Are Bank Capital Ratios Too High or Too Low? Risk Aversion, Incomplete Markets, and Optimal Capital Structure*

Douglas Gale and Onur Özgür

New York University

September 13, 2004

Abstract

We study the effect of relative risk aversion on optimal capital structure in a general-equilibrium model of intermediation with incomplete markets.

JEL Classification: D5, G2.

1 Introduction

Since the inception of the first Basel Accord, capital regulation has been one of the main policy tools used by central bankers to control financial instability.

*The financial support of the C.V. Starr Center for Applied Economics at NYU and the National Science Foundation under Grant No. SES 0095109 is gratefully acknowledged. E-mail: douglas.gale@nyu.edu; onur.ozgur@nyu.edu.

There are two main functions of bank capital. The first is the **incentive function**. Because of the debt-like nature of their liabilities, banks have an incentive to engage in risk shifting or asset substitution, that is, to take on excessive risk knowing that the downside risk is born by their creditors (depositors). Requiring banks to hold a minimum ratio of capital to assets reduces the bank's incentive to take risk. The second function of bank capital is the **risk sharing function**. Capital acts like a buffer that may offset the losses of the creditors (depositors) and allows for the orderly liquidation and disposal of assets in the event of financial distress.

Even if these two functions explain why banks hold capital, they do not explain why bank capital should be regulated. Gale (2003) has argued that the benefits of improved incentives and risk sharing can and will be internalized by the bank. More precisely, to the extent that bank capital improves the bank's incentives to choose the correct level of risk or improves risk sharing between shareholders and depositors, competition for customers will lead the bank to choose the capital structure that optimizes the depositors' welfare.

It is sometimes argued, by contrast, that the existence of **deposit insurance** provides a relevant externality and a rationale for regulation. To the extent that the provision of deposit insurance encourages excessive risk taking, it may be efficient for the regulator to require higher capital levels to offset this distortion. This is, at best, a weak argument in favor of capital regulation. As Gale (2003) has pointed out, two wrongs don't make a right, and it would be better to eliminate the distortionary effects of deposit insurance than to use it as an excuse for further regulation.

In order to find an externality that will justify the imposition of capital regulation, one has to go beyond the microeconomic analysis of a single bank and consider the efficiency of risk sharing in the financial sector or the economy as a whole. Financial fragility provides one possible justification. The failure of one bank imposes costs on other banks, either directly in terms of default on its obligations or indirectly through its effect on asset prices. To the extent that this pecuniary externality is not internalized by the bank, there may be an argument for intervention. One has to be very careful, of course, because pecuniary externalities do not necessarily impose deadweight costs. Allen and Gale (2004) provide an analogue of the classical theorems of welfare economics for a general-equilibrium model of financial intermediation and financial markets — they show that the incidence of financial crises is constrained efficient as long as markets for aggregate risk are complete — so the existence of pecuniary externalities does not imply a role for welfare-improving intervention by the central bank.

The model of Allen and Gale (2004) does not contain a role for bank capital, but Gale (2003) shows how it can easily be incorporated in a model of banking. Gale (2004) extends this model to consider the efficiency of bank capital under conditions analogous to Allen and Gale (2004). Briefly, one can summarize the conclusions as follows:

- With complete markets, the equilibrium capital structure is efficient and indeterminate. This is a variant of the Modigliani-Miller Theorem: complete markets are a perfect substitute for capital, so there are many Pareto-optimal capital structures.

- When markets for capital are incomplete, the capital structure is determinate but may still be Pareto-optimal, for example, if there is a representative intermediary, that is, each intermediary serves a representative sample of the population. Essentially, each intermediary acts as a central planner and achieves first-best risk sharing in autarky.

- When markets for capital are incomplete and intermediaries are ex post heterogeneous, the missing markets are essential and the equilibrium capital structure is constrained-inefficient. This result is analogous to the well known result that general equilibrium with incomplete markets is generically constrained-inefficient (Geanakoplos and Polemarchakis, 1986).

Thus, incomplete markets can in principle provide us with a rationale for capital regulation. However, there is an important difference between demonstrating the existence of a potentially welfare-improving policy intervention and identifying or characterizing the welfare-improving policy. What is the central bank supposed to do? General equilibrium models are complex and even in simple cases the interaction of general-equilibrium effects may contradict our intuitions, which tend to be based on partial-equilibrium reasoning. Gale (2004) makes this point concretely, by analyzing an example in which depositors have a degree of risk aversion equal to unity, i.e., logarithmic utility functions. Our intuition tells us that, because banks do not take into account the effect of their actions on financial fragility or asset price volatility, they should be forced to hold more capital. This is the "axiom" that lies at the heart of the Basel system. Gale (2004) shows, by contrast, that, whereas the laisser-faire equilibrium with incomplete markets has a capital ratio that is lower than the first best, in order to improve welfare it is necessary to lower the capital stock still further. In other words, there is too much capital not too little.

How are we to make sense of this counter-intuitive result and under what circumstances, if any, would minimum capital requirements be welfare-improving? The answers to these two questions are the subject of this paper.

The remainder of the paper is organized as follows. Section 2 sketches the model from Gale (2004). Section 3 describes the equilibrium and the bank's contracting problem. Section 4 contains the results of our simulations, which show the dependence of the comparative statics properties on the degree of relative risk aversion.

2 A model of risk

As a vehicle for our analysis we use a variant of the model found in Gale (2003). Time is divided into three periods or **dates** denoted by $t = 0, 1, 2$. At each date there is a single all-purpose good that can be used for consumption or for investment. There are two assets, a short-term asset that matures after one period and a long-term asset that matures after two periods.

- The **short asset** is represented by a storage technology: one unit of the good invested at date t yields one unit at date $t + 1$.

- The **long asset** is represented by a constant-returns-to-scale investment technology: one unit invested at date 0 yields $R > 1$ units at date 2 (and nothing at date 1).

In choosing the optimal combination of the two assets, there is a tradeoff between liquidity and the rate of return. The short asset provides greater liquidity (immediate access to returns), but the long asset provides a higher return per unit invested. These properties of the assets play a critical role when markets are incomplete.

The economic agents in this economy are divided into two groups: risk-averse consumers who provide a demand for liquidity and risk-neutral investors who supply the capital. There is a continuum of identical, risk-neutral investors with unit mass. Although investors are risk neutral, we assume that their consumption must be non-negative at each date. Otherwise, it is impossible to make sense of limited liquidity. The investor's utility function is defined by

$$u(c_0, c_1, c_2) = \rho c_0 + c_1 + c_2,$$

where $c_t \geq 0$ denotes the investor's consumption at date $t = 0, 1, 2$. The constant $\rho > R$ represents the investor's opportunity cost of funds. An investor's endowment consists of an unboundedly large amount of the good at date 0 and nothing at dates 1 and 2.

There is a continuum of identical, risk-averse consumers with unit mass. Each consumer has an endowment of one unit of the good at date 0 and nothing at dates 1 and 2. At date 1, each consumer receives a preference shock: with probability λ he becomes an **early consumer**, who only values consumption at date 1, and with probability $1-\lambda$ he becomes a **late consumer**, who only values consumption at date 2. A consumer's period utility function $U : \mathbf{R}_{+} \rightarrow \mathbf{R}$ is twice continuously differentiable and satisfies the usual neoclassical properties,

$$U'(c) > 0, U''(c) < 0, \lim_{c \searrow 0} U'(c) = \infty.$$

The consumer's risk aversion, together with his uncertainty about the preference shock (early or late), creates a demand for insurance. By pooling these risks, it is possible to provide liquidity to consumers at date 1 while holding a smaller amount of the short asset.

We assume that the fraction of early consumers is equal to the probability λ that an individual agent turns out to be an early consumer. The probability λ is a random variable satisfying

$$\lambda = \begin{cases} \lambda_H & \text{w. pr. } 0.5 \\ \lambda_L & \text{w. pr. } 0.5, \end{cases}$$

where $0 < \lambda_L < \lambda_H < 1$. This is the only source of aggregate uncertainty.

We assume the simplest form of heterogeneity among intermediaries: at date 1 all the consumers in a particular intermediary are of the same type, early or late. Then the proportion of early consumers in intermediary i is denoted by λ_i and defined by

$$\lambda_i = \begin{cases} 1 & \text{w. pr. } \lambda_s \\ 0 & \text{w. pr. } 1 - \lambda_s \end{cases}$$

in states $s = H, L$.

We assume that there are only spot markets for the good and for assets. The good serves as the numeraire at each date. The price of any asset in positive net supply is equal to unity at date 0, since assets are produced then. At date 1, there is a spot market for the long asset.

All uncertainty is resolved at the beginning of date 1, when the state of nature λ is observed and each agent discovers whether he is an early or late consumer.

3 Equilibrium

A consumer deposits his entire endowment with a single intermediary who offers him a consumption contract $c = \{(c_{1s}, c_{2s}) : s = H, L\}$ in exchange, where c_{1s} is consumption offered to an early consumer at date 1 in state s and c_{2s} is the consumption offered to a late consumer at date 2 in state s.

The intermediary writes a contract $e = \{(e_0, e_{1s}, e_{2s}) : s = H, L\}$ with investors, where $e_0 \geq 0$ is the amount of capital invested at date 0, $e_{1s} \geq 0$ is the amount of the good promised to investors in state s if all the depositors are early consumers and $e_{2s} \geq 0$ is the amount of the good promised in state s if all the depositors are late consumers. There is no loss of generality in assuming that the payments to investors are made at date 2.

Since there is a perfectly elastic supply of capital, the investors do not get any surplus, but the contract e must satisfy the participation constraint

$$\sum_{s=H,L} \frac{1}{2} \{\lambda_s e_{1s} + (1 - \lambda_s) e_{2s}\} \geq \rho e_0, \tag{1}$$

which says that the expected utility of future consumption offered by the intermediary must equal the utility of the capital supplied at date 0.

Free entry ensures that the intermediaries receive no profit. Competition among intermediaries forces them to maximize the expected utility of the typical depositor,

$$\sum_{s=H,L} \{\lambda_s U(c_{1s}) + (1-\lambda_s)U(c_{2s})\} \tag{2}$$

subject to the intermediary's budget constraints and the investors' participation constraint.

At date 0 the intermediary has $1+e_0$ of the good to invest in short and long assets. We denote the investment in the short asset by θ and the investment in the long asset by $1+e_0-\theta$. There are no short sales, so the intermediary's budget constraint is satisfied if $0 \leq \theta \leq 1+e_0$.

Let p_s denote the implicit price of future consumption in state s at date 1, so the value of the long asset is $p_s R$. In each state $s = H, L$, the intermediary's budget constraint at date 1 is

$$c_{1s} + p_s e_{1s} = \theta + p_s(1+e_0-\theta)R \tag{3}$$

if the depositors are early consumers and

$$p_s\left(c_{2s} + e_{2s}\right) = \theta + p_s(1+e_0-\theta)R \tag{4}$$

if the depositors are late consumers. In both cases, the left hand side is the present value of the consumption promised to the depositors and investors and the right hand side is the liquidated value of the intermediary's portfolio.

The intermediary's decision problem is to choose (c, e, θ) to maximize the objective function (2) subject to the participation constraint (1) and the budget constraints (3) and (4).

Market clearing at dates 0 and 2 is guaranteed by intermediary's budget constraint and the fact that there is essentially one good. Market clearing at date 1 requires $p_s \leq 1$. If $p_s > 1$, the short asset dominates the long asset between dates 1 and 2, no one will be willing to hold the long asset, and the market for the long asset cannot clear. Note that $p_s < 1$ implies that the long asset dominates the short asset between date 1 and date 2, so no one will hold the short asset. This is compatible with market clearing if the demand for consumption is equal to the supply of the short asset. If demand for the good is strictly less than the supply of the short asset, some intermediaries must be willing to hold the short asset between dates 1 and 2. It will be optimal to hold the short asset if $p_s = 1$. In short, market-clearing requires that either

$$\lambda_s c_{1s} + (1 - \lambda_s) c_{2s} = \theta \text{ and } p_s \leq 1 \tag{5}$$

or

$$\lambda_s c_{1s} + (1 - \lambda_s) c_{2s} < \theta \text{ and } p_s = 1. \tag{6}$$

We define an equilibrium to consist of a contract (c, e, θ) and a price vector $p = (p_H, p_L)$ such that (c, e, θ) solves the intermediary's decision problem, given the price vector p, and the market-clearing conditions (5) and (6) are satisfied.

A necessary condition for (c, e, θ) to be a solution to the intermediary's decision problem is that, for the given values of θ and e_0, the consumption allocation

(c, e) solves the problem:

$$\begin{array}{ll} \max & \sum_{s=H,L} \lambda_s U(w_s - p_s e_{1s}) + (1-\lambda_s) U\left(\frac{w_s}{p_s} - e_{2s}\right) \\ \text{s.t.} & \sum_{s=H,L} \lambda_s e_{1s} + (1-\lambda_s) e_{2s} \geq 2\rho e_0, \end{array} \quad (7)$$

where $w_s \equiv \theta + (1 + e_0 - \theta)R$. The necessary and sufficient conditions for a solution to (7) are

$$U'(w_s - p_s e_{1s}) p_s \geq \mu, \ \ s = H, L \quad (8)$$

with equality if $e_{1s} > 0$, and

$$U\left(\frac{w_s}{p_s} - e_{2s}\right) \geq \mu, \ \ s = H, \quad (9)$$

with equality if $e_{2s} > 0$. Note that in the case where $p_s = 1$, we will have $c_{1s} = c_{2s}$ and $e_{1s} = e_{2s}$.

4 Results

To illustrate the equilibrium properties of the model, we use a parametric example. The period utility $U(c)$ is assumed to satisfy

$$U(c) = \frac{1}{1-\sigma} c^{1-\sigma},$$

where the constant $\sigma > 0$ is the degree of relative risk aversion. The liquidity shocks are assumed to be $\lambda_H = 0.6$ and $\lambda_L = 0.4$. The opportunity cost of funds for the investors is $\rho = 2$ and the return on the long asset is $R = 1.8$. These parameter values produce results that are typical of other simulations. Our main interest here is in tracing the effects of the degree of relative risk aversion on the equilibrium values of the endogenous variables. These relationships are described in Figures 1 through 4.

Figure 1 shows the equilibrium consumption allocation $(c_{1H}, c_{2H}, c_{1L}, c_{2L})$ for different values of σ. As expected, when σ is small, there is a large variation in consumptions across aggregate states and intermediary types. As σ increases, consumption levels gradually converge. For values of σ above 2.5, the distribution of consumption has a two-point support $c_{1H} < c_{2H} = c_{1L} = c_{2L}$. As σ increases further, these two values gradually converge to a single point. Consumption is at its lowest level for early consumers in state H. The more risk averse consumers are, the more important it is to raise the value of c_{1H}. An intermediary does this by adjusting its portfolio and changing its capital structure as σ increases. There is a cost, of course, which is that the other consumption level $c_{2H} = c_{1L} = c_{2L}$ declines.

Figure 2 shows the equilibrium returns to capital $(e_{1H}, e_{2H}, e_{1L}, e_{2L})/e_0$ for different values of σ. Although in principle e_{ts} can be positive for any type t and state s, in practice we see that most of the payments are very small. For small values of σ, most of the repayment to the investors comes from e_{1H}. Since c_{1H} is the lowest consumption level, we can see that the capital structure is increasing consumption risk. For high values of σ, the risk-sharing effect is dominant, and the repayment is largely coming from e_{2H}, which corresponds to the highest consumption state c_{2H}.

Figure 3 shows us the level of capital e_0 for different values of σ. The striking feature of the graph is how rapidly the level of capital increases for values of σ

greater than 2.5. Figure 3 also shows the first-best level of "capital," that is, the date-0 contribution from risk neutral investors in a Pareto-efficient allocation. The equilibrium and Pareto-optimal levels are both in the neighborhood of 0.05 for small values of σ but diverge for larger values of σ. For very high values of σ, the first-best level of capital declines toward zero: since fluctuations in capital are being eliminated by holding more of the short asset, there is no need for capital to provide insurance.

Figure 4 shows us the equilibrium asset prices associated with different values of σ. The value of the long asset is highest when the demand for liquidity is low. As σ increases, the amount of liquidity in the economy increases and this raises prices in both aggregate states. For values of σ above 2.5, there is so much liquidity at date 1 that $p_L = 1$ and some intermediaries hold the short asset between dates 1 and 2.

The equilibrium described in these figures is inefficient because markets are incomplete. In particular, intermediaries would like to insure against liquidity shocks. Instead, intermediaries that need liquidity at date 1 are forced to sell their long assets at the random price p_s which is low when they are most likely to need liquidity (λ_s is high) and high when they are least likely to need it (λ_s is low). This is the opposite of what efficient risk sharing requires. Now it is neither surprising nor especially interesting that an equilibrium with incomplete markets is Pareto-inefficient. The more interesting fact is that an equilibrium with incomplete markets is *constrained-inefficient*: even if a central planner is constrained to use the available markets to allocate resources at dates 1 and

2, by changing the agents' decisions at date 0 the planner make some agents better off without making others worse off. How is this possible? To be sure, the intermediaries, consumers and investors are making individually optimal choices in equilibrium, but they do so taking the equilibrium prices (p_H, p_L) as given. The planner, by manipulating the choices of agents at date 0 is able to change the equilibrium prices and this will have some impact on welfare. When markets are complete, these pecuniary externalities sum to zero, that is, they cannot make everyone better off. But when markets are incomplete, agents do not have the same marginal rates of substitution and it is generically possible to generate pecuniary externalities that do make everyone better off (Geanakoplos and Polemarchakis, 1986).

The scope for welfare-improving capital regulation depends on the ability of the regulator to manipulate prices in a way that makes consumers better off. Suppose that the regulator can dictate the amount of capital e_0 raised by each intermediary. The intermediaries take the level of capital e_0 and the market-clearing prices (p_H, p_L) as given and choose the optimal repayments $(e_{1H}, e_{2H}, e_{1L}, e_{2L})$ and consumption plan $(c_{1H}, c_{2H}, c_{1L}, c_{2L})$, subject to their budget constraints and the investors' participation constraint. For each value of e_0, there will be an equilibrium in which intermediaries are maximizing the representative consumers' expected utility and markets clear at the prices (p_H, p_L). In this model, welfare is unambiguously measured by the expected utility of the representative consumer, since investors receive no surplus and intermediaries earn zero profits in equilibrium. Then the regulator will vary the required level of capital e_0 in order to maximize the consumers' expected utility. We refer to this capital level and the corresponding values of the endogenous variables $(e_{1H}, e_{2H}, e_{1L}, e_{2L})$, $(c_{1H}, c_{2H}, c_{1L}, c_{2L})$, and (p_H, p_L) as a *constrained optimum*.

The graph in Figure 3 shows the value of e_0 that maximizes welfare (consumers' expected utility) for each value of σ. Strikingly, while the constrained-optimal value of e_0 is close to the equilibrium value for values of σ less than 2.5, the two values diverge for large values of σ and the constrained-optimal value of e_0 is less than the equilibrium value. In other words, there is too much capital in the incomplete-markets equilibrium.

We can get more insight into the reasons why it makes sense to reduce the amount of capital by looking at Figure 4, which shows the prices corresponding to the the constrained optimum. For values of σ greater than 2.5, where $p_L = 1$, the reduction in capital increases the asset price in state H and thus increases the consumption of the early consumers, ceteris paribus. For values of σ less than 2.5, the constrained optimal value of e_0 does not differ greatly from the equilibrium value, but prices are reduced in both states. So there appears to be a critical value of σ above which it is optimal to increase asset prices and below which it is optimal to reduce asset prices.

The change in prices is critical for understanding the change in welfare. In equilibrium, the contracts e and c are chosen optimally from the point of view of consumers' welfare so the envelope theorem tells us that the change in e and c in response to a change in e_0 does not have a first-order impact on welfare. Only price changes have a first-order impact on welfare. Allen and Gale (2004) study the effects of regulation on welfare in the context of a model that has a similar structure but does not allow for bank capital and show that an increase (decrease) in asset prices increases welfare if and only if the degree of relative risk aversion is greater (less) than one. Here the presence of capital complicates the analysis and presumably explains why the critical value of σ is greater than one, so that a reduction in prices is called for in the interval between 1 and 2.5.

To understand why a reduction in capital is required to increase asset prices (for values of σ greater than 2.5), note that reducing the level of capital reduces the total amount of assets held by the intermediaries but if we look at the composition of the portfolio we see that it decreases the amount of the long asset by more than the amount of the short asset. Thus, it increases the liquidity of the market and raises prices in the bad state (Allen and Gale, 1994).

We are left with a counter-intuitive result: in order to improve risk sharing one needs to restrict the capital levels of the intermediaries. It ceases to be surprising that reducing capital can improve welfare, however, once one realizes that this is simply a means of controlling intermediaries' portfolios and that, because of general equilibrium effects, one cannot predict in general whether the relative proportions of long- and short-term assets will increase or decrease as capital ratios are raised. This observation also suggests that directly regulating the proportions of short- and long-term assets in the intermediaries' portfolios might be a more effective and more easily predictable policy than regulating levels of capital.

A final caveat to this analysis is that the changes in welfare from implementing the optimal policy are small. Assuming that banks can only use incomplete contracts, such as deposit contracts, would introduce the possibility of default and that might increase the deadweight cost of incomplete markets. This would not necessarily change the qualitative features of the analysis. Changes in prices are still critical to the determination of the welfare impact of capital regulation, so unless the relationship between the change in e_0 and changes in prices turns out to be much simpler in a model with the added complexity of default, one would not expect to be able to predict the direction of the welfare-improving policy.

Figure 1: CONSUMPTION PROFILES

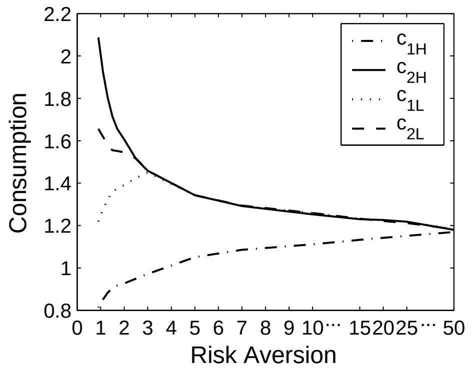


Figure 2: ASSET−RETURN PROFILES

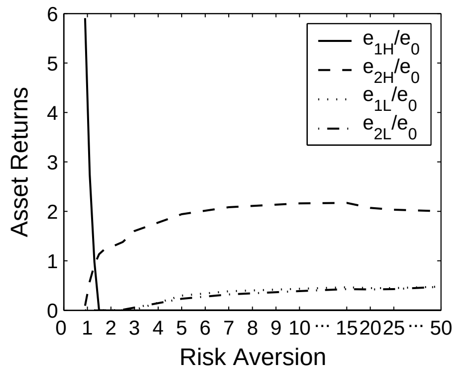


Figure 3: CAPITAL LEVEL

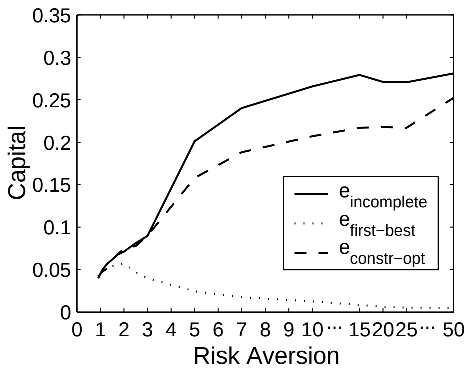
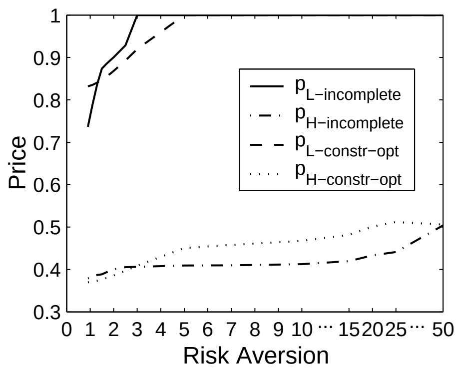


Figure 4: PRICES



References

Allen, Franklin and Douglas Gale (2003). "Capital Adequacy Regulation: In Search of a Rationale" in Richard Arnott, Bruce Greenwald, Ravi Kanbur and Barry Nalebuff (eds). *Economics for an Imperfect World: Essays in Honor of Joseph Stiglitz.* Cambridge, MA: MIT Press.

— and — (2004). "Financial Intermediaries and Markets," *Econometrica* **72**, 1023-1061.

Gale, Douglas (2003). "Financial Regulation in a Changing Environment," in Thomas Courchene and Edwin Neave (eds) *Framing Financial Structure in* *an Information Environment.* Kingston, Ontario: John Deutsch Institute for the Study of Economic Policy, Queen's University.

— (2004). "Notes on Optimal Capital Regulation," New York University, unpublished.

Geanakoplos, J. and H. Polemarchakis (1986). "Existence, Regularity, and Constrained Suboptimality of Competitive Allocations When the Asset Market Is Incomplete," in Walter, Heller, Ross Starr, and David Starrett eds. *Essays in honor of Kenneth J. Arrow: Volume 3, Uncertainty, information, and communication.* Cambridge, New York and Sydney: Cambridge University Press, 65-95.