Asset Prices and Insurance Loadings

Semyon Malamud  Eugene Trubowitz

First version: July 2008
Current version: July 2008

This research has been carried out within the NCCR FINRISK project on “Mathematical Methods in Financial Risk Management”
Asset Prices and Insurance Loadings

Semyon Malamud† and Eugene Trubowitz‡

July 1, 2008

Abstract

We analyze an equilibrium model in which agents exposed to idiosyncratic risk can purchase insurance policies in addition to financial assets. The price of an insurance contract depends nonlinearly on the claims and explicitly contains safety loadings, proportional to variance. We consider random loading factors that vary with business cycles.

To extract economic information from the equilibrium, we expand the stochastic discount factor (SDF) in the volatility of consumption growth, a natural small parameter. We obtain the true dependence of the SDF on idiosyncratic risk and insurance loadings. Our expression for the stochastic discount factor displays a threshold loading such that the equilibrium SDF is monotone increasing in the cross-sectional variance of idiosyncratic risk for loadings above the threshold. The SDF is monotone decreasing for loadings below the threshold.

One consequence, for loadings below the threshold, is that countercyclical idiosyncratic risk has the opposite influence on the SDF to that of the classic Constantinides and Duffie (1996) tradeless equilibrium.

Contents

1 Introduction 1
2 Setup 3
3 Hedgeable filtration, hedgeable claims and the size of unhedgeable idiosyncratic risk 5
4 Market consistent pricing of insurance liabilities 6

---

*Financial support by the National Centre of Competence in Research ”Financial Valuation and Risk Management” (NCCR FINRISK) is gratefully acknowledged. We also thank Paul Embrechts and Mario Wüthrich for useful comments and remarks.

†ETHZ and Swiss Finance Institute, D-MATH, HG G 33.1, Rämistrasse 101, CH–8092 Zürich, Switzerland (semka@math.ethz.ch)

‡ETHZ, D-MATH, HG G 34.1, Rämistrasse 101, CH–8092 Zürich, Switzerland (trub@math.ethz.ch)
5 The problem of individual optimal asset-insurance allocation 7
6 The law of large numbers and CRRA utilities 9
7 Optimal asset-insurance allocation for small idiosyncratic risk 11
8 Equilibria for weak uncertainty. Expansion in the size of uncertainty 13
9 Heterogeneous classes 15
10 Conditionally independent idiosyncratic shocks 17
11 Calibration with i.i.d. loadings and Markov endowment 20
12 Conclusions 22
Appendices:
A The derivatives of the optimal consumption and claims streams 24
B Expansion of the optimal consumption stream for small uncertainty 25
C Expansion of the SDF for geometric random walks 25
References 27
1 Introduction

We analyze a general equilibrium model that allows us to determine the explicit dependence of the stochastic discount factor (SDF) on exogenous insurance loadings. Our analysis is in the context of an idiosyncratically incomplete, multiperiod market, where one can purchase insurance policies in addition to financial assets. This work is a preliminary step in the development of a unified, market consistent, theory of asset and liability pricing. In Malamud, Trubowitz, and Wüthrich (2007), we introduce a framework that endogenizes insurance loadings.

Traditionally, one distinguishes between the notion of a complete consumption insurance model and its counterpart, an incomplete consumption insurance model.

A complete, consumption insurance model is functionally a complete market model, albeit with a small conceptual twist. One imagines that the market is complete because each agent can buy insurance against any state contingent shock. There is, however, the unspoken and upon consideration, the unpalatable assumption that financial assets and insurance policies are priced by the same mechanism. Of course, the mechanisms are not the same. The price of an asset is equal to its expected discounted payoff. The price of an insurance policy is equal to its expected discounted payoff plus a safety loading.

An incomplete, consumption insurance model is characterized by the property that an agent can only hedge against aggregate shocks, or, equivalently, cannot hedge against idiosyncratic shocks. Incomplete consumption insurance models have been studied intensely. They are also known as idiosyncratically incomplete market models. A great deal of the common wisdom about the influence of idiosyncratic shocks on asset prices is based on the tradeless equilibria of Constantinides and Duffie (1996). In the absence of trade, they obtain the explicit dependence of the SDF on idiosyncratic risk. A general solution to the problem, for equilibria with trade, is given for small idiosyncratic risk in Malamud and Trubowitz (2006a).

The main prerequisite for the formulation of our model is an exogenous Ansatz for the safety loading. Of course, some fraction of the safety loading is intrinsically exogenous, imposed by a regulator. The endogenous component depends on the capital of the insurance company, the risk preferences of the shareholders and the market. See, Schlesinger and Doherty (1985), Schlesinger (2000), Dana and Scarsini (2007), Malamud, Trubowitz, and Wüthrich (2007) for microeconomic theory of insurance premia and loadings.

Naturally, the pricing of an insurance liability must be market consistent in the sense that it is decomposed into a hedgeable and an unhedgeable component. The latter is orthogonal to the market subspace. The hedgeable component is priced linearly by arbitrage. The unhedgeable component is priced nonlinearly. Its price is the safety loading. There is a large literature dedicated to market consistent pricing of insurance liabilities. See, e.g., Bühlmann,
Delbaen, Embrechts, and Shiryaev (1998), Rouge and El Karoui (2000), Embrechts (2000), Henderson and Hobson (2004), Malamud, Trubowitz, and Wüthrich (2007). The simplest, commonly accepted Ansatz for the loading is a multiple, $\alpha^{-1}_t$, of the variance.\footnote{If the size of insurance claims is small, we can approximate the highly nonlinear insurance premium by its second order Taylor polynomial. Then, obviously, the first order term is the expected value of claims and the second order is the variance. Thus, the variance Ansatz makes a perfect sense when insurance claims are small. This is precisely the case that we analyze in this paper.} We adopt the Ansatz that the loading at each period is proportional to the variance of the unhedgeable component of the claims. The exogenously specified loading factor $\alpha^{-1}_t$ is random and depends on the time $t$.\footnote{This allows to account for the empirically observed cyclicity of insurance premia.} We will refer to $\alpha^{-1}_t$ as the loading factor.

We assume, as in Constantinides and Duffie (1996), that the economy is populated by an infinite number of agents with identical risk aversion, discount factors and independent, identically distributed idiosyncratic income shocks. But, by contrast to Constantinides and Duffie, we allow for an arbitrary aggregate endowment process and arbitrary idiosyncratic shocks.

The first step towards analyzing equilibria is to solve the utility maximization problem for each agent.\footnote{See, Malamud and Trubowitz (2006b).} In our model, an agent faces a difficult problem of finding an optimal asset-insurance allocation. Namely, he must choose an optimal combination of a portfolio of financial assets and an insurance contract which, by construction, is priced nonlinearly. The first order conditions for this utility maximization problem constitute a highly nonlinear system of equations that is very difficult to analyze. But, when idiosyncratic shocks and the aggregate uncertainty are both small, we can use standard perturbation theory (Taylor polynomial approximation) to get exact expressions for the optimal consumption stream. See, Samuelson (1970) and Judd and Guu (2001) for the description of the perturbative method. See also, Schlesinger and Doherty (1985), Briys (1986), Moridaira, Urrutia, and Witt (1992), Aase (1993), Gollier (1994), Meier and Ormiston (1995), Aase (forth), Schlesinger (2000), Dana and Scarsini (2007), Lee (2007) for other work on optimal asset-insurance choice.

We expand the equilibrium state price densities in the standard deviation $\sigma$ of the growth rate of aggregate endowment.\footnote{The standard deviation $\sigma$ is a natural small parameter. For example, the commonly accepted value of $\sigma$ for the US economy is $\sigma \approx 0.036$.} Using the implicit function theorem, we show that equilibrium state price densities are analytic functions of $\sigma$ and the expansion converges in a small neighborhood. See, Kogan and Uppal (2001), Chan and Kogan (2002), Leippold, Trojan, and Vanini (2006) for other applications of perturbation theory to problems in economics.

Our convergent expansion allows complete control over equilibria. We explicitly calculate the first two terms of this expansion, those of order $\sigma$ and $\sigma^2$, and show how the presence of insurance screens the precautionary

1\footnotetext{If the size of insurance claims is small, we can approximate the highly nonlinear insurance premium by its second order Taylor polynomial. Then, obviously, the first order term is the expected value of claims and the second order is the variance. Thus, the variance Ansatz makes a perfect sense when insurance claims are small. This is precisely the case that we analyze in this paper.}

2\footnotetext{This allows to account for the empirically observed cyclicity of insurance premia.}

3\footnotetext{See, Malamud and Trubowitz (2006b).}

4\footnotetext{The standard deviation $\sigma$ is a natural small parameter. For example, the commonly accepted value of $\sigma$ for the US economy is $\sigma \approx 0.036$.}
savings effect, generated by idiosyncratic shocks and may even drive bond prices down, relative to the background complete market.

To be able to make concrete economic predictions, based on our expression for the pricing kernel, we make (following Constantinides and Duffie (1996)) the simplifying assumption that both aggregate endowment and idiosyncratic shocks follow geometric random walks. In this case, the extremely complicated expression for the pricing kernel becomes much more transparent and many important predictions can be made. As an illustration, we discuss here the implications for the equity premium.

The general consensus that equity returns are typically procyclical implies that a highly countercyclical SDF is necessary for a large equity premium (see, Brav, Constantinides, and Geczy (2002), Storesletten, Telmer, and Yaron (2007)). This argument is the justification for the commonly accepted belief that only countercyclical idiosyncratic shocks can increase equity premium. There are new, economically significant effects in the presence of insurance because the multiplicative factor appearing in the expression for the SDF in front of the cross-sectional variance of the growth of idiosyncratic shocks, changes sign when the loading factor $\alpha_t$ crosses an explicitly given threshold. If, as is common, we make the assumption that the cross-sectional variance is monotone decreasing in the aggregate consumption growth, then the contribution to SDF from idiosyncratic risk is countercyclical when $\alpha_t$ lies below the threshold and procyclical when it lies above. In particular, sufficiently small insurance loadings imply that countercyclical idiosyncratic risk actually reduces equity premium.

2 Setup

We assume a discrete time, pure exchange economy with a single perishable consumption good. The economy is endowed with a stream of $W = (W_t, t = 1, \cdots, T)$ of the single good, referred to as the aggregate endowment. Equity is the Lucas tree asset whose dividend process coincides with the aggregate endowment. The supply of equity is normalized to one. Apart from equity, there are zero coupon risk free bonds of all maturities. All risky assets apart from equity are assumed to be in zero net supply.

Fix a probability space $(\Omega, \mathcal{B})$. In our model there are $T$ time periods and two filtrations $\mathcal{F} = (\mathcal{F}_t, t = 0, \cdots, T)$ and $\mathcal{I} = (\mathcal{I}_t, t = 0, \cdots, T)$ of the underlying sigma algebra $\mathcal{B}$ satisfying $\mathcal{I}_t \supset \mathcal{F}_t$ for each $t = 0, \cdots, T$. We imagine that the filtration $\mathcal{F}$ contains information only about ”aggregate” events and that the dominating filtration $\mathcal{I}$ contains additional information about individuals, that is, idiosyncratic events.

Note that a conditional expectation operator is a projection. Namely, in the space $L_2(\Omega, \mathcal{B})$ the operator of conditional expectation on sigma algebra

---

5Most equilibrium asset pricing models make the random walk assumption. See, e.g., Campbell and Cochrane (1999).
\( \mathfrak{A} \) is the orthogonal projection onto \( L_2(\Omega, \mathfrak{A}) \) of \( \mathfrak{A} \)-measurable random variables. For this reason it will be useful for us to use the notation \( P^\mathcal{F}_t \) and \( P^\mathcal{G}_t \) for the conditional expectations on \( \mathcal{F}_t \) and \( \mathcal{G}_t \) respectively. When working with processes, we will be often dealing with direct sums

\[
P^\mathcal{F}_t = \bigoplus_{t=1}^T P^\mathcal{F}_t \quad \text{and} \quad P^\mathcal{G}_t = \bigoplus_{t=1}^T P^\mathcal{G}_t
\]

It will be technically useful to assume that knowledge of idiosyncratic events at time \( t \) does not give us any information about aggregate events at time \( t + 1 \). We make the formal

**Assumption 1** For each \( t = 1, \ldots, T \), and for every integrable random variable \( Y \) measurable with respect to \( \mathcal{F}_{t+1} \),

\[
E [ Y | \mathcal{F}_t ] = E [ Y | \mathcal{G}_t ] \tag{2.1}
\]

We assume that

- the probability space \( \Omega \) is finite;
- the market is complete with respect to the aggregate events;
- all asset price and dividend processes are adapted to the aggregate filtration \( \mathcal{F} \);
- there is a unique aggregate state price density process \( M = (M_t, t = 1, \ldots, T) \) through which all assets are priced.\(^6\)

An important consequence of Assumption 1 is

**Proposition 2.1** Under Assumption 1,

\[
q_{it} = E \left[ \frac{M_{t+1}}{M_t} (q_{it+1} + d_{it+1}) | \mathcal{F}_t \right] = E \left[ \frac{M_{t+1}}{M_t} (q_{it+1} + d_{it+1}) | \mathcal{G}_t \right]
\]

for any asset \( i \) in the economy.

As is common in the modern finance literature (see, e.g., Constantinides and Duffie (1996)), we will be working directly with the pricing kernel \( M \) and then derive all asset prices from \( M \).\(^7\)

---

\(^6\)In fact, the last item is a direct consequence of the first one. The assumption of aggregate completeness is implies that idiosyncratic risk is the only source of market incompleteness.

\(^7\)When aggregate endowment is a binomial process (the event tree is binomial), equity and one period risk free bonds are enough to make the market dynamically complete. In general, it may happen that equity and zero coupon bonds of all maturities are not enough to make the market dynamically complete. In this case we can, for example, add several other multinomial processes to complete the market.


3 Hedgeable filtration, hedgeable claims and the size of unhedgeable idiosyncratic risk

Let \( w = (w_t, t = 0, \ldots, T) \) be the \( \mathcal{G} \) adapted individual endowment process of an agent in our economy.

An agent trades in the market and chooses a \( \mathcal{G} \) adapted portfolio strategy \( x = (x_1, \ldots, x_L) \). Here, \( x_j = (x_{j0}, \ldots, x_{jT-1}, 0) \) and the random variable \( x_{jt} \) counts the number of shares of asset \( A_j \) held at time \( t + 1 \) before dividends are paid and assets are traded. The last component 0 formalizes the convention that no investments are made at the final time period \( T \). The dividend process \( D_x \) generated by the portfolio strategy \( x \) is

\[
D_{x,t} = \sum_{j=1}^L (D_{jt} + q_{jt}) x_{jt-1} - \sum_{j=1}^L q_{jt} x_{jt}
\]

for \( t = 0, \ldots, T \), where \( D_j = (D_{jt}) \) and \( q_j = (q_{jt}) \) are the dividend and price processes of the asset \( A_j \). In particular, the initial investment is \( D_{x,0} = -\sum_{j=1}^L q_{j0} x_{j0} \). Our assumption of aggregate completeness allows for a very simple and elegant description of the set of dividend processes of portfolio strategies. The following result is an immediate consequence of Assumption 1 and the standard properties of state price densities.

**Lemma 3.1** Let for each \( t = 1, \ldots, T \), \( \mathcal{H}_t = \sigma(\mathcal{F}_t, \mathcal{G}_{t-1}) \) be the minimal sigma algebra containing \( \mathcal{F}_t \) and \( \mathcal{G}_{t-1} \). Then, for any \( \mathcal{H} \)-adapted process \( X_t \) there exists a portfolio strategy \( x \), such that

\[
X_t = \sum_{j=1}^L (D_{jt} + q_{jt}) x_{jt-1} = E \left[ \sum_{\tau=t}^T D_{x,\tau} M_{\tau} \middle| \mathcal{H}_t \right]
\]

for each \( t \).

Filtration \( \mathcal{H}_t \) has a very clear economic meaning: it is the hedgeable filtration. Precisely, since the market is complete with respect to \( \mathcal{F}_t \), it is possible to hedge against any event in \( \mathcal{H}_t \) by a \( \mathcal{G}_{t-1} \) measurable investment at time \( t-1 \). As above, we will use \( P_{\mathcal{H}}^t \) to denote the conditional expectation operator and \( P_{\mathcal{H}} = \oplus_{t=1}^T P_{\mathcal{H}}^t \) to denote the direct sum. The difference \( Q = P_{\mathcal{G}} - P_{\mathcal{H}} \) will also play an important role in our analysis. By the above, this is the "unhedgeable" projection.

In this paper we will be dealing with insurance policies. An insurance policy is a contract between the insurer and the insured. The insurer guarantees to pay a stream \( Z = (Z_t, t = 0, \ldots, T) \) of insurance claims to the insured (an agent in our economy), and the agent pays an insurance premia stream to the insurer, a representative insurance company.

Recall that a claims stream \( Z = (Z_t, t = 1, \ldots, T) \) is called hedgeable if there exists a portfolio strategy \( x \) such that \( Z_t = D_{x,t} \) for all \( t \geq 1 \). The following Lemma is a direct consequence of Lemma 3.1:
Proposition 3.2 A claims stream $Z$ is hedgeable if and only if
\[(I - P^T) \mathbb{E} \left[ \sum_{\tau=t}^T Z_\tau M_\tau \mid \mathcal{G}_t \right] = 0 \iff \text{Var}_{\mathcal{G}^t} \left( P^T \sum_{\tau=t}^T Z_\tau M_\tau \right) = 0 \quad (3.2)\]
for all $\tau = 1, \cdots, T$.

Formulae (3.2) mean that a stream is hedgeable if and only if the value of the future part of the stream at each moment of time stream has no unhedgeable component. Intuitively, the conditional variance in (3.2) reflects the size of unhedgeable idiosyncratic risk.

\section{4 Market consistent pricing of insurance liabilities}

In this section we describe the insurance contracts, available for purchasing in our economy. Insurance market is assumed to be complete: agents are able to buy insurance against any claims process.

\textbf{Definition 4.1} An insurance contract consists of two nonnegative random processes, the insurance claims stream $Y = (Y_t, t = 0, \cdots, T)$ and insurance premia stream $\Pi := (\pi_t, t = 0, \cdots, T)$. We refer to $Y - \Pi$ as the effective dividend process of the insurance contract. We always assume $Y_0 = 0$ (there are no claims at time zero).

\textbf{Definition 4.2} An insurance premia stream is a map $X \rightarrow \Pi(X)$ mapping insurance claim streams to insurance premia. We assume that $\Pi$ is convex for each $t$:
\[\pi_t \left( \frac{1}{2} (Y_1 + Y_2) \right) \leq \frac{1}{2} (\pi_t(Y_1) + \pi_t(Y_2))\]

\textbf{Definition 4.3} Market consistency An insurance premia stream $\Pi = (\pi_t)$ is market consistent if for all $t \geq 0$

\begin{enumerate}
\item it is linear on the subspace of hedgeable claim streams: for any hedgeable $Z_1, Z_2$ we have $\pi_t(a Z_1 + a Z_2) = a \pi_t(Z_1) + b \pi_t(Z_2)$;
\item for any claims stream $Z$
\[\mathbb{E} \left[ \sum_{t=0}^T \pi_t(Z) M_t \right] \geq \mathbb{E} \left[ \sum_{t=1}^T Z_t M_t \right]\]
and the identity holds if and only if $Z$ is hedgeable.
\item for any claims stream $Y$ and any hedgeable claims stream $Z$ we have
\[\pi_t(Y + Z) = \pi_t(Y) + \pi_t(Z)\]
\end{enumerate}
Items (1) and (2) mean that hedgeable claims are priced just like assets in the financial market. Item (3) means that we can always subtract the hedgeable component from the claims and price it linearly because it can be perfectly hedged by trading in the financial market. Furthermore, inequality of item (2) means that insurance loading (=risk premium) for the unhedgeable component of the claims stream is positive and nonlinear.

From now on we will make the variance Ansatz for the premia streams. Proposition 3.2 motivates us to make the following

**Assumption 2** The insurance premia are determined by

\[ \pi_0(Y) := \sum_{t=0}^{T} E\left[ Y_t M_t \right] \]

and

\[ \pi_t(Y) := \frac{1}{2} \alpha_t \text{Var}_{\mathcal{F}_t} \left( M_t^{-1} P_{\mathcal{H}_t} \sum_{\tau=t}^{T} Y_\tau M_\tau \right) \]

for all \( t = 1, \cdots, T \) where \( \mathcal{A} := (\alpha_t, t = 1, \cdots, T) \) is a \( \mathcal{F} \)-adapted positive process, \( \alpha_t > 0 \) for all \( t \).

Proposition 3.2 immediately yields that the above premia are market consistent. Furthermore, they can be viewed as a second order Taylor polynomial approximation to the ”real life” premia when insurance claims are small. This is precisely the case that we analyze in this paper.

5 The problem of individual optimal asset-insurance allocation

**Definition 5.1** An agent with the endowment stream \( \mathbf{w} \) maximizes

\[ U(c) := E \left[ \sum_{t=0}^{T} e^{-\rho t} u(c_t) \right] \]

over the budget set

\[ \mathbf{B}(\mathbf{w}) := \left\{ \mathbf{c} = \mathbf{w} + D_x + \mathbf{Y} - \Pi(\mathbf{Y}) > 0 \mid \mathbf{Y} \text{ is } \mathcal{G} \text{-adapted and } x \text{ is a } \mathcal{G} \text{adapted portfolio strategy} \right\} \quad (5.1) \]

That is, the agent optimally allocates his wealth between a portfolio of financial assets and an insurance contract.

Market consistency of the premia streams defined in Assumption 2 allows us to reduce the above utility maximization problem to a simpler form.
Proposition 5.2 A consumption stream $c \in B(w)$ if and only if there exists a $\mathcal{G}$ adapted process $Z = (Z_t, t = 0, \cdots, T)$ such that

$$c_t(Z) = w_t + Z_t - \pi_t(Z)$$

for all $t \geq 1$ and

$$\sum_{t=0}^{T} E[Z_t M_t] = 0 \quad (5.2)$$

By construction, both insurance and financial markets are assumed to be perfect. There are no transaction costs or intermediary fees. Consequently, since insurance premia are market consistent, an agent is indifferent between directly trading in the financial market and buying an insurance contract that would give him the payoff of the corresponding portfolio strategy. Therefore, everywhere in the sequel we will refer to $Z$ as the optimal claims stream, even though it in fact incorporates the dividend stream of the optimal portfolio strategy.\(^8\)

Note that $U(c(Z))$ is a concave function of $Z$ since insurance premia are convex functions of claims. Since the probability space is finite, utility $u$ satisfies the Inada condition $u'(0) = +\infty$ and we are maximizing over a convex set defined by (5.2), an interior solution always exists. Inada condition does not allow the optimal consumption stream come close to zero. Thus, to understand the structure of the optimal consumption stream, we have to write down the first order conditions.

Proposition 5.3 Let $Z$ be the optimal claims stream maximizing $U(c(Z))$ and let $c = (c_t) = (c_t(Z))$ be the corresponding optimal consumption stream. Then,

$$e^{-\rho t} u'(c_t) M_t^{-1} = \lambda + \sum_{\tau=1}^{t} \frac{1}{M_{\tau}^{2}} (P_{\mathcal{H}_\tau} e^{-\rho \tau} u'(c_\tau)) Q^{\tau} \sum_{\theta=\tau}^{T} Z_\theta M_\theta \quad (5.3)$$

Here, $\lambda$ is the Lagrange multiplier, uniquely determined by the budget constraint (5.2).

First order conditions (5.3) allow us to obtain a lot of important information about the optimal consumption stream. The following is true.

Theorem 5.4 The quotient of the marginal utility process $e^{-\rho t} u'(c_t)$ and aggregate state price density process $M_t$ is a martingale relative to the filtration $(\mathcal{H}_{t+1}, t = 0, \cdots, T - 1)$. That is,

$$P_{\mathcal{H}_t} e^{-\rho t} u'(c_t) M_t^{-1} = e^{-\rho (t-1)} u'(c_{t-1}) M_{t-1}^{-1} \quad (5.4)$$

\(^8\)Of course, in the real world both insurance and financial markets are not perfect. Intermediary costs may be very high. On the other hand, it may be easier and cheaper for a large insurance/financial company to trade in the market, then for a "small" agent. The effect of market imperfections is a topic for future research.
In fact,

\[ e^{-\rho t} u'(c_t) M_t^{-1} = e^{-\rho(t-1)} u'(c_{t-1}) M_{t-1}^{-1} \]

\[ + \frac{1}{M_t^2 \alpha_t} \left( P_t^g e^{-\rho t} u'(c_t) \right) Q^t \sum_{\tau=t}^{T} Z_{\tau} M_{\tau} \]  

(5.5)

Here, as above, \( Q_t = P_t^g - P_{t}^g \) is the unhedgeable projection.

Note that, in a complete market the marginal utility process is proportional to the unique (aggregate) state price density process. When the market is incomplete, the marginal utility is equal to the agent’s own, subjective state price density. Nevertheless, the agent tries to push his marginal utility to the “least risky”, aggregate state price densities \( M \). Identity (5.5) provides an explicit expression for the difference

\[ e^{-\rho t} u'(c_t) M_t^{-1} - e^{-\rho(t-1)} u'(c_{t-1}) M_{t-1}^{-1} \]

That is, the "barrier" to market completeness. Clearly, this is the unedgeable component

\[ Q^t \sum_{\tau=t}^{T} Z_{\tau} M_{\tau} \]

of the optimal claims stream \( Z \).

Identity (5.5) provides some economic intuition about the structure of the optimal consumption. Conditioned on hedgeable events in \( \mathcal{H}_t = \sigma(\mathcal{F}_t, \mathcal{G}_{t-1}) \), monotonicity of \( u' \) and (5.5) imply that consumption \( c_t \) is a monotone decreasing function of

\[ \frac{1}{\alpha_t} Q^t \sum_{\tau=t}^{T} Z_{\tau} M_{\tau} , \]

the unedgeable part of idiosyncratic shocks times the insurance loading, but modulo the state prices. If the unedgeable risk is large and insurance is expensive, consumption will be small. On the other hand, the sensitivity also depends on \( u'(c_{t-1}) \). If past consumption \( c_{t-1} \) is small, \( u'(c_{t-1}) \) is large and therefore, today’s consumption is very sensitive to idiosyncratic shocks.

Equation (5.5) is also useful for factor analysis (regression). Using real consumption data to uncover marginal utility of consumption (see, e.g., Attanasio and Jappelli (2001) for an econometric analysis), we can use (5.5) as a theoretical basis for regressing marginal utility on insurance loadings and/or idiosyncratic shocks.

6 The law of large numbers and CRRA utilities

By construction, the markets we are considering are idiosyncratically incomplete. The only way this can happen is when the aggregate demand is adapted
to $\mathcal{F}$, the aggregate filtration. In this case, the aggregate demand does not depend on individual idiosyncratic shocks, but, rather, on the cross-sectional distribution of these shocks. This can only happen if each individual agent is of measure zero and idiosyncratic shocks are independent and identically distributed across agents, conditioned on the aggregate filtration (this assumption is standard in the literature on idiosyncratic shocks. See, e.g., Mankiw (1986), Constantinides and Duffie (1996)).

We will assume that the following strengthened form of the Inada condition holds: the marginal utility blows up as a power of $c$ at zero. Namely, we assume that there exists a positive $\gamma$ such that

$$\lim_{c \to 0} \frac{u'(c)}{c^{-\gamma}} = K \quad (6.1)$$

for some positive, finite constant $K$. This assumption is satisfied by all standard utilities, including, of course, the CRRA class.\footnote{We would like to thank the referee, who stimulated us to consider equilibria with general utilities.}

Consider a class of agents $j = 1, \cdots, N$ with identical utilities $u$, identical discount rates $\rho$, the same aggregate component $N^{-1}W$ of individual endowments

$$w = N^{-1}(W + w_j)$$

and assume that the idiosyncratic shocks $N^{-1}w^1(j)$ are i.i.d. across the agents $j = 1, \cdots, N$, when conditioned on $\mathcal{F}$.

The following is true.

**Proposition 6.1** Let $c(j) = c(w^1(j))$ be the optimal consumption stream for the agent $j$ with utility $u(c)$ and let $c(\gamma, W + w^1(1))$ be the optimal consumption stream for an agent with the CRRA utility $c^{1-\gamma}$ and the same endowment. Then,

$$\lim_{N \to \infty} \sum_{j=1}^{N} c_j(N^{-1}(W + w_j)) = \mathbb{P}_\mathcal{F} c(\gamma, W + w^1) \quad (6.2)$$

**Proof.** Proposition 6.1 follows from (6.1) by an application of the (conditional) law of large numbers and homogeneity of power utility.

As we send the number $N$ of agents to infinity, the individual endowment $N^{-1}(W + w_j)$ of agent $j$ goes to zero, and therefore individual consumption also goes to zero. But, when consumption is small, (6.1) implies that the agent behaves approximately like the one with a power utility $c^{1-\gamma}$. (This follows directly from Theorem 5.4, because the optimal consumption stream is uniquely determined by the first order conditions, and for small $c$ these conditions coincide with those for the power utility). Because power utility is homogeneous, consumption is homogeneous of degree one and hence,

$$c(\gamma, N^{-1}(W + w_j)) \approx N^{-1}c(\gamma, W + w_j^1)$$
There are some minor technical issues, arising because the premium \( \Pi(Z) \) is nonlinear. But, it is quadratic and generates corrections of order \( 1/N^2 \) that disappear in the limit.

A direct application of the law of large numbers implies

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} c(j) = \lim_{N \to \infty} N^{-1} \sum_{j=1}^{N} c(\gamma, W + w_j^1) = P\mathcal{F} c(\rho, \gamma, W + w^1, M)
\]

which is what had to be proved.

The result of Proposition 6.1 is somewhat surprising: even if we allow for general utilities, the aggregate demand and, consequently, equilibrium asset prices, will look like those for an economy, populated by agents with CRRA utilities. This fact allows us to confine our analysis to the benchmark case of CRRA utilities.

## 7 Optimal asset-insurance allocation for small idiosyncratic risk

Basing on Proposition 6.1, we will confine our subsequent analysis to the case of CRRA, power utility functions.

In general, the dependence of the optimal consumption stream \( c \) on idiosyncratic income shocks \( w \) is very complicated. In this section we calculate Taylor polynomial approximations to the optimal consumption stream when uncertainty is small. Therefore, in this section we will assume that the aggregate state price density process \( M \) is deterministic, that is \( M_t \) is a constant, depending on \( t \). This simple case is the main ingredient of our perturbative analysis of the equilibrium asset prices for weak uncertainty.

We will need several notations.

It is known (see, e.g., Wang (1996)) that

\[
c_m = (c_{m_t}, t = 1, \ldots, T) = (e^{-\rho t} M_t^{-b} c_0, t = 1, \ldots, T)
\]

is the optimal consumption stream of an agent with discount rate \( \rho \) and risk aversion \( \gamma \) in a complete market.

**Definition 7.1** Let \( A = (\alpha_t, t = 1, \ldots, T) \) and define the process \( N = (N_t, t = 1, \ldots, T) \) via

\[
N_t := N_t(A) = \gamma M_t \alpha_t + \sum_{\tau=t}^T c_{m_{\tau}} M_{\tau}
\]

Let \( H = \oplus_{t=1}^T L_2(\Omega, \mathcal{G}_t) \) be the space of \( \mathcal{G} \)-adapted processes. Let \( J : H \to H \) be defined by

\[
(J(x))_t = \sum_{\tau=1}^T x_\tau
\]
Any endowment process \( w \in H \) admits an orthogonal decomposition \( w = w^A + \sigma w^I \) into the aggregate component \( w^A = P_x w \) and the idiosyncratic component \( \sigma w^I = (I - P_x) w \). We assume that the parameter \( \sigma \), determining the size of idiosyncratic shocks, is small. In applications to equilibrium asset prices, \( \sigma \) will be the volatility of the aggregate consumption growth, which is known to be small in the empirical data. The following is true.

**Theorem 7.2** Let \( w(\sigma) = w^A + \sigma w^I \) and let \( c(\sigma) = c(w(\sigma)) \) be the optimal consumption stream defined in Theorem 5.4. Then, \( c(\sigma) \) is a real analytic function of \( \sigma \) for sufficiently small \( \sigma \). Define the operator

\[
B = cmJN^{-1}QJ^*M \tag{7.2}
\]

Then, the conditional expectation of the optimal consumption stream relative to the aggregate events in \( \mathcal{F} \) is given by

\[
P_x c(\sigma) = cm + \frac{1}{2} \sigma^2 \left( (\gamma + 1) cm^{-1}(B w^I)^2 - 2 (M^2 A)^{-1} P_x \left( Q J^* M cm JN(A)^{-1} N(0)^{-1} A Q J^* M w^I \right)^2 \right) + O(\sigma^3) \tag{7.3}
\]

See, Theorem B.2 in the Appendix for the expansion of the whole consumption stream \( c \).\(^\text{10}\) As the number of agents in the economy goes to infinity, the law of large numbers implies that the aggregate consumption converges to the conditional expectation \( P_x c \) (see, e.g., Constantinides and Duffie (1996)). Therefore, (7.3) is sufficient for the subsequent equilibrium analysis.

The terms of order \( \sigma^2 \) appearing in (7.3) are weighted combinations of conditional variances of idiosyncratic shocks relative to the aggregate events. Naturally, variance is the simplest and most popular proxy for the size of idiosyncratic risk (see, e.g., Constantinides and Duffie (1996) and Jacobs and Wang (2004)). Formula (7.3) shows explicitly how optimal consumption depends on the variance of idiosyncratic shocks.

As the insurance loadings \( A^{-1} \) become infinite \((A_t \to 0)\), insurance becomes so expensive that nobody buys it and we arrive at a pure financial market model without insurance. In this case, the term in the second line of (7.3) vanishes and we get for \( A = 0 \),

\[
P_x c = cm + \frac{1}{2} \sigma^2 \left( (\gamma + 1) cm^{-1}(B w^I)^2 \right) + O(\sigma^3) \tag{7.4}
\]

\(^{10}\)This expansion is quite complicated contains a lot of terms, that are difficult to interpret.
Consequently, $P\mathcal{F}c \geq cm$ for sufficiently small $\sigma$. This is the well known precautionary savings effect: for utility functions with a positive third derivative, idiosyncratic risk forces the agent to increase savings and drives future consumption up. See, e.g., Leland (1968), Sandmo (1970), Kimball (1990), Huggett and Ospina (2001). On the other hand, in the zero loadings limit $A \to +\infty$, insurance becomes so cheap that the market becomes effectively complete and $c \to cm$. In particular, there are no precautionary savings in this case.

Of course, the most interesting effects appear in the more realistic, intermediate regime $0 << A << \infty$. The possibility to buy insurance diminishes the effect of idiosyncratic risk and generates the second line in (7.3). In particular, the right hand side of (7.3) is always smaller than that of (7.4). That is, insurance always diminishes precautionary savings. In fact, there are economies, for which the effect has the opposite sign.

**Proposition 7.3** There exists a open set of insurance loadings $A$ such that, for loadings in this set, idiosyncratic risk generates **precautionary borrowing effect**: in the presence of idiosyncratic risk it is optimal for the agent to consume more at time zero and, consequently, reduce future consumption.

### 8 Equilibria for weak uncertainty. Expansion in the size of uncertainty

We are now ready to write down and analyze the equilibrium equations. The first step is to use the law large numbers and send the number of agents to infinity. This is what makes the aggregate demand adapted to the aggregate filtration $\mathcal{F}$. This is standard for models with idiosyncratic risk. See, e.g., Constantinides and Duffie (1996).

Each of the $N$ agents in the economy has the same risk aversion $\gamma$ and discount rate $\rho$. Agent $j$ is endowed with the fraction $1/N$ of the aggregate endowment and each agent $j$ has an idiosyncratic income shocks process $N^{-1}w^j$. The processes $w^j$ are independent and identically distributed relative to the aggregate filtration $\mathcal{F}$ and satisfy $P\mathcal{F}w^j = 0$ for all $j$. An application of the law of large numbers yields\(^{11}\)

**Proposition 8.1** Let $c_j = c(\rho, \gamma, N^{-1}(W + w^j), M)$ be the optimal consumption stream of agent $j$. Let also $w^1 = w^1$. Then,

$$\lim_{N \to \infty} \sum_{j=1}^N c_j = P\mathcal{F}c(\rho, \gamma, W + w^1, M)$$

\(^{11}\)There are some technical issues, arising because the premium $II(Z)$ is nonlinear. But, it is quadratic and generates corrections of order $1/N^2$ that disappear in the limit.
Definition 8.2 We assume that endowment of the representative insurance company is an exogenously specified, $\mathcal{F}$-adapted process $I$. A state price density process $M$ is an idiosyncratically incomplete Arrow-Debreu equilibrium if all asset markets clear\(^{12}\) and the insurance market clears, that is, the aggregate premium paid by consumers net the aggregate claims equals $I$.

We assume that $P_{\mathcal{F}} W^1 = 0$, so that idiosyncratic shocks do not affect the size of the aggregate endowment. Standard arguments imply that the aggregate state price density process $M$ is an equilibrium if and only if the market for the consumption good clears, that is,

$$P_{\mathcal{F}} c(\rho, \gamma, W + w^1, M) = W - I$$

(8.2)

It is interesting to note that, even in the case $w^1 = 0$ of no idiosyncratic risk, we can exogenously specify $I \neq 0$. Since insurance premia are market consistent, agents will use the insurance company as an external supplier of financial assets and the equilibrium will be just a complete market equilibrium with some assets in negative net supply, so that their aggregate dividend equals $-I$.

Applying Theorem 7.2 and the implicit function theorem to (8.2), we arrive at

Theorem 8.3 Assume that the aggregate uncertainty and idiosyncratic shocks are small, that is,

$$W(\sigma) - I(\sigma) = W^{(0)} + \sigma W^{(1)} + \frac{1}{2} \sigma^2 W^{(2)} + O(\sigma^3)$$

where $W^{(0)}$ is a deterministic process and

$$w^1(\sigma) = \sigma w^{(1)} + \frac{1}{2} \sigma^2 w^{(2)} + O(\sigma^3)$$

Then, there exists a positive number $\varepsilon > 0$ such that for all $\sigma \in (-\varepsilon, \varepsilon)$ there exists a unique, real analytic in $\sigma$, SPD process $M(\sigma)$ satisfying $M(0) = (e^{-\rho t} W_t^{-\gamma})$. Furthermore, this SPD process $M = M(\sigma)$ is given by

$$M = M_h \left( 1 + \gamma W^{-1} \left( \frac{1}{2} \frac{1}{W} (\gamma + 1) P_{\mathcal{F}} \left( B w^1 \right)^2 \right) 
- (M_h^2 A)^{-1} P_{\mathcal{F}} \left( Q J^* M_h c m J N(A)^{-1} N(0)^{-1} A Q J^* M_h w^1 \right)^2 \right) + O(\sigma^3)$$

(8.3)

Here, $M_h = M_h(\sigma) = (e^{-\rho t} W_t^{-\gamma})$ is the state price density process in the complete market economy without idiosyncratic risk.

\(^{12}\)By assumption, all assets are in zero net supply.
Expansion (8.3) is a direct analog of (7.3). The first line represents the response to idiosyncratic shocks and the second line - to insurance loadings. It is known (see, e.g., Huggett and Ospina (2001)) that idiosyncratic shocks drive equilibrium bond prices up, thus decreasing the risk free rates. Insurance is able to change the sign of this effect. Namely, the following analog of Proposition 7.3 is true.

**Proposition 8.4** Let $\sigma$ be sufficiently small. Then, in the presence of insurance, the prices of risk free bonds are always smaller than those in a pure idiosyncratically incomplete market. Furthermore, there exists an open set of insurance loadings $A$ such that, for loadings in this set, insurance completely screens the effect of idiosyncratic risk and bond prices are smaller than those in the background complete market.

### 9 Heterogeneous classes

It is possible to extend the analysis of the previous section to the case of a heterogeneous population. In this section we assume that the economy is populated by heterogeneous classes of agents.

Inside each class $i$, $i = 1, \cdots, n$, all agents have the same risk aversion $\gamma_i$, same discount rate $\rho_i$, and the same idiosyncratic shocks processes $w_i^I$, that are independent and identically distributed among the agents in the same class. The last assumption is necessary, because we need to apply the law of large numbers inside each class to insure that the aggregate demand is adapted to the aggregate filtration and does not depend on individual, idiosyncratic shocks.

We also suppose that the class $i$, $i = 1, \cdots, n$ is endowed with a fraction $\eta_i$ of the aggregate endowment,

$$\sum_{i=1}^{n} \eta_i = 1$$

Even when the market is complete, analyzing a heterogeneous economy becomes a very difficult problem. See, e.g., Wang (1996), Chan and Kogan (2002), Malamud (2008), Malamud (forth.).

We will assume that the size of preferences heterogeneity is small and controlled by the same small parameter $\sigma$. Namely,

$$\gamma_i = \gamma + \sigma \Gamma_i , \quad \rho_i = \rho + \sigma \mathcal{R}_i$$

where $\Gamma_i$ and $\mathcal{R}_i$ are heterogeneous. When $\sigma = 0$, we are back to the benchmark, complete market, homogeneous economy. We will choose the approximating homogeneous economy so that

$$\sum_{i} \gamma_i \eta_i = \gamma , \quad \sum_{i} \rho_i \eta_i = \rho$$
We will need the following notation.

**Notation.** Let $\Gamma = (\Gamma_1, \ldots, \Gamma_n)$ and $\mathcal{R} = (\mathcal{R}_1, \ldots, \mathcal{R}_n)$. The wealth weighted variances of risk aversion and discount rates and their wealth-weighted covariance are defined via

$$\text{Var}_\eta(\Gamma) = \sum_i \eta_i \Gamma_i^2 - \left( \sum_i \eta_i \Gamma_i \right)^2,$$

$$\text{Var}_\eta(\mathcal{R}) = \sum_i \eta_i \mathcal{R}_i^2 - \left( \sum_i \eta_i \mathcal{R}_i \right)^2,$$

and

$$\text{Cov}_\eta(\Gamma, \mathcal{R}) = \sum_i \eta_i \mathcal{R}_i \Gamma_i - \sum_i \eta_i \mathcal{R}_i \sum_i \eta_i \Gamma_i.$$

The following analog of Theorem 8.3 is true.

**Theorem 9.1** Assume that the aggregate uncertainty and idiosyncratic shocks are small, that is,

$$W(\sigma) - I(\sigma) = W^{(0)} + \sigma W^{(1)} + \frac{1}{2} \sigma^2 W^{(2)} + O(\sigma^3)$$

where $W^{(0)}$ is a deterministic process and

$$w_i(\sigma) = \sigma w_i^{(1)} + \frac{1}{2} \sigma^2 w_i^{(2)} + O(\sigma^3)$$

for each class $i = 1, \ldots, n$. Then, there exists a positive number $\varepsilon > 0$ such that for all $\sigma \in (-\varepsilon, \varepsilon)$ there exists a unique, real analytic in $\sigma$, SPD process $M(\sigma)$ satisfying $M(0) = (e^{-\rho t} W_t^{-\gamma})$. Furthermore, this SPD process $M = M(\sigma)$ is given by

$$M = M_h \left( 1 + \sigma^2 M^\Pi_2 + M^I_2 + O(\sigma^3) \right) \quad (9.1)$$

Here, $M_h = M_h(\sigma) = (e^{-\rho t} W_t^{-\gamma})$ is the state price density process in the complete market economy without idiosyncratic risk,

$$M^\Pi_2 = Y_1 \text{Var}_\eta(\mathcal{R}) + Y_2 \text{Var}_\eta(\Gamma) + Y_3 \text{Cov}_\eta(\Gamma, \mathcal{R})$$

is the second order response of state prices to preferences heterogeneity and

$$M^I_2 = \sum_{i=1}^n \eta_i^{-1} \gamma W^{-1} \left( \frac{1}{2} \mathcal{W} (\gamma + 1) P_{\mathcal{X}} \left( \mathcal{B} w_i \right)^2 
- (M^i_h A)^{-1} P_{\mathcal{X}} \left( Q J^* \mathcal{M}_h \mathcal{C} M J N(A)^{-1} N(0)^{-1} A Q J^* \mathcal{M}_h w_i \right)^2 \right) \quad (9.2)$$

is the second order response to idiosyncratic shocks. The coefficients $Y_1, Y_2, Y_3$ are given by

$$Y_1 := 2^{-1} \left( t \gamma^{-1} - 1 \right), \quad (9.3)$$

$$Y_2 := 2^{-1} \gamma^{-1} \log W_t \left( 2 + \log W_t \right), \quad (9.4)$$

$$Y_3 := \gamma^{-1} \left( 1 + t \right) \log W_t. \quad (9.5)$$
There are several interesting features of the second order approximation of Theorem 9.1. First, the effects of preferences heterogeneity (generated by $M_{\Pi}^2$) are completely decoupled from the effects of idiosyncratic risk (generated by $M_{I}^2$). Namely, the cross-sectional correlations of idiosyncratic risk and risk aversion/patience parameters do not affect state prices to second order.

Second, the effects of individual idiosyncratic shocks are decoupled among classes: the cross-sectional correlations of idiosyncratic shocks do not affect state prices to second order. Here, one must be careful: even though the cross-correlations $\text{Corr}(w_{I}^i, w_{I}^j)$ do not enter the expansion of state price densities, they will play an important role for the joint dynamics of asset prices and returns. Correlations of asset prices and returns will be affected by cross-correlations of idiosyncratic shocks between classes. Allowing for multiple classes with complicated cross-sectional dynamics of idiosyncratic shocks is able to generate a very rich dynamic behavior.

In this paper, we restrict ourselves to the case of a one class economy because our main goal is to understand the effects of insurance loadings on asset prices.

10 Conditionally independent idiosyncratic shocks

The model that we present in this section can be viewed as an extension of the Constantinides and Duffie (1996) model to the case when agents can buy insurance. But, we would like to emphasize a crucial difference between our model and that of Constantinides and Duffie. In their model there is no trade and, consequently, no need to determine the optimal consumption stream: agents simply consume their own endowment. By contract, there is trade in our model, agents purchase financial assets and insurance contracts, and the problem of determining the optimal stream becomes highly nontrivial. As a result, only approximate solutions (see, Theorem 7.2) are possible.

The only restriction that we impose is that the idiosyncratic shocks are independent, when conditioned on the aggregate filtration, generated by the aggregate endowment. In particular, no restrictions on the structure of the aggregate endowment are needed.

From now on we assume that

- Aggregate endowment $W = X_1 \cdots X_t$ and $X_t = e^{\mu_t + \sigma \xi_t}$ where the volatility $\sigma$ is small and $\xi_t$ is an arbitrary, $\mathcal{F}$-adapted process;

- In analogy with Constantinides and Duffie (1996), idiosyncratic shocks are given by $w_{I}^t = W_t X_{I}^1 \cdots X_{I}^t$ where $X_{I}^1$ are independent (with

\footnotesize

\begin{itemize}
  \item As we explain above (see, (7.4)), pure idiosyncratic incompleteness arises in the special case of infinite loadings.
  \item In the first version of this paper, we imposed a very restrictive assumption that both aggregate endowment and idiosyncratic shocks are geometric random walks. We would like to thank the anonymous referee, whose valuable comments and remarks helped us to remove this assumption.
\end{itemize}

17
respect to the aggregate filtration) idiosyncratic income shocks satisfying $E[X^i_t] = 1$ and $E[(X^i_t)^2 | \mathcal{F}_t] = e^{Y(X^i_t)}$, where $Y$ is an arbitrary, non-anticipative functional (in applications, we will assume that $Y$ is a deterministic function, independent of the past). We also assume that $Y(X^i_t) = O(\sigma^2)$.

The horizon $T$ is assumed to be very large compared with $t$. Therefore, the approximation $T = \infty$ is often used to replace a finite sum of a geometric progression by its infinite sum. To state the next theorem, we will need several notations.

**Theorem 10.1** Under the above assumptions,

$$
\log \frac{M_{t+1}}{M_t} = -\rho - \gamma \mu - \sigma \gamma \xi_{t+1} + \gamma^3 \left( \frac{S}{\gamma \lambda_{t+1} + S} \right)^2 Y(X_{t+1}) \left( \frac{1}{2} \gamma^2 + \lambda_{t+1} \right) + \gamma^3 \left( \frac{S}{\gamma \lambda_t + S} \right)^2 Y(X_t) \lambda_t + O(\sigma^3) \tag{10.1}
$$

Here,

- $\rho$ (resp. $\gamma$) is the common discount rate (resp. risk aversion) of all agents in the economy,
- $\mu$ (resp. $\sigma$) is the expectation (resp. volatility) of the growth rate $\sigma \xi_t + \mu$ of aggregate endowment,
- $\lambda_t = \alpha_t e^{-\mu t}$ is the exogenous insurance loading per background aggregate endowment,
- $S = \left( 1 - e^{-\rho} + (1 - \gamma) \mu \right)^{-1}$,
- $Y(X_{t+1})$ is the cross sectional variance of the growth of idiosyncratic shocks. The form of $Y$ determines the cyclical variation of idiosyncratic risk,
- $O(\sigma^3)$ is an analytic function of sufficiently small $\sigma$ that goes to zero at the rate of $\sigma^3$.

By definition, the loading $\lambda^{-1}_t$ per aggregate endowment represents the degree of market incompleteness. It is the discrepancy between asset prices and insurance liability prices. As the loadings $\lambda^{-1}_t$ tend to zero (i.e., $\lambda_t \to \infty$), we recover market completeness and the limiting log SDF becomes the log SDF of the standard, homogeneous complete market economy, $-\rho - \gamma \mu - \sigma \gamma \xi_{t+1}$. On the other hand, as the loadings $\lambda^{-1}_t$ tend to infinity
(i.e., $\lambda_t \to 0$), we recover pure idiosyncratic incompleteness and obtain the Constantinides and Duffie (1996) tradeless equilibrium

$$\lim_{\lambda_t \to 0} \log \frac{M_{t+1}}{M_t} = -\rho - \gamma \mu - \sigma \gamma \xi_{t+1} + \frac{1}{2} \gamma (\gamma + 1) Y(X_{t+1}) \quad (10.2)$$

that only depends on idiosyncratic shocks at time $t+1$. Thus, our equilibrium with insurance naturally interpolates between the complete and idiosyncratically incomplete markets.

We now interpret the terms appearing in (10.1) and extract economic information. The most important contribution to the log SDF is

$$\gamma^3 \left( \frac{S}{\gamma \lambda_{t+1} + S} \right)^2 Y(X_{t+1}) \left( \frac{1}{2} \frac{\gamma + 1}{\gamma^2} - \lambda_{t+1} \right) \quad (10.3)$$

in the second line of (10.1). It only depends on idiosyncratic shocks and loadings at time $t + 1$ and determines the correlation of the log SDF with the growth rate of aggregate endowment, conditioned on information at time $t$. The importance of this conditional correlation is revealed by the standard identity

$$\frac{1}{r_{F,t+1}} E_t [r_{A,t+1}] = 1 - \text{Cov}_t \left( \frac{M_{t+1}}{M_t}, r_{A,t+1}^A \right) \quad (10.4)$$

Here, $r_{A,t+1}^A$ is the return on the asset $A$ and $r_{F,t+1}^F$ is the risk free rate. Clearly, the magnitude and sign of the conditional covariance $\text{Cov}_t \left( \frac{M_{t+1}}{M_t}, r_{A,t+1}^A \right)$ are principally determined by (10.3) in our model.

The general consensus that equity returns are typically procyclical implies that a highly countercyclical SDF is necessary for a large equity premium (see, Brav, Constantinides, and Geczy (2002), Storesletten, Telmer, and Yaron (2007)). This argument is the justification for the commonly accepted belief that only countercyclical idiosyncratic shocks can increase equity premium. This intuition is a direct consequence of (10.2). To make (10.2) decrease faster as a function of $\xi_{t+1}$, one must choose $Y(X_{t+1})$ monotone decreasing. There are new economically significant effects in the presence of insurance because the multiplicative factor $\left( \frac{1}{2} \frac{\gamma + 1}{\gamma^2} - \lambda_{t+1} \right)$ appearing in (10.3) changes sign when $\lambda_{t+1}$ crosses the threshold $\frac{1}{2} \frac{\gamma + 1}{\gamma^2}$. If, as is common, we make the assumption that $Y(X_{t+1})$ is monotone decreasing, then the term (10.3) is countercyclical when $\lambda_{t+1}$ lies below the threshold and procyclical when it lies above. In particular, sufficiently small insurance loadings imply that countercyclical idiosyncratic risk actually reduces equity premium.

The third line

$$\gamma^3 \left( \frac{S}{\gamma \lambda_t + S} \right)^2 Y(X_t) \lambda_t = \gamma^3 \frac{S^2 \lambda_t}{(\gamma \lambda_t + S)^2} Y(X_t) \quad (10.5)$$

in (10.1) also generates new effects. It depends on idiosyncratic shocks and loadings at time $t$. There is no analog of this term in tradeless equilibria (see,
This term is responsible for cyclical variation of the SDF. For example, equity premia in tradeless equilibria (10.2) are deterministic and independent of time. That is, absolute constants. Insurance loadings generate (10.5) and, in turn, produce random and time dependent equity premia.

11 Calibration with i.i.d. loadings and Markov endowment

Modeling aggregate consumption as a geometric random walk has become popular in modern financial economics. See, e.g., Constantinides (1990), Campbell and Cochrane (1999), Campbell (2003), Mehra and Prescott (2003).

Nevertheless, there is empirical evidence that the aggregate consumption is not a random walk. For example, Mehra and Prescott (1985) find a small autocorrelation in consumption growth. In a recent work, Bansal and Yaron (2004) find very important long run risk components in the aggregate consumption growth and show that these components are responsible for numerous interesting economic phenomena.

Following Mehra and Prescott (1985) and Bansal and Yaron (2004), we will model aggregate consumption growth $X_t$ as a Markov process.

In this section we calculate the price of equity (Lucas tree asset), the asset whose dividend coincides with the aggregate endowment process $W_t$. We do this under the simplifying assumption that the loading factors per background aggregate endowment, $\lambda_t$, are independent and identically distributed.

We make the

**Assumption 3** There exists a function $f(x)$ such that $\alpha_t W_t^{-1} = f(X_t)$.

Let $K((x_1, x_2), (y_1, y_2))$ be the transition kernel of the two-dimensional Markov process $(X_{t+1}, X_t)$:

$$\text{Prob}[(X_{t+1}, X_t) = (x_1, x_2) \mid (X_t, X_{t-1}) = (y_1, y_2)] = K((x_1, x_2), (y_1, y_2))$$

Define the linear operator

$$A(f(y_1, y_2))(x_1, x_2) = \sum_{y_1, y_2} K((x_1, x_2), (y_1, y_2)) F(y_1, y_2) f(y_1, y_2)$$

with

$$F(y_1, y_2) = e^{-\rho} y_1^{-\gamma} \left( \gamma^3 \left( \frac{s}{\gamma f(y_1) + s} \right)^2 Y(y_1) \left( \frac{\gamma + 1}{\gamma - 1} - f(y_1) \right) + \gamma^3 \left( \frac{s}{\gamma f(y_2) + s} \right)^2 Y(y_2) f(y_2) \right)$$

(11.1)
Proposition 11.1 Let 

\[ P_t = E \left[ \sum_{\tau=1}^{\infty} \frac{M_{t+\tau}}{M_t} W_{t+\tau} \mid \mathcal{F}_t \right] \]

be the price of equity at time \( t \). The equity price dividend ratio is given by

\[ \frac{P_t}{W_t}(X_t) = E_t[(I - A)^{-1}(X_{t+1})] + O(\sigma^3) \]

In particular, if \( X_t \) are i.i.d., the price dividend ratio takes a simpler form

\[ \log \frac{P_t}{W_t} = C + \gamma f(X_t) \left( \frac{S\gamma}{\gamma f(X_t) + S} \right)^2 Y(X_t) + O(\sigma^3) \]

where \( C \) is an explicitly given constant (see, Appendix).

We calibrate our model to the real data, using the simple, binomial process of Mehra and Prescott (1985).

We assume that \( X_t \) takes values \( u = 1.054 \) and \( d = 0.982 \) with equal probability 0.5.\(^{15}\)

Furthermore, we take \( \rho = 0.02 \) for the discount rate and \( \gamma = 4 \) for the risk aversion. Adopting the findings of Storesletten, Telmer, and Yaron (2004), we use the calibration \( Y(u) = 0.032 \) and \( Y(d) = 0.184 \) for the (countercyclical) variance of idiosyncratic risk in booms and recessions respectively.

Finally, we need a model for the insurance loadings. Recall that, in a static framework, the quantity \( \pi \), defined via

\[ u(w_0 - \pi) = E[(w_0(1 + w))] \]

is the risk premium, required by an agent with the utility function \( u \) for taking the (multiplicative) risk \( w \). In the Arrow-Pratt approximation (see, Gollier (2001), p.22-23), \( \pi \approx 0.5R\text{Var}[w] \) when the risk \( w \) is small. Here, \( R \) is the relative risk aversion. If we imagine that \( u \) is the utility of a representative shareholder of the insurance company, we could think of loadings \( \lambda_t^{-1} \) per endowment as the effective, relative risk aversion of the shareholder. Thus, it is natural to assume that \( \lambda_t^{-1} \) takes values in the interval \([0.5, 5]\), where a reasonable risk aversion should lie. Let \( f(u) = l_1 \), \( f(d) = l_2 \) \( \in [0.2, 2] \), be the values of inverse loadings in the two possible states.

Note that, without insurance loadings, both price dividend ratio and equity premium are constant. We will analyze the impact of insurance loadings on two stylized facts: (1) price dividend ratios are procyclical with ; (2) equity premia are large and countercyclical. For this stylized fact, we will use the

\[^{15}\text{In fact, they allow for a more general Markov structure, described above and account for a small autocorrelation in the aggregate consumption growth. We discuss the effects of this autocorrelation below.}\]
standard measure of equity premium, the Hansen and Jaganathan (1991) bound for the Sharpe ratios.

(1) Substituting the calibration, we get $S \approx 14$. The interval $[0.2, 2]$ always lies below $S/\gamma = 3.5$. Thus, *only countercyclical loadings* (increasing $f$) can overcome the effect of countercyclical idiosyncratic shocks and make the price dividend ratio move *procyclically*. There is some empirical evidence, suggesting that loadings do move countercyclically (see, e.g., Danielsson and Jonsson (2005), Grace and Hotchkiss (1995), Meier and Outreville (2006)). Another source of countercyclicity is the regulator. One expects that the regulatory contribution to $\alpha_t^{-1}$ is countercyclical, rising in times of recession and falling in times of expansion. See, Embrechts, McNeil, and Frey (2005), p. 13 and Embrechts (2001), Danielsson and Jonsson (2005), Grace and Hotchkiss (1995), Meier and Outreville (2006). Changing the values of $l_1, l_2$ we can achieve the observed size of cyclical variation.

(2) Conditional variance of the SDF can be increased by introducing loadings of reasonable size, $l_1, l_2 \in [0.2, 1]$. But, to make this conditional variance countercyclical, we need procyclical loadings, that would imply countercyclical price dividend ratios. Thus, countercyclical equity premia are incompatible with procyclical price-dividend ratios in this simple model. Loadings processes with autocorrelation and/or an endowment processes with stochastic volatility can help solving this problem.

Finally, if we, as Mehra and Prescott (1985), allow for a small autocorrelation in $X_t$, this will generate additional cyclical variation of equity premia and price-dividend ratios. Unfortunately, these effects are very small and do not change the picture. Autocorrelations in consumption growth can have large effects on asset prices through the so-called long run risk factors, but this only becomes relevant when we consider the more general, recursive utility on non-time-additive type. See, e.g., Bansal and Yaron (2004). When the elasticity of intertemporal substitution and risk aversion are decoupled, the whole future stream of aggregate risks appears in the current SDF. By contrast, Theorem 10.1 shows that, because of time additivity, only current risks are important for the current SDF.

Extending our model to the case of recursive utility is a topic of future research.

12 Conclusions

We have analyzed an equilibrium model in which agents can purchase insurance contracts against their idiosyncratic income shocks in addition to financial assets.

The main new mechanism is that the price of an insurance contract depends *nonlinearly* on the claims. We choose the simplest possible variance Ansatz and introduce insurance safety loadings, proportional to variance of unhedgeable risks. The loading factors are random and vary with business
cycles.

We expand the stochastic discount factor (SDF) in the volatility of consumption growth, a natural small parameter and obtain the true dependence of the SDF on idiosyncratic risk and insurance loadings. Our expression allows us to qualify and quantify the effects of cyclicity of insurance loadings on asset prices and, in particular, on the equity premium.
APPENDICES

A The derivatives of the optimal consumption and claims streams

Let \( \Delta = \text{diag}(e^{-\rho t}) \). Note that, for any \( \mathcal{G}_t \)-measurable \( S_t \) with \( P_{\mathcal{G}_t}^t S_t = 0 \) we always have \( Q^t \cdot S_t \cdot P_{\mathcal{G}_t}^t = S_t \cdot P_{\mathcal{G}_t}^t \).

**Lemma A.1** Define the map \( F : H \rightarrow H \) via

\[
F(Z) = \Delta c^{-\gamma} - M J \Delta (M^2 A)^{-1} \cdot (P_{\mathcal{G}_t} c^{-\gamma}) Q J^* M Z
\]

with

\[
c(w, Z) = w + Z - (2 M^2 A)^{-1} P_{\mathcal{G}_t} (Q J^* M Z)^2
\]

Let for all \( t = 1, \cdots, T, \tau := Q^t \sum_{\tau=1}^T Z \cdot M_{\tau} \). Then, the map \( F \) is monotone increasing and the optimal claims stream \( Z \) is determined via

\[
Z = F_Z^{-1}(c_{\gamma}^0 M)
\]

The partial derivatives of \( F \) are given by

\[
\gamma^{-1} \frac{\partial F_t}{\partial Z_t} = -e^{-\rho t} c_t^{-\gamma - 1} + M_t e^{-\rho t} \frac{1}{M_t^2 \alpha_t} c_t^{-\gamma - 1} P_{\mathcal{G}_t}^t \cdot S_t \cdot Q^t
\]

\[
\quad + e^{-\rho t} M_t \frac{1}{M_t^2 \alpha_t} Q^t \cdot S_t \cdot P_{\mathcal{G}_t}^t \cdot c_t^{-\gamma - 1}
\]

\[
\quad - \sum_{\tau=1}^t M_t^2 e^{-\rho \tau} \frac{1}{(M_{\tau}^2 \alpha_{\tau})^2} S_{\tau} \cdot P_{\mathcal{G}_\tau}^\tau \cdot c_{\tau}^{-\gamma - 1} \cdot S_{\tau}
\]

\[
\quad - \gamma^{-1} M_t^2 \sum_{\tau=1}^t e^{-\rho \tau} \frac{1}{M_{\tau}^2 \alpha_{\tau}} (P_{\mathcal{G}_\tau}^\tau c_{\tau}^{-\gamma}) \cdot Q^\tau
\]

and for \( r > t \)

\[
\gamma^{-1} \frac{\partial F_t}{\partial Z_r} = e^{-\rho t} \frac{1}{M_t^2 \alpha_t} c_t^{-\gamma - 1} P_{\mathcal{G}_t}^t \cdot S_t \cdot Q^t \cdot M_{t}
\]

\[
\quad - M_t \sum_{\tau=1}^t e^{-\rho \tau} \frac{1}{(M_{\tau}^2 \alpha_{\tau})^2} S_{\tau} \cdot P_{\mathcal{G}_\tau}^\tau \cdot c_{\tau}^{-\gamma - 1} \cdot S_{\tau} \cdot M_{\tau}
\]

\[
\quad - \gamma^{-1} M_t \sum_{\tau=1}^t e^{-\rho \tau} \frac{1}{M_{\tau}^2 \alpha_{\tau}} (P_{\mathcal{G}_\tau}^\tau c_{\tau}^{-\gamma}) \cdot Q^\tau \cdot M_{\tau}
\]

and \( \frac{\partial F}{\partial Z} \) is, of course, selfadjoint.

Note that the products above must be understood as products of operators and not products of functions.
B  Expansion of the optimal consumption stream for small uncertainty

**Lemma B.1** The Jacobian of the optimal consumption map \( c(w) \) is

\[
\frac{\partial c(w)}{\partial w} = I + \frac{\partial Z}{\partial w} - (M^2 A)^{-1} P_{\mathcal{F}} S Q J^* M \frac{\partial Z}{\partial w} \tag{B.1}
\]

and the second differential is

\[
\frac{\partial^2 c(w)}{\partial w^2}(y, y) = \frac{\partial^2 Z}{\partial w^2} - (M^2 A)^{-1} P_{\mathcal{F}} (Q J^* M \frac{\partial Z}{\partial w}(y))^2 - (M^2 A)^{-1} P_{\mathcal{F}} S Q J^* M \cdot \frac{\partial^2 Z}{\partial w^2} \tag{B.2}
\]

In the complete market case with no uncertainty,

\[
\frac{\partial c(w)}{\partial w} = cm J N^{-1} Q J^* M \quad \text{and} S = 0
\]

and hence

\[
\frac{\partial^2 c(w)}{\partial w^2}(y, y) = \frac{\partial^2 Z}{\partial w^2} - (M^2 A)^{-1} P_{\mathcal{F}} (Q J^* M \frac{\partial Z}{\partial w}(y))^2
\]

Even if there is aggregate uncertainty (that is, there is randomness in the process \( M \)), we still have the important identity \( P_{\mathcal{F}} \frac{\partial c(w)}{\partial w^2} = 0 \).

**Theorem B.2** We have

\[
c(\sigma) = cm + \sigma B w^1 + \frac{1}{2} \sigma^2 (I - B) ((\gamma + 1) cm^{-1}(B w_1)^2 - (M^2 A)^{-1} P_{\mathcal{F}} (Q J^* M (B - I) w_1)^2 + 2 cm J (M^2 A)^{-1} \cdot (P_{\mathcal{F}} (M cm^{-1}(B w_1))) Q J^* M (B - I) w_1) - (M^2 A)^{-1} P_{\mathcal{F}} (Q J^* M (B - I)(w_1))^2 + O(\sigma^3)
\]

C  Expansion of the SDF for geometric random walks

We have (up to an error of order \( \sigma^3 \))

\[
N_t(A) = \gamma M_{h_t} \alpha_t + \sum_{\tau=t}^T \frac{cm_{\tau}}{M_{\tau}}
\]

\[
= e^{-\rho t} W_{t}^{-1-\gamma} P_{\mathcal{F}} \left( \gamma \alpha_t W_{t}^{-1} + \sum_{\tau=t}^T e^{-\rho(\tau-t)} W_{\tau}^{-1-\gamma} \right) + O(\sigma)
\]

\[
= e^{-\rho t} W_{t}^{-1-\gamma} \left( \gamma \alpha_t W_{t}^{-1} + \frac{1}{1 - e^{-\rho} E[X^{1-\gamma}]} \right) \tag{C.1}
\]
Let \( S_0 = E[X^{1-\gamma}] \). A direct calculation shows that
\[
P^t \sum_{\tau=t}^{\infty} e^{-\rho(\tau-t)} \left( \frac{W_\tau}{W_t} \right)^{1-\gamma} \frac{w_\tau^{1-\gamma}}{W_\tau} = \frac{w_t^{1-\gamma}}{W_t} S
\]
with \( S = \frac{1}{1-e^{-\rho S_0}} \) and
\[
(Bw)_t = W \sum_{\tau=1}^{t} \gamma \alpha_{\tau} W_{\tau-1}^{-1} + S \frac{w_{\tau-1}^{1-\gamma}}{W_{\tau-1}^{-1}} (X_{\tau}^{1-\gamma} - 1) \quad (C.2)
\]
Since the projectors \( Q_{t_1}, Q_{t_2} \) are pairwise orthogonal and \( X_t^{1-\gamma} \) are pairwise independent, conditioned on \( F \),
\[
(W^{-2} P_F (Bw)_{-2})_t = \sum_{\tau=1}^{t} \frac{S_{\tau}}{\gamma \alpha_{\tau} W_{\tau-1}^{-1} + S} y_{\tau}^2 + O(\sigma^3)
\]
because \( y_{\tau}^2 = O(\sigma^2) \) for all \( t \). Similar considerations, involving orthogonality and independence, imply that
\[
(Q J^* M_h B w)_t = M_{ht} W_t \frac{S_{\tau}^2}{\gamma \alpha_{\tau} W_{\tau-1}^{-1} + S W_{\tau-1}^{-1}} (X_{\tau}^{1-\gamma} - 1) \quad (C.3)
\]
and thus
\[
M_t = M_{ht} \left( 1 + \frac{\gamma (\gamma + 1)}{2} \sum_{\tau=1}^{t} \frac{S_{\tau}}{\gamma \alpha_{\tau} W_{\tau-1}^{-1} + S} y_{\tau}^2 \right.
\]
\[
- \gamma \alpha_{\tau} W_{\tau-1}^{-1} \left( \frac{S_{\tau} \gamma}{\gamma \alpha_{\tau} W_{\tau-1}^{-1} + S} \right)^2 y_{\tau}^2 \) + O(\sigma^3) \quad (C.4)
\]
Therefore, the stochastic discount factor is given by
\[
\frac{M_{t+1}}{M_t} = \exp \left( -\rho - \gamma \log X_{t+1} \right.
\]
\[
+ \frac{\gamma}{2} \left( \frac{S}{\gamma \alpha_{t+1} W_{t+1}^{-1} + S} \right)^2 \left( (\gamma + 1) - 2 \alpha_{t+1} W_{t+1}^{-1} \gamma^2 \right) y_{t+1}^2
\]
\[
+ \gamma \alpha_{\tau} W_{\tau-1}^{-1} \left( \frac{S_{\tau} \gamma}{\gamma \alpha_{\tau} W_{\tau-1}^{-1} + S} \right)^2 y_{\tau}^2 \) + O(\sigma^3) \quad (C.5)
\]
References


