\sum_{j=1}^N \kappa_j} + \frac{1}{\sum_{j \neq i} \sigma_j^2} \right) \frac{1}{\Sigma_1} \psi}{\alpha_i A_2^2 + \frac{\Sigma_2}{\Sigma_1} \frac{\alpha_i}{\left(\sum_{j=1}^N \kappa_j\right)^2}} - \mu_i \tag{80}$$

We have thus shown that (80) solves $A_2 E_2 + B_2 \Sigma_2 = 0$. Based on our definition of $K\left(c_i\right)$ in (78), this implies that $c_i\left(p_c\right)$ is also the unique solution to the first-order condition:

$$0 = \frac{\partial E\left[U_i\right]}{\partial c_i} = K\left(c_i\right)\left(A_2 E_2 + B_2 \Sigma_2\right),$$

In addition, since $K\left(c_i\right)$ is strictly positive, and (79) is linear and downwards-sloping in c_i, when $c_i < c_i\left(p_c\right)$, $\partial E[U_i]/\partial c_i > 0$, and when $c_i > c_i\left(p_c\right)$, $\partial E[U_i]/\partial c_i < 0$. Thus, the first-order condition (80) is necessary and sufficient for optimality: $c_i\left(p_c\right)$ in (80) is the unique maximizer of $E\left[U_i\right]$, given p^c. (80) thus characterizes consumer contract demand, given p^c.

Expression (80) rearranges to (47), our expression for contract demand, in Proposition 3:

$$c_i\left(p_c\right) = \beta_i\left(\psi - p^c\right) - \mu_i + \zeta_i \psi, \; \beta_i > 0, \zeta_i > 0. \tag{81}$$

Where we define β_i and ζ_i as in (82), repeated here for clarity:

$$\beta_i = \frac{\Sigma_2}{\alpha_i A_2^2 + \frac{\Sigma_2}{\Sigma_1} \frac{\alpha_i}{\left(\sum_{j=1}^N \kappa_j\right)^2}} > 0, \; \zeta_i = \frac{\left(\frac{\alpha_i}{\sum_{j=1}^N \kappa_j} \right) \left(\frac{\alpha_i}{\sum_{j=1}^N \kappa_j} + \frac{1}{\sum_{j \neq i} \sigma_j^2} \right) \frac{1}{\Sigma_1}}{\alpha_i A_2^2 + \frac{\Sigma_2}{\Sigma_1} \frac{\alpha_i}{\left(\sum_{j=1}^N \kappa_j\right)^2}} > 0, \tag{82}$$

and expressions (50), (51), and (52) for Σ_1, Σ_2, A_2 correspond to the expressions (71), (73), and (77) which we previously defined in this appendix. Ultimately, expressions (81) and (82) express contract demand as analytic, though complex, functions of the input parameters $\alpha_j, \kappa_j, \sigma_j$. This concludes the proof of the expression for contract demand, (47).

Equilibrium contract prices. The equilibrium contract price, (48), follows immediately from imposing market clearing (44) on contract demands (47), using (45) that $\sum_{i=1}^{N} \mu_i = 0$.

Futures contracts are Pareto-improving. To show that futures contracts are Pareto-improving relative to spot markets, note that consumers can simply choose $c_i = 0$ in (61), thus attaining their spot market utility:

$$-E\left[\exp\left(-\alpha_i\left(W_i^{SpotEqm}\left(\boldsymbol{x}\right)\right)\right)\right].$$

Thus, the optimal choice of c_i must achieve weakly greater expected utility than i achieves in spot markets alone.

B.1.1 Proof of Lemma 1

Proof. We can substitute the normal PDF for $dF(x)$ on the LHS of (66), to obtain:

$$\int (Ax+B)\exp\left[Dx^2+Ex+G\right]dF(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(Ax+B)\exp\left[\left(D-\frac{1}{2\sigma^2}\right)x^2+Ex+G\right]dx. \quad (83)$$

Clearly, the integral on the RHS converges if $1/\sigma^2 - 2D \leq 0$. We can further rearrange the integral to:

$$\int (Ax+B)\exp\left[Dx^2+Ex+G\right]dF(x) = \frac{\exp\left[G+\frac{E^2}{2/\sigma^2-4D}\right]}{\sigma\sqrt{1/\sigma^2-2D}}\int_{-\infty}^{+\infty}(Ax+B)\sqrt{\frac{1/\sigma^2-2D}{2\pi}}\exp\left[-\left(\frac{1}{2\sigma^2}-D\right)\left(x-\frac{E}{1/\sigma^2-2D}\right)^2\right]dx. \quad (84)$$

Note that $\sqrt{\frac{1/\sigma^2-2D}{2\pi}}\exp\left[-\left(\frac{1}{2\sigma^2}-D\right)\left(x-\frac{E}{1/\sigma^2-2D}\right)^2\right]$ is the density function of the normal distribution $X \sim N\left(\frac{E}{1/\sigma^2-2D}, \frac{1}{1/\sigma^2-2D}\right)$. Hence, we can write the integral on the RHS of (84) as:

$$\int_{-\infty}^{+\infty}(Ax+B)\sqrt{\frac{1/\sigma^2-2D}{2\pi}}\exp\left[-\left(\frac{1}{2\sigma^2}-D\right)\left(x-\frac{E}{1/\sigma^2-2D}\right)^2\right]dx = \mathbb{E}_X\left[AX+B\right] = \frac{AE}{1/\sigma^2-2D}+B.$$

Plugging in to (84) and rearranging, we get:

$$\int (Ax+B)\exp\left[Dx^2+Ex+G\right]dF(x) = \frac{\frac{AE}{1/\sigma^2-2D}+B}{\sigma\sqrt{1/\sigma^2-2D}}\exp\left[G+\frac{E^2}{2/\sigma^2-4D}\right].$$

□

B.2 Justification of Mean-Zero Inventory Shocks

It is without loss of generality to assume (45) – that the aggregate inventory shock, $\sum_{i=1}^{N} x_i$, has mean 0 – because of a redundancy in the way we specify consumers' wealth W_i is specified: the linear term ψ and the inventory shock x_i can be "renormalized" in a way that keeps consumer utility unchanged. We state this in the following simple claim, which we prove in Appendix B.2.1 below.

Claim 1. Consumer i's wealth function, (4), can equivalently be written as:

$$W_i = \tilde{\psi}(\tilde{x}_i+q_i) - \frac{(\tilde{x}_i+q_i)^2}{2\kappa_i}$$

where:

$$\tilde{\psi} \equiv \psi + A \tag{85}$$

$$\tilde{x}_i \equiv x_i - 2\kappa_i A \tag{86}$$

Claim 1 implies that we can "renormalize" the constant term ψ, increasing it by any constant A across all consumers, as long as we correspondingly renormalize inventory shocks x_i. Intuitively, since ψx_i is simply a linear component of preferences, having a higher ψ is equivalent to having a lower inventory shock x_i, by an amount that depends on κ_i. Since the scaling in (86) is linear in A, this immediately implies that, for any set of original inventory shocks x_i which do not have 0 mean across consumers, we can find some A to normalize ψ and inventory shocks, which leads the resultant inventory shocks to have zero mean across consumers. This choice of A is simply:

$$\sum_{i=1}^{N} E[\tilde{x}_i] = 0$$

$$\implies \sum_{i=1}^{N} E[x_i] = A\sum_{i=1}^{N} 2\kappa_i$$

$$\implies A = \frac{\sum_{i=1}^{N} E\left[\tilde{x}_i\right]}{2\sum_{i=1}^{N} \kappa_i}$$

As a result, it is completely without loss of generality – that is, it is simply a renormalization of agents' utility functions – to assume that the expected sum of inventory shocks across consumers is 0, as we do in (45).

B.2.1 Proof of Claim 1

Copying (4) and ignoring the price term, consumers' wealth as a function of x_i and q_i is:

$$W_i = C_i + \psi\left(x_i + q_i\right) - \frac{\left(x_i + q_i\right)^2}{2\kappa_i} \tag{87}$$

where we add a constant C_i, to emphasize that constant shifts in wealth do not affect our model outcomes. Suppose we renormalize the inventory shock x_i as:

$$\tilde{x}_i \equiv x_i - M$$

Hence, (87) becomes:

$$W_i = C_i + \psi\left(\tilde{x}_i + M + q_i\right) - \frac{\left(\tilde{x}_i + M + q_i\right)^2}{2\kappa_i}$$

Expanding and grouping terms, we have:

$$W_i = C_i + \psi M - \frac{M^2}{2\kappa_i} + \left(\psi - \frac{M}{2\kappa_i}\right)\left(\tilde{x}_i + q_i\right) - \frac{\left(\tilde{x}_i + q_i\right)^2}{2\kappa_i}$$

We can write this as:

$$W_i = \tilde{C}_i + \tilde{\psi}_i\left(\tilde{x}_i + q_i\right) - \frac{\left(\tilde{x}_i + q_i\right)^2}{2\kappa_i}$$

where:

$$\tilde{C}_i \equiv C_i + \psi M - \frac{M^2}{2\kappa_i}$$

$$\tilde{\psi} \equiv \psi - \frac{M}{2\kappa_i}$$

We ignore the $\tilde{C}_i$ term, since it is a constant scaling factor for W_i under the assumption of CARA utility, and can be factored out. Defining

$$A \equiv -\frac{M}{2\kappa_i}$$

we have proved Claim 1.