Intertemporal Substitution and Risk Aversion

Lars Peter Hansen  
University of Chicago  
email: l-hansen@uchicago.edu

John Heaton  
University of Chicago  
email: john.heaton@gsb.uchicago.edu

Nikolai Roussanov  
University of Chicago  
email: nroussan@gsb.uchicago.edu

Junghoon Lee  
University of Chicago  
email: junghoon@uchicago.edu

November 28, 2006

\(^1\)Nan Li provided expert research assistance. We benefitted from comments by Nan Li, Sydney Ludvigson and Jesus Santos. Hansen acknowledges support from the National Science Foundation under Award Number SES0519372. Heaton acknowledges support from the Center for Research in Security Prices.
1 Introduction

Households save and invest both for intertemporal reasons and to control exposure to risk. The resulting patterns of consumption, savings and investment, at both the household and the aggregate level, reveal information about the parameters of preferences that govern intertemporal substitution and risk aversion. Prices that clear financial markets must also reflect the demands of investors and hence are affected by their preferences. In this way security market data convey information from asset prices that complements that from microeconomic data sets, from experimental evidence, or from survey evidence. An important aim of this chapter is to understand better how changes in investor preferences alter asset prices. This guides our understanding of the consequences of inputs from external data sources and the value of asset market data for revealing investor preferences.

Risk premia in security returns provide compensation for risk averse investors. These risk premia often have simple characterizations. For instance, in the capital asset pricing model (CAPM), risk premia are proportional to the covariances between the return to the aggregate wealth portfolio and asset returns. More generally, in the consumption-based capital asset pricing model (CCAPM) the covariance between consumption and asset returns determines the riskiness of returns. Since the dynamics of consumption are linked to the dynamics of wealth, this model implies that understanding the riskiness of the wealth of investors is ultimately important in understanding security returns. This linkage is heavily influenced by the assumed form of investor preferences.

While asset market data offer fertile proving ground for theories of investor behavior and market structure, this data source also pose special challenges or puzzles. In the case of the CAPM, differences across securities in the measured covariance of returns with aggregate stock market indices have been shown to have little relationship with average returns (see for example Fama and French (1992)). Similarly there appears to be very little covariance between measures of the aggregate consumption of investors, and asset returns. The empirical results in Grossman and Shiller (1981), Hansen and Singleton (1983), Mehra and Prescott (1985), Shiller (1982) and Hansen and Jagannathan (1991) give alternative characterizations of puzzles that emerge in the study of security market returns and aggregate consumption. Thus, when we look to security market data for information about preference parameters, we are exposed to the empirical challenges posed by this source of data.

Our chapter features alternative and complementary methods of analysis for the study of the macroeconomic underpinning of asset valuation. We describe some alternative ways to characterize model implications, and we show how statistical methods can be put to good use. While we apply some of these methods to illustrate substantive results, our chapter is not intended as comprehensive of empirical evidence. Excellent surveys with more extensive empirical discussions are given in Campbell (2003) and Lettau and Ludvigson (2003).

Alternative specifications of investor preferences and their links to prices are given in sections 2 and 3. Specifically in section 3 we show how to construct stochastic discount
factors used in representing prices for the alternative models of investor preferences described in section 2. While we describe the investor preferences for an array of models, we focus our equilibrium price calculations and quantification on a particular subset of these preferences, the CES Kreps and Porteus (1978) model. This model is rich enough to draw an interesting distinction between risk aversion and intertemporal substitution and to pose important measurement and econometric challenges. Some basic statistical methods for characterizing present-value implications as they relate to asset pricing are developed in section 4. Section 5 develops some analytical results and local approximations designed to reveal how intertemporal substitution and risk aversion alter equilibrium prices. Section 6 uses vector-autoregressive (VAR) statistical models to measure risk aversion from a heterogenous set of asset returns and quantifies the resulting statistical uncertainty. Section 7 develop generalized method of moments (GMM) and related estimation methods and illustrate their use in extracting measures of intertemporal substitution and risk aversion. These latter sections add some important qualifications to the existing empirical literature.
2 Investor Preferences

In this section we survey a variety of models of investor preferences that are used in the literature. These specifications of investor preferences imply, through their intertemporal marginal rates of substitution, stochastic discount factors that represent risk prices and interest rates. This discussion is complementary to the Backus et al. (2004) survey of exotic preferences pertinent to macroeconomics. As in which follows, they provide alternative specifications of intertemporal and risk preferences.\(^1\)

Recursive utility gives a useful framework for parameterizing risk aversion and intertemporal substitution. As advocated by Epstein and Zin (1989a) and Constantinides (1990), it gives a convenient way to introduce parameters that serve distinct roles in decision making. Let \(\{\mathcal{F}_t : t \geq 0\}\) denote the sequence of conditioning information sets (sigma algebras) available to an investor at dates \(t = 0, 1, \ldots\). Adapted to this sequence are consumption processes \(\{C_t : t \geq 0\}\) and a corresponding sequence of continuation values \(\{V_t : t \geq 0\}\) associated with this consumption process. The date \(t\) components \(C_t\) and \(V_t\) are restricted to be in the date \(t\) conditioning information set.\(^2\) The continuation values are determined recursively and used to rank alternative consumption processes.

Consider three approaches. The first approach takes a risk adjustment of the continuation value; the second approach introduces intertemporal complementarities; and the third approach social externalities.

2.1 Risk adjustment

Consider investor preferences that can be represented recursively as

\[ V_t = \psi(C_t, V_{t+1} | \mathcal{F}_t) \]

where \(C_t\) is current consumption. Given a consumption process, this recursion takes future values and maps them into current values. It requires a terminal condition for the continuation value to initiate a backward induction. A familiar example is:

\[ V_t = (1 - \beta)U(C_t) + \beta E(V_{t+1} | \mathcal{F}_t) \]

where \(U\) is a concave utility function. This recursion is additive in expected utility. More general depictions of recursive utility, provide a way to allow for alternative adjustments to risk and uncertainty.

\(^1\)While Backus et al. (2004) do an admirable job of describing a broad class of preference specifications and their use in macroeconomics, the empirical challenge is how to distinguish among these alternatives. As Hansen (2004) emphasizes, some specifications are inherently very difficult to distinguish from one another.

\(^2\)More formally, \(C_t\) and \(V_t\) are restricted to be \(\mathcal{F}_t\) measurable.
2.1.1 A smooth adjustment

Following Kreps and Porteus (1978) and Epstein and Zin (1989a) introduce a strictly increasing, smooth concave function $h$. In applications this function is typically:

$$h(V) = \begin{cases} \frac{V^{1-\gamma}-1}{1-\gamma} & \gamma > 0, \gamma \neq 1 \\ \log V & \gamma = 1 \end{cases}$$

Then a risk adjusted version of the continuation value is:

$$R(V|\mathcal{F}) = h^{-1}(E[h(V)|\mathcal{F}]).$$

The presumption is that $V_t$ depends on the continuation value through the risk adjustment $R(V_{t+1}|\mathcal{F}_t)$, which is a restriction on function $\psi$:

$$V_t = \psi(C_t, V_{t+1}|\mathcal{F}_t) = \psi^*[C_t, R(V_{t+1}|\mathcal{F}_t)].$$

The function $h$ is strictly increasing and adjusts for the riskiness of the continuation value for the consumption profile $\{C_{t+\tau} : \tau = 1, 2, \ldots\}$. It imposes a nontrivial preference over lotteries indexed by calendar time. The parametric form of $h$ gives a convenient way to parameterize risk preferences.

Consider the special case in which the continuation value is perfectly predictable, implying that $E(V_{t+1}|\mathcal{F}_t) = V_{t+1}$. Then $R(V_{t+1}|\mathcal{F}_t) = V_{t+1}$ so that the function $h$ has no bearing on the specification of preferences over perfectly forecastable consumption plans. The incremental risk adjustment does alter the implications for intertemporal substitution for predictable consumption plans.

Examples of $\psi^*$ function are as follows:

**Example 2.1.**

$$\psi^*(C, R) = (1 - \beta)U(C) + \beta R$$

for some increasing concave function $U$.

The concavity of $U$ already induces some degree of risk aversion, but it also has an impact on intertemporal substitution.

It is often convenient to work with an aggregator that is homogeneous of degree one. Curvature in $U$ can be offset by transforming the continuation value. In the case of a constant elasticity of substitution (CES) specification this gives:

**Example 2.2.**

$$\psi^*(C, R) = \left[(1 - \beta)(C)^{1-\rho} + \beta(R)^{1-\rho}\right]^\frac{1}{1-\rho}$$

for $\rho > 0$. The case in which $\rho = 1$ requires taking limits, and results in the Cobb-Douglas specification:

$$\psi^*(C, R) = C^{1-\beta}R^\beta.$$

The parameter $\rho$ is the reciprocal of the elasticity of intertemporal substitution.
Example 2.3. There is an extensive literature in control theory starting with the work of Jacobson (1973) and Whittle (1990) on introducing risk sensitivity into control problems. Hansen and Sargent (1995) suggest a recursive version of this specification in which

\[ \psi^*(C, R) = U(C) + \beta R \]

as in example 2.1 with the incremental risk adjustment given by

\[ R(V_{t+1} | F_t) = - \frac{1}{\theta} \log E \left[ \exp(-\theta V_{t+1}) | F_t \right]. \]

The parameter \( \theta \) is the risk sensitivity parameter. As emphasized by Tallarini (1998), this specification overlaps with the CES specification when \( \rho = 1, U(C) = \log C \) and \( \theta = \gamma - 1 \).

To verify this link, take logarithms of the continuation values in the CES recursions. The logarithmic function is increasing and hence ranks of hypothetical consumption processes are preserved.

Although it is convenient to make a risk adjustment of the continuation value, there is an alternative transformation of the continuation value that depicts preferences as a nonlinear version of expected utility. Let

\[ \tilde{V}_t = h(V_t). \]

Then

\[ \tilde{V}_t = h \left[ \psi^* \left( C_t, h^{-1} \left[ E \left( \tilde{V}_{t+1} | F_t \right) \right] \right) \right] = \tilde{\psi} \left[ C_t, E \left( \tilde{V}_{t+1} | F_t \right) \right] \]

The introduction of \( h \) can induce nonlinearity in the aggregator \( \tilde{\psi} \). Kreps and Porteus (1978) use such a nonlinear aggregator to express a preference for early and late resolution of uncertainty. When \( \tilde{\psi} \) is convex in this argument there is a preference for early resolution of uncertainty and conversely when \( \tilde{\psi} \) is concave. We will show that the intertemporal composition of risk also matters for asset pricing.

2.1.2 A version without smoothness

The Epstein and Zin (1989a) recursive formulation was designed to accommodate more fundamental departures from the standard expected utility model. This includes departures in which there are kinks in preferences inducing first-order risk aversion. First-order risk aversion is used in asset pricing as a device to enhance risk aversion.

Examples of applications in the asset pricing include Bekaert et al. (1997) and Epstein and Zin (1990), but we shall feature a more recent specification due to Routledge and Zin (2003). Routledge and Zin (2003) propose and motivate an extension of Gul (1991)’s preferences for disappointment aversion. These preferences are based on a different way to
compute the risk adjustment to a continuation value and induce first order risk aversion. Continuation values are risk adjusted in accordance to:

\[ h(\hat{V}) = E[h(V)|\mathcal{F}] + \alpha E \left( 1_{\{V-\delta\hat{V} \leq 0\}} \left[ h(V) - h(\delta\hat{V}) \right] |\mathcal{F} \right) \]

which is an implicit equation in \( \hat{V} \). In this equation, \( 1 \) is used as the indicator function of the subscripted event. The random variable \( h(\hat{V}) \) is by construction less than or equal to the conditional expectation of \( h(V) \) with an extra negative contribution coming because of the averaging over the bad events defined by the threshold \( h(V) \leq h(\delta V) \). The risk adjusted value is defined to be:

\[ R(V|\mathcal{F}) = \hat{V}. \]

The \( h \) function is used as a risk adjustment as in our previous construction, but the parameters \( 0 < \delta < 1 \) and \( \alpha > 0 \) capture a notion of disappointment aversion. While the Gul (1991) specification assumes that \( \delta = 1 \), this limits the preference kink to be on the certainty line. By allowing \( \delta \) to be less than one, the disappointment cutoff is allowed to be lower.

### 2.2 Robustness and uncertainty aversion

Investors may be unsure about the probability used to evaluate risks. Instead of using one model, they may choose a family of such models. In some circumstances this also leads to what looks like a risk adjustment in the continuation value to a consumption plan. We illustrate this using the well known close relationship between risk sensitivity and robustness featured in the control theory literature starting with the work of Jacobson (1973). As in Hansen and Sargent (1995) we may formulate this recursively as:

\[ v_t = (1 - \beta)U(C_t) + \min_{q_{t+1} \geq 0, E[q_{t+1}|\mathcal{F}_t] = 1} \left[ \beta E(q_{t+1}v_{t+1}|\mathcal{F}_t) + \beta\theta E(q_{t+1}\log q_{t+1}|\mathcal{F}_t) \right] \]

where \( \theta \) is a penalization parameter and \( q_{t+1} \) is a random variable used to distort the conditional probability distribution. The minimization is an adjustment for uncertainty in the probability model, and \( E[q_{t+1}(\log q_{t+1})|\mathcal{F}_t] \) is a discrepancy measure for the probability distortion called conditional relative entropy.

The solution to the minimization problem is to set:

\[ q_{t+1} \propto \exp\left( -\frac{V_{t+1}}{\theta} \right) \]

where the proportional constant is conditional on \( \mathcal{F}_t \) and chosen so that \( E(q_{t+1}|\mathcal{F}_t) = 1 \). This solution gives an exponential tilt to the original conditional probability distribution based on the continuation value and penalty parameter \( \theta \). Substituting this minimized choice of \( q_{t+1} \) gives the recursion:

\[ v_t = (1 - \beta)U(C_t) + \beta h^{-1} E[h(v_{t+1})|\mathcal{F}_t] \]

(1)
where
\[ h(V) = \exp(-V/\theta). \]

Hence this setting is equivalent to assuming an exponential risk adjustment in the continuation value function.

As emphasized by Tallarini (1998), when \( U \) is the logarithmic function, we may transform the continuation value of (1) to obtain the Cobb-Douglas recursion in example 2.2 with \( \theta = \frac{\gamma - 1}{\gamma} \) and \( V_t = \exp(v_t) \). Maenhout (2004) and Skiadas (2003) give a characterization of this link in more general circumstances that include the CES specification in a continuous time version of these preferences by making the penalization depend on the endogenous continuation value (see also Hansen (2004)).

Strictly speaking, to establish a formal link between inducing a concern about model misspecification and a concern about risk required a special set of assumptions. These results illustrate, however, that it may be difficult in practice to disentangle the two effects. What may appear to be risk aversion emanating from asset markets may instead be a concern that a conjectured or benchmark probability model is inaccurate. Risk aversion from asset market data may be different from risk aversion in an environment with well defined probabilities.

There are other ways to model uncertainty aversion. Following Epstein and Schneider (2003) we may constrain the family of probabilities period by period instead penalizing deviations. If we continue to use relative entropy, the constrained worst case still entails exponential tilting, but \( \theta \) becomes a Lagrange multiplier that depends on date \( t \) information. The recursion must subtract off \( \beta \theta \) times the entropy constraint. As demonstrated by Petersen et al. (2000) and Hansen et al. (2006), a time invariant parameter \( \theta \) may be interpreted as a Lagrange multiplier of an intertemporal constraint, in contrast to the specifications advocated by Epstein and Schneider (2003).

The challenge for empirical work becomes one estimating penalization parameters or alternatively the size of constraints on the families of probabilities. These objects replace the incremental risk adjustments.

### 2.3 Intertemporal Complementarity and Social Externalities

Consider next a specification with intertemporal complementarities. Introduce a habit stock, which we model as evolving according to:

\[ H_t = (1 - \lambda)C_t + \lambda H_{t-1} \]

where \( \lambda \) is a depreciation factor and \( H_t \) is a geometric average of current and past consumptions. In building preferences, form an intermediate object that depends on both current consumption and the history of consumption:

\[ S_t = \left[ \delta(C_t)^{1-a} + (1 - \delta)(H_t)^{1-a} \right]^{-\frac{1}{a}} \]
where $\alpha > 0$ and $0 < \delta < 1$. Construct the continuation value recursively via:

$$V_t = [(1 - \beta)(S_t)^{1-\rho} + \beta[R(V_{t+1}|F_t)]^{1-\rho}]^{\frac{1}{1-\rho}}.$$ 

Alternatively, $H_t$ may be used as a subsistence point in the construction of $S_t$ as in:

$$S_t = C_t - \delta H_t.$$ 

Typically $R(\cdot|F) = E(\cdot|F)$, and this specification is used as a distinct way to separate risk aversion and intertemporal substitution. Intertemporal substitution is now determined by more than just $\rho$: in particular the preference parameters $(\delta, \alpha)$ along with $\rho$ and the depreciation factor $\lambda$. The parameter $\rho$ is typically featured as the risk aversion parameter.

Preferences of this general type in asset pricing have been used by Novales (1990), Constantinides (1990), Heaton (1995) and others. Novales used it to build an equilibrium model of real interest rates, but used a specification with quadratic adjustment costs in consumption. Instead of using CES specification, Constantinides and Heaton use $H_t$ to shift the subsistence point in the preferences to study the return differences between equity and bonds. It remains an open issue as to how important these various distinctions are in practice.

When the consumer accounts for the effect of the current consumption choice on future values of the habit stock, the habit effects are internal to the consumer. Sometimes the habit stock $H_t$ is taken to be external and outside the control of the consumer. The habit stock serves as a social reference point. Examples include Abel (1990) and Campbell and Cochrane (1999).
3 Stochastic Discount Factors

In this section we describe how investor preferences become encoded in asset prices via stochastic discount factors. Our use of stochastic discount factor representations follows Harrison and Kreps (1979) and Hansen and Richard (1987) and many others. For the time being we focus on one-period pricing and hence one-period stochastic discount factors; but subsequently we will explore multi-period counterparts. Multi-period stochastic discount factors are built by forming products of single period stochastic discount factors.

3.1 One-period Pricing

Consider the one-period pricing of elements $X_{t+1}$ in a space of asset payoffs. An asset payoff is a bundled (across states) claim to a consumption numeraire over alternative states of the world that are realized at a future date. Thus payoffs $x_{t+1} \in X_{t+1}$ depend on information available at $t+1$. Mathematically they are depicted as a random variable in the date $t+1$ conditioning information set of investors. The time $t$ price of $x_{t+1}$ is denoted by $\pi_t(x_{t+1})$ and is in the date $t$ information set $\mathcal{F}_t$ of investors.

Hansen and Richard (1987) give restrictions on the set of payoffs and prices for there to exist a representation of the pricing function of the form:

$$E(S_{t,t+1}x_{t+1}|\mathcal{F}_t) = \pi_t(x_{t+1}) \quad (2)$$

where $\mathcal{F}_t$ is the current conditioning information set which is common across investors. These restrictions allow investors to use information available at date $t$ to trade in frictionless markets. The positive random variable $S_{t,t+1}$ is a stochastic discount factor used to price assets. It discounts asset payoffs differently depending on the realized state in a future time period. Consequently, this discounting encompasses both the discounting of known payoffs using a risk-free interest rate and the adjustments for risk. As argued by Harrison and Kreps (1979) and others, the existence of a positive stochastic discount factor follows from the absence of arbitrage opportunities in frictionless markets.

A common and convenient empirical strategy is to link stochastic discount factors to intertemporal marginal rates of substitution. We illustrate this for a two-period economy, but we will deduce formulas for dynamic economies in subsequent presentation.

Example 3.1. Suppose that investor $j$ maximizes the utility function:

$$E[w^j(c_t^j, c_{t+1}^j)|\mathcal{F}_t]$$

by trading financial claims. Let $(\tilde{c}_t^j, \tilde{c}_{t+1}^j)$ be the optimal consumption choices for this consumer. Consider a perturbation of this consumption bundle in the direction $(\tilde{c}_t^j - c_t^j, \tilde{c}_{t+1}^j - c_{t+1}^j).$
\[ r \pi_t(x_{t+1}), \bar{c}_{t+1} + r x_{t+1} \] which is parameterized by the real number \( r \). Notice that this change in consumption is budget neutral for all choices of \( r \). Differentiating with respect to \( r \), at the optimal choices we have:

\[
E[u_1(c_t, \bar{c}_t)|F_t] \pi_t(x_{t+1}) = E[u_2(c_t, \bar{c}_t+1) x_{t+1}|F_t].
\]

As a result

\[
E(M_{t,t+1} x_{t+1} | F_t) = \pi_t(x_{t+1})
\]

(3)

where the intertemporal marginal rate of substitution:

\[
M_{t,t+1} = \frac{u_2(c_t, \bar{c}_t+1)}{E[u_1(c_t, \bar{c}_t)|F_t]}.
\]

This same argument applies to any feasible perturbation and hence (3) is applicable to any payoff as long as the perturbation away from the optimal that we explored is permitted.

This gives a link between important economic quantities and asset prices.

Note that

\[
E[(M_{t,t+1} - M_{i,t+1})x_{t+1}|F_t] = 0
\]

for all investors \( j \) and \( i \). Therefore any difference in the marginal rates of substitution across agents are orthogonal to the payoff space \( X_{t+1} \).

Suppose now that \( X_{t+1} \) includes any bounded function that is measurable with respect to a sigma algebra \( G_{t+1} \) that is contained in \( F_{t+1} \). Then this orthogonality implies:

\[
E(M_{t,t+1} | G_{t+1}) = S_{t,t+1}
\]

for all \( j \). The stochastic discount factor is unique if it is restricted to be measurable with respect to \( G_{t+1} \). More generally, any of the intertemporal marginal rates of substitution of the investors can be used as a stochastic discount factor to depict prices. One economically important example of the difference between \( G_{t+1} \) and \( F_{t+1} \) is the case where there are traded claims to aggregate uncertainty but claims to individual risk are not. Therefore there is limited risk-sharing in financial markets in this economy.\(^4\)

Suppose that investors can trade contracts contingent on any information that is available as of date \( t + 1 \). Further suppose that these investors do not face any trading frictions such as transactions costs or short-sale constraints. Under this complete market specification \( G_{t+1} = F_{t+1} \) and \( F_{t+1} \) includes all individuals’ information. In this case

\[ M_{t,t+1} = S_{t,t+1} \]

and \( S_{t,t+1} \) is unique. The marginal rates of substitution are equated across investors.

For pedagogical simplicity we compute shadow prices. That is we presume that consumption is fixed at some determined process. Subsequently, we will have to add specificity to this process, but for the time being we remain a bit agnostic. It can be the outcome of a decentralized production economy, but we place production considerations on the back burner.

\(^4\) See, for example, Constantinides and Duffie (1996).
3.2 CES Benchmark

Consider an economy with complete markets and investors with identical preferences of this CES type. In what follows we will use the common preference specification to deduce a formula for the stochastic discount factor. For the recursive utility model with a CES specification, it is convenient to represent pricing in two steps. First we value a contingent claim to next period’s continuation value. We then change units from continuation values to consumption by using the next-period marginal utility for consumption. In all cases, marginal utilities are evaluated at aggregate consumption. The CES specification makes these calculations easy and direct.

Because the CES recursion is homogeneous of degree one in its arguments, we can use Euler’s Theorem to write:

$$V_t = (MC_t)C_t + E[(MV_{t+1})V_{t+1}|\mathcal{F}_t].$$

(4)

Claims to future continuation values $V_{t+1}$ can be taken as substitutes for claims to future consumption processes. When current consumption be the numeraire, equilibrium wealth is given by $W_t \equiv V_t/MC_t$. Divide (4) by $MC_t$ to obtain a recursive expression for wealth:

$$W_t = C_t + E[S_{t,t+1}W_{t+1}|\mathcal{F}_t].$$

The marginal utility of consumption is:

$$MC_t = (1 - \beta)(C_t)^{-\rho}(V_t)^\rho,$$

and the marginal utility of next-period continuation value is:

$$MV_{t+1} = \beta(V_{t+1})^{-\gamma}[R(V_{t+1}|\mathcal{F}_t)]^{-\rho}(V_t)^\rho.$$

(5)

Forming the intertemporal marginal rate of substitution gives:

$$S_{t,t+1} = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\rho} \left[ \frac{V_{t+1}}{R(V_{t+1}|\mathcal{F}_t)} \right]^{\rho - \gamma}.$$

When we incorporate kinks in preferences as in setting suggested by Routledge and Zin (2003), the marginal utility of next-period continuation value is:

$$MV_{t+1} = \beta(V_{t+1})^{-\gamma}[R(V_{t+1}|\mathcal{F}_t)]^{-\rho}(V_t)^\rho \left[ 1 + \alpha \mathbf{1}_{\{(V_{t+1})^{1-\gamma} \leq [\delta R(V_{t+1}|\mathcal{F}_t)]^{1-\gamma}\}} \left( 1 - \delta^{1-\gamma} \alpha E \left( (V_{t+1})^{1-\gamma} \mathbf{1}_{\{(V_{t+1})^{1-\gamma} \leq [\delta R(V_{t+1}|\mathcal{F}_t)]^{1-\gamma}\}|\mathcal{F}_t} \right) \right) \right].$$

Combining these terms, the one-period intertemporal marginal rate of substitution is

$$S_{t,t+1} = \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\rho} \left[ \frac{V_{t+1}}{R(V_{t+1}|\mathcal{F}_t)} \right]^{\rho - \gamma} \left[ 1 + \alpha \mathbf{1}_{\{(V_{t+1})^{1-\gamma} \leq [\delta R(V_{t+1}|\mathcal{F}_t)]^{1-\gamma}\}} \left( 1 - \delta^{1-\gamma} \alpha E \left( (V_{t+1})^{1-\gamma} \mathbf{1}_{\{(V_{t+1})^{1-\gamma} \leq [\delta R(V_{t+1}|\mathcal{F}_t)]^{1-\gamma}\}|\mathcal{F}_t} \right) \right) \right].$$

The stochastic discount factor depends directly on current consumption, and indirectly on future consumption through the continuation value.

We now consider some special cases of the CES version of the Kreps-Porteus model:
Example 3.2. Let \( \rho = \gamma \) and \( \alpha = 0 \). Then the contribution to the continuation value drops out from the stochastic discount factor. This is the model of Lucas (1978) and Breeden (1979).

Example 3.3. Consider the special case with \( \rho = 1 \) and \( \alpha = 0 \), but allow \( \gamma \) to be distinct from one. Then the counterpart to the term: \( \left( \frac{V_{t+1}}{R_t(V_{t+1}|\mathcal{F}_t)} \right)^{\rho-\gamma} \) entering the stochastic discount factor is

\[
\frac{(V_{t+1})^{1-\gamma}}{E[(V_{t+1})^{1-\gamma}|\mathcal{F}_t]}.
\]

Notice that this term has conditional expectation equal to unity.

Example 3.4. Consider the special case in which \( \gamma = 1 \) and \( \alpha = 0 \), but allow \( \rho \) to be distinct from one. In this case the counterpart to the term \( \left( \frac{V_{t+1}}{R_t(V_{t+1}|\mathcal{F}_t)} \right)^{\rho-\gamma} \) entering the stochastic discount factor is:

\[
\left[ \frac{V_{t+1}}{\exp E (\log V_{t+1}|\mathcal{F}_t)} \right]^{\rho-1}
\]

The logarithm of this term has expectation zero.
4 Empirical Observations from Asset Returns

Time series observations of asset returns and consumption are needed to identify the parameters governing the preferences of consumers. The stochastic discount factor developed in section 3 and its implication for security prices impose a set of joint restrictions on asset prices and consumption. Before analyzing these restrictions, we first display some important empirical regularities from asset markets alone. Besides standard sample statistics for asset returns we also examine some standard decompositions of prices. These are based on a log-linear approximation and the present-value relationship.

This decomposition was proposed by Campbell and Shiller (1988a, 1988b) and Cochrane (1992). The methods have been used extensively in the finance literature to summarize statistical evidence about dividend-price ratios, dividend growth and returns. We develop these methods and show their link to related work in the macroeconomics literature by Hansen et al. (1991). We then apply these decompositions to an important set of test assets.

4.1 Log linear approximation and present values

The price of a security at time \( t \) is given by \( P_t \). The return to this security from time \( t \) to time \( t+1 \) is determined by the cash flow received at time \( t+1 \), denoted \( D_{t+1} \) and the price of the security at time \( t+1 \), denoted \( P_{t+1} \). The return is given by:

\[
R_{t+1} = \frac{D_{t+1} + P_{t+1}}{P_t} = \left( \frac{D_{t+1}}{D_t} \right) \left( \frac{1 + P_{t+1}/D_{t+1}}{P_t/D_t} \right). \tag{6}
\]

The cash flow, \( D_{t+1} \) is the dividend in the case of stocks or a coupon in the case of bonds. Although many individual companies do not pay dividends, our empirical analysis is based on the analysis of portfolios of stocks and these dividends will be positive.

This allows us to take logarithms of (6). Using lower case letters to denote logarithms of each variable we have:

\[
r_{t+1} = (d_{t+1} - d_t) - (p_t - d_t) + \log[1 + \exp(p_{t+1} - d_{t+1})]. \tag{7}
\]

We view this as a difference equation for the logarithm of the price dividend ratio with forcing processes given by the returns and dividend growth rate. The use of returns as a forcing process allow us to deduce some statistical restrictions. The valuation models of section 3 determines both the prices and the returns endogenously.

To make (7) a linear difference equation, consider the approximation:

\[
\log[1 + \exp(p_{t+1} - d_{t+1})] \approx \log[1 + \exp(\mu_{p-d})] + \kappa (p_{t+1} - d_{t+1} - \mu_{p-d}) \tag{8}
\]

where

\[
\kappa = \frac{\exp(\mu_{p-d})}{1 + \exp(\mu_{p-d})} < 1.
\]
and $\mu_{p-d}$ is a constant centering point for the linearization. This point is usually taken to be the mean of the logarithm of the dividend price ratio which will be different for alternative cash flows because of differences in cash flows and discount rates.

Substituting approximation (8) into the difference equation (7) and rearranging terms:

$$ p_t - d_t = (d_{t+1} - d_t) - r_{t+1} + \kappa (p_{t+1} - d_{t+1}) + c $$

(9)

where

$$ c = \log[1 + \exp(\mu_{p-d})] - \kappa \mu_{p-d}. $$

For the remainder of this section, we will ignore the approximation error and treat (9) as the difference equation of interest.

Solving (9) forward gives:

$$ p_t - d_t = \sum_{j=0}^{\infty} (\kappa)^j [d_{t+j+1} - d_{t+j} - r_{t+j+1}] + \frac{c}{1 - \kappa}. $$

Notice that the constant term in this solution satisfies the approximation:

$$ \frac{c}{1 - \kappa} \approx \mu_{p-d}. $$

### 4.1.1 Moving-average models

The implications of the linear difference equation for returns will be examined using simply linear time series models. We therefore assume that there is a first-order Markov process for a state vector $x_t$ where the dynamics are given by:

$$ x_{t+1} = Ax_t + Bw_{t+1} $$

(10)

where $\{w_{t+1} : t = \ldots, 0, 1, \ldots\}$ is a sequence of iid normally distributed random vectors with $E w_{t+1} = 0$ and $E(w_{t+1} w_{t+1}') = I$. The matrix $A$ is assumed to have eigenvalues with absolute values less than 1. This assumption implies a stochastic steady state for $x_t$ where $x_t$ is a moving-average of current and past shocks:

$$ x_t = \sum_{j=0}^{\infty} A^j B w_{t-j} = \sum_{j=0}^{\infty} A^j BL^j w_t \equiv \mathcal{A}(L) w_t $$

where $L$ denotes the “lag” operator.

Dividends and returns and prices are linked to the state vector $x_t$ via:

$$ d_{t+1} - d_t = \mu_d + G_d x_t + H_d w_{t+1} $$

$$ r_{t+1} = \mu_r + G_r x_t + H_r w_{t+1} $$

$$ p_t - d_t = \mu_{p-d} + G_{p-d} x_t. $$
The present-value model implies restrictions on this representation, which we now explore. We will derive these restrictions in two ways. Substitute these depictions into (9) and obtain:

\[ G_{p-d}x_t = (G_d - G_r + \kappa G_{p-d}A)x_t \]
\[ 0 = (H_d - H_r + \kappa G_{p-d}B)w_{t+1} \]

Since these restrictions must hold for all realized values of \( x_t \) and \( w_{t+1} \), these two equations restrict directly the representation for dividends, returns and the price-dividend ratio.

To obtain an alternative perspective on these restrictions, we use the implied moving-average representations. In stochastic steady state, dividends and returns satisfy:

\[ d_{t+1} - d_t = \delta(L)w_{t+1} + \mu_d \]
\[ r_{t+1} = \rho(L)w_{t+1} + \mu_r \]
\[ p_t - d_t = \pi(L)w_t + \mu_{p-d} \]

where

\[ \delta(z) = \sum_{j=0}^{\infty} \delta_j z^j \quad \sum_{j=0}^{\infty} |\delta_j|^2 < \infty \]
\[ \rho(z) = \sum_{j=0}^{\infty} \rho_j z^j \quad \sum_{j=0}^{\infty} |\rho_j|^2 < \infty \]
\[ \pi(z) = \sum_{j=0}^{\infty} \pi_j z^j \quad \sum_{j=0}^{\infty} |\pi_j|^2 < \infty \]

The variable \( z \) is introduced so that we may view \( \delta(z), \rho(z), \pi(z) \) as power series. They are sometimes referred to as the \( z \)-transforms of the moving-average coefficients. The coefficients of the power series are the moving-average coefficients. The power series converge at least on the domain \( |z| < 1 \).

In this case, the coefficients of the power series \( \delta(z) \) and \( \rho(z) \) are given by:

\[ \delta_0 = H_d \quad \rho_0 = H_r \]
\[ \delta_j = G_d A^{j-1} B \quad \rho_j = G_r A^{j-1} B. \]

Hence

\[ \delta(z) = H_d + z G_d (I - zA)^{-1} B \]
\[ \rho(z) = H_r + z G_r (I - zA)^{-1} B. \]

Difference equation (9) implies that

\[ z\pi(z) = \delta(z) - \rho(z) + \kappa \pi(z). \]  

This is an equation that restricts the moving average coefficients. We may evaluate these functions at \( z = \kappa \):

\[ \kappa \pi(\kappa) = \delta(\kappa) - \rho(\kappa) + \kappa \pi(\kappa). \]

This implies that

\[ \delta(\kappa) = \rho(\kappa). \]
Using the power series representation of $\rho$ and $\delta$, this implies that the discounted (by $\kappa$) impulse responses for returns and cash flow growth rates must be equal. This is the present-value-budget-balance restriction of Hansen et al. (1991). This restriction is necessary in order that the future shocks to cash flow growth rates and to returns net out so that the price-dividend ratio only depends on current and past shocks.

Under the Markov representation of the state variable $x_t$, the restriction:

$$\rho(\kappa) = \delta(\kappa)$$

becomes:

$$H_r + \kappa G_r (I - \kappa A)^{-1}B = H_d + \kappa G_d (I - \kappa A)^{-1}B .$$

The moving average representation for the price-dividend ratio is obtained by solving equation (11) for $\pi$:

$$\pi(z) = \frac{\delta(z) - \rho(z)}{z - \kappa} .$$

(13)

Because of the denominator term, the right-hand side looks like it explodes at $z = \kappa$. This is not the case, however. The numerator is also zero at $z = \kappa$. After dividing out the common zero at $\kappa$, $\pi$ will have a well defined power series for $|z| < 1$, and formula (13) for $\pi(z)$ is a valid formula for the $z$-transform of the moving-average coefficients. Performing this division is consistent with the formula:

$$G_{p-d} = (G_d - G_r)(I - \kappa A)^{-1}$$

used in representing the price-dividend ratio.

This “solution” is a bit unusual. It takes returns and dividend growth as given and solves for the price-dividend ratio. A structural asset pricing model does in fact have different primitives. Even when cash flows are given exogenously, returns and price-dividend ratios are both determined endogenously. The rationale for “solving” the model in this manner is instead a way to allow for prices or returns to reveal additional information used by investors to forecast future cash flows. It is a restriction imposed on a moving-average representation of the shocks that are pertinent to the investors’ decision-making.

4.1.2 Decompositions

This solution for $\pi$ is often used to motivate empirical decompositions of prices and measurement of return risk.

1. Return decomposition - The risk in returns from time $t$ to time $t+1$ is captured by the term $\rho_0 w_{t+1}$. Since $\rho(\kappa) = \delta(\kappa)$,

$$\rho_0 = \delta(\kappa) - \sum_{j=1}^{\infty} \kappa^j \rho_j .$$
Hence one period exposure to risk has both a discounted cash flow component and a component due to return predictability. When return predictability is not very strong, the discounted impact of shocks on future dividends is the most important source of risk. In addition if $\kappa$ is close to one, $\delta(\kappa)$ measures the accumulated impact of current shocks on dividends far into the future. This measure of long-run risk is featured in the work of Bansal et al. (2005) and Hansen et al. (2005).

2. Price-dividend decomposition - Using $\rho(\kappa) = \delta(\kappa)$, express $\pi$ as

$$\pi(z) = \left[ \frac{\delta(z) - \delta(\kappa)}{z - \kappa} \right] - \left[ \frac{\rho(z) - \rho(\kappa)}{z - \kappa} \right].$$

The first term is the discounted expected future cash flow growth and the second is the discounted expected future returns both net of constants. This decomposition is used to measure the importance of discounted cash flows in accounting for variation in the price-dividend ratio. This decomposition was originally proposed by Campbell and Shiller (1988a, 1988b).

4.1.3 Identifying shocks

For the restriction on the joint dynamics of returns, dividends and prices to be testable, we must be able to identify shocks. Vector autoregressive (VAR) methods are commonly used in conjunction with other restrictions to identify shocks. Hansen et al. (1991) show that there is tension, however, between the use of VAR methods to identify shocks and the present-value-budget-balance implications that are imposed in the log-linear model.

Let $y_t$ be a vector of observables with moving average representation:

$$y_{t+1} = B(L)w_{t+1} + \mu_y.$$ 

To construct $w_{t+1}$ from $y_{t+1}, y_t, ...$ requires that $B(z)$ be of full rank for $|z| < 1$. In vector autoregressive applications, it is typically assumed that $y$ and $w$ have the same number of entries. In this case $B(z)$ must be nonsingular for $z < 1$, and, in particular, $B(\kappa)$ must be nonsingular. If $y_{t+1}$ contains $d_{t+1} - d_t$ and $r_{t+1}$ as the first two entries, then $\delta(\kappa) = \rho(\kappa)$ implies that

$$\begin{bmatrix} 1 & -1 & 0 \end{bmatrix} B(\kappa) = 0$$

which violates the restriction that $B(z)$ be nonsingular. Returns do not contain enough information to reveal shocks along with dividend growth. This is the dividend-return counterpart to a claim established in Hansen et al. (1991), and it gives a warning against using VAR methods in conjunction with dividends and returns alone.

Let $y_{t+1}$ include cash flow growth rates $d_{t+1} - d_t$ and the price dividend ratio $p_{t+1} - d_{t+1}$. Given the implied moving-average representations from a state-space model or a VAR form:

$$p_t - d_t = \pi(L)w_t + \mu_{p-d}$$
\[ d_{t+1} - d_t = \delta(L) w_{t+1} + \mu_d. \]

In this case construct the moving-average representation for the approximate return via:

\[ r_{t+1} = \rho(L) w_{t+1} + \mu_r \]

where \( \rho(z) = \delta(z) + (\kappa - z) \pi(z) \). This necessarily satisfies the present-value restriction (12). Thus we sidestep the informational inconsistency mentioned previously by using prices to reveal shock components instead of returns.

4.2 Test Assets

To illustrate the construction of these returns we use the prices, returns and dividends constructed from six portfolios. The portfolios returns and dividends are constructed as in Hansen et al. (2005).

The first portfolio is a market portfolio of stocks traded on the NYSE and NASDAQ. The other portfolios are constructed by sorting stocks on the basis of book value relative to market value of equity as in Fama and French (1992). Five portfolios with equal numbers of stocks in each portfolio are constructed from the entire universe of stocks. Dividends are then constructed from the return series for each portfolio with and without dividends. This construction is done on a quarterly basis from 1947 to 2005. Because of the pronounced seasonalness in dividends, dividends are smoothed over a year. Details of the data construction can be found in Hansen et al. (2005).

Table 1 reports summary statistics for the five book-to-market portfolios (portfolios “1” through “5”). Notice that portfolio 1 has the lowest average book-to-market value (B/M) and the highest average price-divided ratio (P/D) and the lowest average return. Moving from portfolio 1 to portfolio 5, the average book-to-market value increases, the average price-divided ratio declines and the average return increases. As we will see, differences in the average returns are not explained by exposure to contemporaneous covariance with consumption.

4.2.1 Vector Autoregression

We first consider a statistical decomposition of the price-dividend ratio for each portfolio using vector autoregressions. To do this let:

\[ y \equiv \begin{bmatrix} p_t - d_t \\ d_t - d_{t-1} \end{bmatrix}. \]

We fit a VAR of the form:

\[ y_t = A_0 + A_1 y_{t-1} + \cdots + A_l y_{t-l} + B w_t \]

where the two-dimensional shock vector \( w_t \) has mean zero and covariance matrix \( I \). Further \( A_0 \) is a two-dimensional, the matrices \( A_j, j = 1, 2, \ldots, l \) and \( B \) are two by two.
Properties of Portfolios Sorted by Book-to-Market

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-period Exp. Return (%)</td>
<td>6.79</td>
<td>7.08</td>
<td>9.54</td>
<td>9.94</td>
<td>11.92</td>
<td>7.55</td>
</tr>
<tr>
<td>Long-Run Return (%)</td>
<td>8.56</td>
<td>8.16</td>
<td>10.72</td>
<td>10.84</td>
<td>13.01</td>
<td>8.77</td>
</tr>
<tr>
<td>Avg. B/M</td>
<td>0.32</td>
<td>0.61</td>
<td>0.83</td>
<td>1.10</td>
<td>1.80</td>
<td>0.65</td>
</tr>
<tr>
<td>Avg. P/D</td>
<td>51.38</td>
<td>34.13</td>
<td>29.02</td>
<td>26.44</td>
<td>27.68</td>
<td>32.39</td>
</tr>
</tbody>
</table>

Table 1: Data are quarterly from 1947 Q1 to 2005 Q4 for returns and annual from 1947 to 2005 for B/M ratios. Returns are converted to real units using the implicit price deflator for nondurable and services consumption. Average returns are converted to annual units using the natural logarithm of quarterly gross returns multiplied by 4. “One-period Exp. Return,” we report the predicted quarterly gross returns to holding each portfolio in annual units. The expected returns are constructed using a separate VAR for each portfolio with inputs: $(c_t - c_{t-1}, e_t - c_t, r_t)$ where $r_t$ is the logarithm of the gross return of the portfolio. We imposed the restriction that consumption and earnings are not Granger caused by the returns. One-period expected gross returns are calculated conditional on being at the mean of the state variable implied by the VAR. “Long-Run Return” reports the limiting value of the logarithm of the expected long-horizon return from the VAR divided by the horizon. “Avg. B/M” for each portfolio is the average portfolio book-to-market over the period computed from COMPSTAT. “Avg. P/D” gives the average price-dividend for each portfolio where dividends are in annual units.
We further impose the normalization that $B$ is lower triangular so that the second shock (the second element of $w_t$) does not impact the price-divided ratio contemporaneously.

This VAR implies linear dynamics for the Markov process $x_t$. To see this, let

$$\mu \equiv E(y_t) = (I - A_1 - \cdots A_l)^{-1}A_0$$

and

$$y_t^* \equiv y_t - \mu .$$

Then $x_t$ is given by:

$$x_t = \begin{bmatrix} y_t^* \\ y_{t-1}^* \\ \vdots \\ y_{t-l}^* \end{bmatrix},$$

$$G \equiv \begin{bmatrix} A_1 & A_2 & \cdots & A_l \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix}$$

and

$$H \equiv \begin{bmatrix} B \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$  

For each portfolio we estimate a VAR with $l = 5$ and consider the properties of portfolio cash flows and prices using estimated impulse response functions.

Figure 1 reports the impulse response functions for the market. The top panel of the figure reports the response of the level of log dividends to the two shocks. The first shock has an immediate effect on dividends and then the response builds going forward. The second shock has a very small effect on dividends. The second panel of the figure reports the response of the log price-dividend ratio to the shocks. Notice that the first shocks has a very little effect on the price-dividend ratio, while the second shock increases the price dividend ratio and the impact persists for many periods. The pattern of responses indicates that the two shocks can be labeled as a dividend shocks and a separate price-dividend shock. Shocks to the price-dividend ratio are long-lasting and have little ability to forecast future dividends. This reflects the well known inability of the price-dividend ratio at the aggregate level to forecast future dividends.

The bottom panel of figure 1 reports the implied response of returns to the two shocks. To better understand the effects of the shocks, the results are reported for the cumulative impact of the shocks on returns. Notice that the dividend shock (shock 1) has little effect on returns while the price-dividend shock has an initial positive impact on returns followed by a slowly building negative impact on returns in the future. For
Figure 1: Impulse Response Functions for the Market Portfolio. Top panel: response of log dividends to shocks. Middle panel: response of the log price-dividend ratio to shocks. Bottom panel: response of returns to shocks. – depicts impulse responses the first shock. — depicts impulse responses to the second shock.
the market portfolio, variation in the price-dividend ratio has some predictive ability for future returns, while variation in dividends that have no effect on prices, has little ability to forecast future returns.

These results are interpreted by Campbell and Shiller (1988a, 1988b) and others as implying that variation in future returns is the most important factor explaining variation in the price-dividend ratio. Further this variation is empirically independent of variation in future dividends. This implies that for this aggregate portfolio variation in the price-dividend ratio must be driven by required returns. This has potentially important implications for the stochastic discount factor of section 3.

The corresponding impulse response functions for portfolios 1 and 5 are reported in figures 2 and 3 respectively. Notice that for these portfolio the labeling of the two shocks as dividend and return shocks is not clear. For example shocks to dividends now have an ability to forecast future returns. As portfolio returns and dividends are disaggregated the predictability of dividends rises. This fact is emphasized in the work of Vuolteenaho (2002).
Figure 3: Impulse Response Functions for the Portfolio 5. Top panel: response of log dividends to shocks. Middle panel: response of the log price-dividend ratio to shocks. Bottom panel: response of returns to shocks. — depicts impulse responses the first shock. —. depicts impulse responses to the second shock.
5 Intertemporal Substitution and Pricing

To understand how investor preference parameters and the stochastic environment influence asset prices, we explore further the solution of the CES version of the Kreps-Porteus model for fixed, prespecified consumption process as in a Lucas-style *endowment economy*. We derive some approximation results where we approximate around a unitary intertemporal substitution parameter $\rho = 1$ for an arbitrary value of $\gamma > 0$. Thus we feature the role of this parameter in our characterizations. As in Restoy and Weil (1998) consumption dynamics plays a central role in these characterizations. For some specifications of consumption dynamics we obtain a structural model of the type analyzed in section 4.

Our expansion in $\rho$ follows in part the work of Kogan and Uppal (2001).\(^5\) The economy we study is different from that of Kogan and Uppal (2001), but they suggest that extensions such as those developed here would be fruitful. By approximating around $\rho$, we are approximating around a stochastic economy with a constant consumption wealth ratio. As we will see, the $\rho = 1$ limit economy leads to other less dramatic simplifications that we exploit in characterizing asset prices and risk premia. The simplifications carry over the $\rho$ derivatives that we calculate for asset prices and returns. While Campbell and Viceira (2002) (chapter 5) show the close connection between approximation around the utility parameter $\rho = 1$ and approximation around a constant consumption-wealth ratio for portfolio problems, there are some interesting differences in our application. Moreover, $\rho = 1$ is inconveniently ruled out in the parameterization of recursive utility considered by Restoy and Weil (1998) and others because of their use of the return-based Euler equation.

We consider first a family of discrete-time economies with log-linear consumption dynamics indexed by $\rho$. When we introduce stochastic volatility in consumption, we find it more convenient to explore an family of economies specified in continuous time. We illustrate these economies using some parameter values extracted from existing research.

5.1 Discrete time

The initial step in our calculation is the first-order expansion of the continuation values in the parameter $\rho$. Let $v^\rho_t$ denote the logarithm of the continuation value for intertemporal substitution parameter $\rho$, and let $c_t$ denote the logarithm of consumption. We rewrite the CES recursion as

$$v^\rho_t = \frac{1}{1 - \rho} \log \left[ (1 - \beta) \exp[(1 - \rho)c_t] + \beta \exp \left[ (1 - \rho) Q_t(v^\rho_{t+1}) \right] \right],$$

(14)

where $Q_t$ is:

$$Q_t(v_{t+1}) = \frac{1}{1 - \gamma} \log E \left[ \exp \left[ (1 - \gamma)v_{t+1} \right] | F_t \right].$$

\(^5\)Our $\rho$ derivatives will be heuristic in the sense that we will not provide a rigorous development of their approximation properties.
When $\rho = 1$ this recursion simplifies to:

$$v_t^1 = (1 - \beta)c_t + \beta Q_t(v_{t+1}^1).$$  \hspace{1cm} (15)

### 5.1.1 Continuation values

We compute the first-order expansion:

$$v_t^\rho \approx v_t^1 + (\rho - 1)Dv_t^1$$

where $v_t^1$ is the continuation value for the case in which $\rho = 1$ and the notation $D$ denotes differentiation with respect to $\rho$. We construct an approximate recursion for $Dv_t^1$ by expanding the logarithm and exponential functions in (14) and including up to second-order terms in $c_t$ and $Q_t$. The approximate recursion is:

$$v_t^\rho \approx (1 - \beta)c_t + \beta Q_t(v_t^\rho) + \beta(1 - \rho)(1 - \beta)\left[Q_t(v_{t+1}^\rho) - c_t\right]^2. \hspace{1cm} (16)$$

As is evident from (15), this approximation is exact when $\rho = 1$.

Our aim to construct an exact recursion for the derivative of $v_t$ with respect to $\rho$. One way to do this is to differentiate directly (14). It is simpler to differentiate the approximate recursion (16) for the logarithm of the continuation value $v_t$ with respect to $\rho$. This is valid because the approximation error in the recursion does not alter the derivative with respect to $\rho$. Performing either calculation gives:

$$Dv_t^1 = -\beta(1 - \beta)\left[Q_t(v_{t+1}^1) - c_t\right]^2 + \beta E^*(Dv_{t+1}^1|F_t)$$

$$= -\frac{(1 - \beta)(v_t^1 - c_t)^2}{2\beta} + \beta E^*(Dv_{t+1}^1|F_t) \hspace{1cm} (17)$$

where $E^*$ is the distorted expectation operator associated with a Radon-Nikodym derivative

$$M_{t,t+1} = \frac{\exp [(1 - \gamma)v_{t+1}^1]}{E(\exp [(1 - \gamma)v_{t+1}^1]|F_t)}. \hspace{1cm} (18)$$

The Radon-Nikodym derivative is a measure-theoretic notion of a derivative. Since $M_{t,t+1}$ is a positive random variable with conditional expectation one, it induces a distorted probability by scaling random variables. For instance, the distorted expectation of a random variable is:

$$E^*(z_{t+1}|F_t) = E(M_{t,t+1}z_{t+1}|F_t)$$

Solving recursion (17) forward gives the derivative $Dv_t^1$. This derivative is necessarily negative. By using the distorted expectation operator $E^*$ to depict the recursion for $Dv_t^1$, the recursion has a familiar form that is convenient for computing solutions.

To implement this approach we must compute $v_t^1$ and the distorted conditional expectation $E^*$, which will allow us to solve (17) for $Dv_t^1$. Later we give some examples when this is straightforward.
5.1.2 Wealth expansion

When $\rho$ is different from one, the wealth-consumption ratio is not constant. Write

$$W_t = \frac{V_t^\rho}{(1-\beta)(C_t)^{\rho}p^\rho} = \frac{(C_t)^\rho(V_t^\rho)^{1-\rho}}{1-\beta}.$$  

A first-order expansion of the continuation value implies a second-order expansion of the wealth-consumption ratio. This can be seen by taking logarithms and substituting in the first-order approximation for the continuation value.

$$\log W_t - \log C_t = -\log(1-\beta) + (1-\rho) [v_t^1 - c_t + (\rho-1)Dv_t^1]$$  

$$= -\log(1-\beta) - (\rho-1)(v_t^1 - c_t) - (\rho-1)^2 Dv_t^1.$$  \hspace{1cm} (19)

The first-order term of (19) compares the logarithm of the continuation value for $\rho = 1$ with the logarithm of consumption. The continuation value is forward looking and time varying. Thus when future looks good relative to the present, the term $v_t^1 - c_t$ can be expected to be positive. When the intertemporal elasticity parameter $\rho$ exceeds one, the first order term implies that a promising future relative to the present has an adverse impact on equilibrium wealth and conversely when $\rho$ is less than one. As we will see, the term $v_t^1$ is very similar to (but not identical to) the term typically used when taking log-linear approximations.\(^6\)

By construction, the second-order term adjusts the wealth consumption ratio in a manner that is symmetric about $\rho = 1$, and it is always positive.

5.1.3 Stochastic discount factor expansion

Consider next the first-order expansion of the logarithm of the stochastic discount factor:

$$s_{t+1,t}^\rho \approx s_{t+1,t}^1 + (\rho-1)Ds_{t+1,t}^1.$$  

Recall that the log discount factor is given by:

$$s_{t+1,t}^\rho = \log \beta - \rho (c_{t+1} - c_t) + (\rho - \gamma) [v_{t+1}^\rho - Q_t(v_{t+1}^\rho)].$$

Differentiating with respect to $\rho$ gives:

$$Ds_{t+1,t}^1 = - (c_{t+1} - c_t) + [v_{t+1}^1 - Q_t(v_{t+1}^1)] + (1-\gamma) [Dv_{t+1}^1 - E^* (Dv_{t+1}^1|F_t)].$$  \hspace{1cm} (20)

Thus we obtain the approximation:

$$s_{t,t+1}^\rho \approx s_{t,t+1}^1 + (\rho-1)Ds_{t+1,t}^1 = \log \beta - \rho (c_{t+1} - c_t) + (\rho - \gamma) [v_{t+1}^1 - Q_t(v_{t+1}^1)].$$

\(^6\)In log-linear approximation the discount rate in this approximation is linked to the mean of the wealth consumption ratio. In the $\rho$ expansion, the subjective rate of discount is used instead.
\[+(1 - \gamma)(\rho - 1) \left[ Dv^1_{t+1} - E^* (Dv^1_{t+1}|\mathcal{F}_t) \right] \]

This shows how changes in \( \rho \) alters one period risk prices. For instance consider approximating one period prices of contingent claim \( z_{t+1} \) to consumption:

\[ E \left[ \exp(s^1_{t+1}) z_{t+1} | \mathcal{F}_t \right] = E \left[ \exp(s^1_{t+1}) z_{t+1} | \mathcal{F}_t \right] + (\rho - 1) E \left[ \exp(s^1_{t+1}) Ds_{t+1} z_{t+1} | \mathcal{F}_t \right] . \]

We will explore the ramifications for local risk prices subsequently when we consider a continuous time counterpart to this expansions. This will provide us with formulas for how \( \rho \) alters risk premia.

### 5.1.4 Log-linear dynamics

To show the previous formulas can be applied, consider the following evolution for consumption in the log linear Markov economy:

\[ x_{t+1} = Ax_t + Bw_{t+1} \]
\[ c_{t+1} - c_t = \mu_c + G'x_t + H'w_{t+1} \]

where \( \{w_{t+1} : t = 0, 1, \ldots\} \) is an iid sequence of standard normally distributed random vectors. Recall that for \( \rho = 1 \), the continuation value must solve:

\[ v^1_t = (1 - \beta)c_t + \beta Q_t(v^1_{t+1}). \]

Conjecture a continuation value of the form:

\[ v^1_t = U^v x_t + \mu_v + c_t. \]

Given this guess and the assumed normality,

\[ Q_t(v^1_{t+1}) = U^v A' x_t + \mu_v + \mu_c + G'x_t + c_t + \frac{1 - \gamma}{2} |U^v B + H'|^2 \]

Thus

\[ U^v = \beta A' U^v + \beta G \]

and

\[ \mu_v = \beta \left[ \mu_c + \mu_v + \frac{1 - \gamma}{2} |U^v B + H'|^2 \right] . \]

Solving for \( U^v \) and \( \mu_v \),

\[ U^v = \beta (I - \beta A')^{-1} G, \]
\[ \mu_v = \frac{\beta}{1 - \beta} \left[ \mu_c + \frac{(1 - \gamma)}{2} |H' + \beta G'(I - A\beta)^{-1} B|^2 \right] . \]  

(21)
For $\rho = 1$ the formulas for the continuation value have simple interpretations. The formula for $U_v$ is also the solution to the problem of forecasting the discounted value of future consumption growth:

$$U_v \cdot x_t = \sum_{j=1}^{\infty} \beta^j E (c_{t+j} - c_{t+j-1} - \mu_c | x_t) = (1 - \beta) \sum_{j=1}^{\infty} \beta^j E (c_{t+j} | F_t) - \beta c_t - \left( \frac{\beta}{1 - \beta} \right) \mu_c.$$ 

Therefore,

$$v^1_t = (1 - \beta) \sum_{j=0}^{\infty} \beta^j E (c_{t+j} | F_t) + \frac{\beta(1 - \gamma)}{2(1 - \beta)} | H' + \beta G'(I - A\beta)^{-1}B|^2$$

The log of the continuation value is a geometric weighted average of current and future logarithm of consumption using the subjective discount factor in the weighting. In addition there is a constant risk adjustment. When consumption growth rates are predictable, they will induce movement in the wealth-consumption ratio as reflected in formula (19). The coefficient on the first-order term in $\rho - 1$ compares the expected discounted average of future log consumption to that of current log consumption. If this geometric average future consumption is higher than current consumption and $\rho$ exceeds one, the optimistic future induces a negative movement in the wealth-consumption ratio. Conversely a relatively optimistic future induces a positive movement in the wealth-consumption ratio when $\rho$ is less than one.

The constant risk correction term

$$\frac{\beta(1 - \gamma)}{2(1 - \beta)} | H' + \beta G'(I - A\beta)^{-1}B|^2$$

entering the continuation value is negative for large values of $\gamma$. Consequently, this adjustment enhances the wealth consumption ratio when $\rho$ exceeds one. In the log-linear consumption dynamics, this adjustment for risk induced by $\gamma$ is constant. An important input into this adjustment is the vector:

$$H + \beta B'(I - \beta A')^{-1}G.$$  

(22)

To interpret this object, notice that the impulse response sequence for consumption growth to a shock $w_{t+1}$: $H'w_{t+1}, G'Bw_{t+1}, G'ABw_{t+1}, ...$. Then (22) gives the discounted impulse response vector for consumption. It is the variance of this discounted response vector (discounted by $\beta$) that enters the constant term of the continuation value as a measure of the risk.

The formulas that follow provide the ingredients for the second-order adjustment in the wealth-consumption ratio and the first-order adjustment in risk adjusted prices.

We use the formula for the continuation value to infer the distorted expectation operator. The contribution of the shock $w_{t+1}$ to $(1 - \gamma)v^1_{t+1}$ is given by $(1 - \gamma)(H +
$B'U_v')w_{t+1}$. Recall that $w_{t+1}$ is a multivariate standard normal. By a familiar complete-the-square argument:

$$
\exp \left[ (1 - \gamma)(H + B'U_v)'w - \frac{1}{2}w'w \right]
\propto \exp \left( -\frac{1}{2} [w - (1 - \gamma)(H + B'U_v)]' [w - (1 - \gamma)(H + B'U_v)] \right).
$$

The left-hand side multiplies the standard normal by the distortion implied by (18) up to scale. The right-hand side is the density of the normal up to scale with mean $(1 - \gamma)(H + B'U_v)$ and covariance matrix $I$. This latter probability distribution is the one used for the distorted expectation operator $E^*$ when computing the derivative of the continuation value. Under this alternative distribution, we may write

$$
w_{t+1} = (1 - \gamma)(H + B'U_v) + w^*_{t+1}
$$

where $w^*_{t+1}$ is a standard normal distribution. As consequence, consumption and the Markov state evolve as:

$$
x_{t+1} = Ax_t + (1 - \gamma)B(H + B'U_v) + Bw^*_t + Bw^*_{t+1}
$$
$$
c_{t+1} - c_t = G'x_t + \mu + (1 - \gamma)H'(H + B'U_v) + H'w^*_t + H'w^*_{t+1}.
$$

5.1.5 Example Economies

To illustrate the calculations we consider two different specifications of consumption dynamics that include predictable components to consumption growth rates. One of these is extracted from Bansal and Yaron (2004) but specialized to omit time variation in volatility. Later we will explore specifications with time varying volatility after developing a continuous time counterpart to these calculations. This specification is designed to capture properties of consumption variation of the period 1929 to 1998 and is specified at a monthly frequency. The second specification is obtained from an estimation in Hansen et al. (2005). In this specification quarterly post World War II data is used. This data is described in appendix D.

The first specification is:

$$
c_{t+1} - c_t = .0015 + x_t + [.0078 0] w_{t+1}
$$
$$
x_{t+1} = .98x_t + [0 .00034] w_{t+1}.
$$

There are two shocks, one directly impacts on consumption and the second one on the conditional mean of consumption. In the Breeden (1979) - Lucas (1978) specification of preferences with power utility, only the first shock will have a local price that is different from zero. In the recursive utility the second shock will also have a nonzero price because of the role of the continuation value to the local prices.

Figure 4 reports the impulse response functions for consumption in reaction to the two shocks. The first shock by construction has a significant immediate impact that is
permanent. The second shock has a relatively small initial impact on consumption but the effect builds to a significant level. With recursive utility this long-run impact can produce a potentially large effect on risk prices especially since the effect can be magnified by choice of the risk aversion parameter $\gamma$.

The second specification is inferred by fitting a vector autoregression of $c_{t+1} - c_t$ and the logarithm of the ratio $c_{t+1} - e_{t+1}$ of consumption to corporate earnings. It is important in this specification that corporate earnings and consumption are cointegrated with a coefficient of one. Most models of aggregate growth yield this restriction. These is also empirical support for our assumption. For example, consider figure 5 which reports an approximate Bayesian posterior distribution for the parameter $\lambda$ where $c_{t+1} - \lambda e_{t+1}$ is assumed to be stationary. This distribution was calculated using the technique described in appendix B. Notice that the distribution of $\lambda$ is centered very close to one. There is some variation around this point but it is very minor so that restricting $\lambda = 1$ is empirically grounded.

In this model there are also two shocks. We identify one as being proportional to the one-step ahead forecast error to consumption scaled to have a unit standard deviation. The second shock is uncorrelated with this first shock and has no immediate impact on
Figure 5: Approximate posterior distribution for cointegration parameter. Construction uses Box-Tiao priors for each equation of the VAR for consumption and corporate earnings. The posterior distribution is for the parameter $\lambda$ where $c_{t+1} - \lambda e_{t+1}$ is assumed to be stationary. The histogram is scaled to integrate to one.
consumption. Figure 6 reports the estimated response of consumption to the two shocks. Notice that both shocks induce important long-run responses to consumption that are different from the short-run impulse. For example, the long-run response of consumption to its own shock is almost twice the immediate response. As in the Bansal-Yaron model, consumption has an important low-frequency component. With recursive preferences this low-frequency component can have important impact on risk premia.

We can identify shocks using an alternative normalization that emphasizes long-run effects. In particular we identify one shock from the VAR that has a transient effect with no impact on consumption in the long run. The other shock is uncorrelated with this transient shock and has permanent consequences for consumption. The impulse response function of consumption to these two shocks is displayed in figure 7. Notice that the long-run response to a permanent shock is almost twice the immediate response to this shock.

Although the VAR does identify an important long-run shock to consumption, there is

---

7 This approach is an adaptation of the identification scheme advocated by Blanchard and Quah (1989).
Figure 7: Impulse Responses of Consumption to Permanent and Temporary Shocks. — depicts impulse response to a permanent shock. –. depicts impulse response to a temporary shock.
substantial statistical uncertainty surrounding this estimate. To assess these uncertainty we use the technique discussed in appendix B. Figure 8 reports the approximate posterior distributions for the immediate response of consumption to the temporary shock along with the long-run response of consumption to a permanent shock. Notice that the long-run response is centered at a larger value but that there is uncertainty about this value. The short-run response is measured with much more accuracy.

5.2 Wealth and Asset Price Variation

As we saw in section 4 pricing models need to imply significant variation in the stochastic discount factor in order to be consistent with some important empirical regularities from financial markets. We also see this when examining aggregate wealth and consumption.

When $\rho = 1$ the ratio of consumption to wealth is constant. As we change $\rho$, this ratio varies. For the alternative models of the dynamics of consumption, we examine whether the pricing model can result in significant variation in the wealth-consumption ratio. This is an important issue because aggregate wealth varies significantly over time due to variation in the market value of wealth. For example in figure 9 we plot the ratio of wealth to consumption quarterly from 1952 to 2005. Aggregate wealth is measured as the difference between financial wealth and financial liabilities for the household sector of the US economy. This measure of wealth does not include other types of wealth such as human capital.

Notice that there is significant variation in the wealth to consumption ratio. Much of this variation is due to the variability of the market value of traded equity. For example during the late 1990 there was a significant increase in the value of the US stock market which resulted in a substantial increase in the wealth to consumption ratio during this period. With the decline in equity values the wealth to consumption ratio has come back down.

5.2.1 Wealth Variation

We now examine the model’s implication for wealth when $\rho$ differs from one. We are interested in the effects of alternative parameter values on the predicted level of wealth, the variation in wealth over time and the response of wealth to shocks.

Consider the implications for the wealth-consumption ratio using the dynamics from the VAR with consumption and corporate earnings. Properties of the log wealth-consumption ratio implied by the VAR and the CES model are given in table 2 for $\gamma$ and $\beta$ fixed at 2 and $0.99^{1/4}$ respectively. Several different values of $\rho$ are considered.

Notice that variation in $\rho$ has a significant impact on the forecasted level of the wealth-consumption ratio. Given a value for $\beta$ this variation could be used to identify $\rho$ based on the observed mean of the ratio. Variation in the mean of the wealth-consumption ratio induced by $\rho$ can be unwound by choice of $\beta$, however. Of interest then is the effect of $\rho$ on the dynamics of the wealth-consumption ratio.
Figure 8: Approximate posterior distributions for responses. The top panel gives the approximate posterior for the immediate response to consumption and the bottom panel the approximate posterior for the long-run response of consumption to the permanent shock. Construction uses Box-Tiao priors for each equation. The histograms are scaled to integrate to one.
Figure 9: Wealth-Consumption Ratio from 1952 to 2006.
Table 2: Properties of the Log Wealth-Consumption Ratio. The parameters $\gamma$ and $\beta$ are fixed at 5 and 0.99^{1/4}$ respectively. Statistics are calculated via simulation based on a times-series simulation with 60,000 draws of the random vector $w_t$. The first 10,000 draws were discarded.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>Mean</th>
<th>STD</th>
<th>STD w/o 2nd order term</th>
<th>Corr. with Consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>9.16</td>
<td>0.0092</td>
<td>0.0086</td>
<td>0.22</td>
</tr>
<tr>
<td>0.67</td>
<td>7.78</td>
<td>0.0060</td>
<td>0.0057</td>
<td>0.22</td>
</tr>
<tr>
<td>0.9</td>
<td>6.39</td>
<td>0.0017</td>
<td>0.0017</td>
<td>0.23</td>
</tr>
<tr>
<td>1.1</td>
<td>5.70</td>
<td>0.0017</td>
<td>0.0017</td>
<td>-0.23</td>
</tr>
<tr>
<td>1.33</td>
<td>5.50</td>
<td>0.0054</td>
<td>0.0057</td>
<td>-0.23</td>
</tr>
<tr>
<td>1.5</td>
<td>5.74</td>
<td>0.0079</td>
<td>0.0086</td>
<td>-0.24</td>
</tr>
</tbody>
</table>

The row “STD” reports the standard deviation of the wealth-consumption ratio which is increasing in the difference between $\rho$ and 1. The row below that ignores term with $(\rho - 1)^2$ in the expansion (19). Notice that this “second-order” term provides little extra variation in the wealth-consumption ratio. Although variation in $\rho$ away from unity does produce variation in the wealth-consumption ratio, this variation is nowhere near the size observed in the data.

The first order term in the wealth-consumption ratio (19) indicates that shocks to the continuation value affect the wealth-consumption ratio and the sign of the effect depends on the value of $\rho$ relative to 1. In the consumption dynamics estimated by HHL, positive shocks to consumption also have positive impact on the continuation value relative to consumption. When $\rho$ is less than 1 this model predicts a positive covariance between shocks to consumption and wealth. This is reflected in the last line of table 2 which reports the correlation between the log wealth-consumption ratio and the log consumption growth. Notice that when $\rho$ is less than 1, this correlation is positive. When $\rho$ is greater than 1, this correlation is negative.

To further examine this effect we report the impulse response of the log wealth-consumption ratio with reaction to the two shocks in the VAR in figure 10. In constructing these impulse response functions we ignored the second order terms in (19).

Consistent with the correlations between consumption growth and the wealth-consumption ratio reported in table 2 we see that when $\rho$ is less than 1 a positive shock to consumption has a positive effect on the wealth-consumption ratio. These shocks have positive risk prices in the model and hence a claim on aggregate wealth has a potentially significant risk premium.

The specification considered by Bansal and Yaron (2004) predicts a similar pattern of responses to shocks. Figure 11 reports the response of wealth-consumption ratio to a one standard deviation shock to predicted consumption. Since the first shock has no impact
Figure 10: Implied Impulse Responses of Wealth-Consumption Ratio, Hansen-Heaton-Li Model. — depicts impulse response to a consumption shock. -.. depicts impulse response to a earnings shock. The parameters $\gamma$ and $\beta$ are set at 5 and $0.99^{1/4}$ respectively.
on the state variable the response of wealth-consumption ratio to it is zero in this model. Notice as in the dynamics estimated by HHL the direction of the response of wealth to a predicted consumption shock depends critically upon the size of \( \rho \) relative to unity. When \( \rho \) is less than one, the wealth-consumption ratio increases with the shock to predicted consumption. As a result this endogenous price moves positively with consumption and the return on the wealth portfolio is riskier than under the assumption that \( \rho = 1.5 \).

Since wealth is linked to the continuation value, observed wealth can also be used to identify long-run shocks to consumption. We estimate a bivariate VAR for logarithm consumption growth and the logarithm of the observed wealth-consumption ratio reported in figure 9. Figure 12 reports the estimated impulse response functions for consumption and wealth implied by this alternative bivariate VAR. As with corporate earnings, the wealth-consumption ratio identifies a potentially important long-run shock to consumption. Notice, however, that the shock to wealth has a very substantial temporary effect on wealth. There is substantial transitory variation in wealth that does not affect consumption as noted by Lettau and Ludvigson (2004).

The relationship between wealth and consumption predicted by the first-order terms of (19) and \( \rho \) imposes a joint restriction on the impulse response functions of wealth and consumption. Because of the substantial response of wealth to its own shock, this restriction cannot be satisfied for reasonable values of \( \rho \). As we will see below the necessary variation in \( \rho \) results in implausible levels of returns and the wealth-consumption ratio. Ignoring this shock we can examine the restriction of (19) based on the consumption shock alone.

To do this we construct the spectral density of \( w_t - c_t - (1 - \rho)(v_{t-1} - c_t) \) implied by the VAR but setting the variance of the wealth shock to zero. The model implies that at the true value of \( \rho \) this density function should be flat. The predicted density is displayed in figure 13 for \( \rho = 0.5 \) and \( \rho = 1.5 \). Smaller values of \( \rho \) come closer to satisfying the restriction than the large values of \( \rho \) as we will see in section 7.

5.2.2 Measurement of Wealth

Inferences drawn from the recursive utility model based on direct measures of aggregate wealth are sensitive to the wealth proxy used. With a fully specified model of the dynamics of consumption, we circumvent this issue since we can construct implied continuation values and the stochastic discount factors needed to price any series of cash flows. We are therefore able to examine the model’s implications for any part of aggregate wealth once we specify the dynamics of the cash flows accruing to the wealth component.

A particularly important part of aggregate wealth is human capital which by its nature is not included in direct measures of wealth. Unobserved human capital may move in a way that offsets variation in measured wealth so that the true wealth to consumption ratio is relatively constant as predicted by the recursive utility model with \( \rho \) close to one. Lustig and Nieuwerburgh (2006) use this idea to infer the dynamics of unobserved human capital. As an alternative we specify a dynamic model of the cash flows produced
Figure 11: Impulse Responses of Wealth-Consumption Ratio to Predicted Consumption Shock, Bansal-Yaron Model. The parameters $\gamma$ and $\beta$ are set at 5 and 0.998 respectively.
Figure 12: Impulse Responses of Consumption and Wealth. Results from bivariate VAR with consumption growth and the wealth-consumption ratio. — depicts the response to a consumption shock. —. depicts the response to a wealth shock.
Figure 13: Spectral Density of $w_t - c_t - (1 - \rho)(v_t^1 - c_t)$. Results are from a bivariate VAR with consumption growth and the wealth-consumption ratio. The variance of wealth shocks is set to zero. — depicts the density for $\rho = 0.5$. -. depicts the density when $\rho = 1.5$. 
Table 3: Summary statistics for corporate and human capital. Statistics are reported for the natural logarithm of each measure of capital relative to consumption.

<table>
<thead>
<tr>
<th>Capital Measure</th>
<th>Standard Deviation</th>
<th>Correlation with Corporate Capital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Capital</td>
<td>0.056</td>
<td>0.56</td>
</tr>
<tr>
<td>Corporate Capital</td>
<td>0.033</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>0.034</td>
<td>0.70</td>
</tr>
</tbody>
</table>

By human capital.

In our analysis we assume that these cash flows are given by labor income. We measure labor income as “Wages and salary disbursements” as reported by the National Income and Product Accounts. As with corporate earnings, we impose the restriction that labor income and aggregate consumption are cointegrated with a unit coefficient. We further assume that $\beta = 0.99^{1/4}$, $\gamma = 5$ and $\rho = 1$.

The model’s implication for the standard deviation of the (log) ratio of human capital to consumption is reported in table 3. For comparison the corresponding standard deviation for the ratio of “corporate capital” to consumption is also calculated by valuing the stream of corporate earnings. This measure of wealth does not correspond to any direct measure of the value of capital held by the corporate sector since corporate earnings do not account for investment. Further earnings are reported after payments to bond holders. Finally in the table “Total” refers to the ratio of the sum of human plus corporate capital to consumption.

Although there are issues of interpretation with these measures of capital, notice that the implied standard deviations are different from zero and that the ratio of human capital to consumption has the greatest variance. In contrast to the analysis of Lustig and Nieuwerburgh (2006), human and corporate capital are predicted to be positively correlated. Further, although the model does predict variation in these measures of wealth relative to consumption, the variation is no where near the level depicted in figure 9. For example, the standard deviation of the logarithm of measured wealth to consumption is 0.24.

This tension is a standard feature of this type of model. Some additional source of variation due to discount factors appears to be necessary to better fit the observed volatility of aggregate wealth and security prices. In the next subsection we add time varying volatility to consumption which provides one potential source of the required variation.

### 5.3 Continuous time

So far we have seen how predictability in consumption is related to movements in the wealth consumption ratio. The intertemporal substitution parameter is an important
ingredient in this relation. In order to study the impact of time varying volatility and permit the risk aversion parameter $\gamma$ to play a more central role in this time series variation, we consider an extension in which consumption displays stochastic volatility. This volatility gives a source of time-variation in risk premia. To capture this we introduce square root process as a model of volatility and shift our analysis to continuous time. The continuous time formulation we now explore simplifies the analysis of volatility.

Suppose that:

\[
\begin{align*}
    dx_t &= Ax_t dt + \sqrt{z_t} dW_t \\
    dz_t &= \tilde{A}(z_t - \mu_z) dt + \sqrt{\tilde{z}_t} d\tilde{W}_t \\
    d\log C_t &= G(x_t) dt + \mu_c dt + \sqrt{z_t} \sigma_{c,t} dW_t + \sqrt{\tilde{z}_t} \tilde{\sigma}_{c,t} d\tilde{W}_t
\end{align*}
\]  

(23)

where the matrix $A$ has eigenvalues with real parts that are strictly negative. The process $z$ is scalar and the coefficient $\tilde{A}$ is negative. The processes $W$ and $\tilde{W}$ are mutually independent standard Brownian motions. The process $W$ can be multivariate and the process $\tilde{W}$ is scalar. The volatility process $\{z_t\}$ follows a Feller square root process and $\tilde{A}\mu_z + \frac{1}{2} \tilde{B}^2 < 0$. In this specification the process $\{z_t\}$ is used to model macroeconomic volatility in an *ad hoc* but convenient manner.

### 5.3.1 Continuous time Bellman equation

Consider a stochastic evolution for the continuation value of the form:

\[
d\log V^\rho_t = \xi^\rho_{v,t} dt + \sqrt{z_t} \sigma^\rho_{v,t} dW_t + \sqrt{\tilde{z}_t} \tilde{\sigma}^\rho_{v,t} d\tilde{W}_t.
\]

For this continuous time diffusion structure, we derive an equation linking the drift $\xi^\rho_{v,t}$ with current consumption and continuation values as well as diffusion coefficients. For this Brownian motion information structure, the continuous time evolution for the continuation value, indexed by $\rho$, must satisfy:

\[
0 = \frac{\delta}{1 - \rho} \left[ \left( \frac{C_t}{V^\rho_t} \right)^{1-\rho} - 1 \right] + \xi^\rho_{v,t} + z_t \left( \frac{1 - \gamma}{2} \right) \left[ \sigma^\rho_{v,t} \cdot \sigma^\rho_{v,t} + (\tilde{\sigma}^\rho_{v,t})^2 \right]
\]

Heuristically this can be obtained by taking limits of the discrete time recursion (14) as the sample horizon shrinks to zero. The rigorous formulation of recursive preferences in continuous time is given by Duffie and Epstein (1992b).

Thus

\[
\xi^\rho_{v,t} = \frac{-\delta}{1 - \rho} \left[ \left( \frac{C_t}{V^\rho_t} \right)^{1-\rho} - 1 \right] + z_t \left( \frac{\gamma - 1}{2} \right) \left[ \sigma^\rho_{v,t} \cdot \sigma^\rho_{v,t} + (\tilde{\sigma}^\rho_{v,t})^2 \right].
\]

In the special case in which $\rho = 1$, the drift is given by:

\[
\xi^1_{v,t} = \delta (v^1_t - \log C_t) + z_t \left( \frac{\gamma - 1}{2} \right) \left[ \sigma^1_{v,t} \cdot \sigma^1_{v,t} + (\tilde{\sigma}^1_{v,t})^2 \right].
\]  

(24)
When $\gamma = 1$, the volatility adjustment for the continuation value vanishes and this recursion coincides with the continuation value for preferences with a logarithmic instantaneous utility function. When $\gamma$ is different from one there is an adjustment for the volatility of the continuation value. In particular, when $\gamma$ exceeds one, there is a penalization for big volatility. Typically we are interested in large values of $\gamma$ to explain the cross section of asset returns.

In what follows we derive the corresponding asset pricing results for a particular endowment economy specified above.\(^8\)

### 5.3.2 Value function when $\rho = 1$

Guess a continuation value of the form:

$$v^1_t = U_v x + \bar{U}_v z_t + c_t + \mu_v.$$

where $v^1_t = \log V^1_t$ as in the discrete-time solution. Thus

$$U_v'Ax + G'x + \bar{U}_v\bar{A}z - \bar{U}_v\bar{A}\mu_c + \mu_c = \delta U_v'x + \delta \bar{U}_v z + \delta \mu_v + z_t \left( \frac{\gamma - 1}{2} \right) \left[ |U_v'B + H'|^2 + (\bar{U}_v\bar{B} + \bar{H})^2 \right].$$

Equating coefficients on $x$ gives:

$$U_v'A + G' = \delta U_v'$$

or

$$U_v = (\delta I - A')^{-1}G.$$ 

This formula for $U_v$ is the continuous time analog of our previously derived discrete time formula given in (21).

Equating coefficients in $z_t$ gives the following equation

$$\bar{U}_v\bar{A} = \delta \bar{U}_v + \frac{\gamma - 1}{2} \left[ (\bar{U}_v\bar{B} + \bar{H})^2 + |U_v'B + H'|^2 \right]$$

in the unknown coefficient $\bar{U}_v$. This equation can be solved using the quadratic formula, provided that a solution exits. Typically there are two solutions to this equation exist, and we select the one that is largest. When $\gamma = 1$, $\bar{U}_v = 0$. Large $\bar{B}$ and large values of $\gamma$ can result in the absence of a solution. On the other hand, shrinking $\bar{B}$ to zero will cause $z_t$ to be very smooth and ensure a solution. The limit can be thought of as giving us the continuous time counterpart to the discrete-time model specified previously in subsection (5.1.4).

Consider the special case in which $\bar{H}$ is zero, and suppose that $\gamma$ exceeds one. Thus there is no immediate impact of the shock $d\bar{W}_t$ on the growth rate of consumption.

---

\(^8\)Asset pricing applications of these preferences are developed by Duffie and Epstein (1992a). They incorporate these preferences into a standard representative agent economy with exogenous asset returns and endogenous consumption in the style of Merton (1973) and Breeden (1979).
When solutions exist, they will necessarily be negative because the quadratic function of $\bar{U}_v$ is always positive for all positive values of $\bar{U}_v$. Thus when volatility increases the continuation value declines. The discrete time wealth-consumption expansion (19) in $\rho$ continues to apply in this continuous time environment. Thus when volatility increases the wealth-consumption ratio will increase as well provided that $\rho$ exceeds one, at least for values of $\rho$ local to unity. Conversely, the ratio declines when $\rho$ is less than one.

Finally, the constant term satisfies:

$$\mu_c - \bar{U}_z \bar{A}_z = \delta \mu_v$$

which determines $\mu_v$.

For future reference, the local shock exposure of $dv^1_t$ is:

$$\sqrt{z_t} (B'U_v + H)' dW_t + \sqrt{z_t} (\bar{B}\bar{U}_v + \bar{H}) d\bar{W}_t.$$

Thus $\sigma^1_{v,t} = (B'U_v + H)'$ and $\bar{\sigma}^1_{v,t} = (\bar{U}_v\bar{B} + \bar{H})$.

### 5.3.3 Derivative with respect to $\rho$

Next we derive the formula for the derivative of the continuation value with respect to $\rho$ evaluated at one. Our aim is to produce formula of the form:

$$v_\rho^t \approx v^1_t + (\rho - 1) Dv_t.$$

The derivative $\{Dv_t\}$ evolves as an Ito process:

$$dDv = D\xi_{v,t} dt + \sqrt{z_t} D\sigma_t dW_t + \sqrt{z_t} D\bar{\sigma}_t d\bar{W}_t,$$

where $D\xi_{v,t}$ is drift coefficient and $D\sigma_t$ and $D\bar{\sigma}_t$ are the coefficients that govern the shock exposures. We obtain these coefficients by differentiating the corresponding coefficients for the continuation value process with respect to $\rho$. For instance,

$$D\xi_{v,t} = \frac{d\xi_{v,t}^\rho}{d\rho} \bigg|_{\rho=1}.$$

Recall the formula for the drift:

$$\xi_{v,t}^\rho = \frac{-\delta}{1 - \rho} \left[ \left( \frac{C_t}{V_t^\rho} \right)^{1-\rho} - 1 \right] + z_t \left( \frac{\gamma - 1}{2} \right) (\sigma^\rho_{v,t} \cdot \sigma^\rho_{v,t} + \bar{\sigma}^\rho_{v,t} \cdot \bar{\sigma}^\rho_{v,t}).$$

Differentiating with respect to $\rho$ gives:

$$D\xi_{v,t} = \delta \frac{(C_t - v^1_t)^2}{2} + \delta Dv_t + z_t (\gamma - 1) \left( D\sigma_{v,t} \cdot \sigma^1_{v,t} + D\bar{\sigma}_{v,t} \cdot \bar{\sigma}^1_{v,t} \right)$$

(25)
To compute this derivative, as in discrete time it is convenient to use a distorted probability measure. Thus we use

\[ dW_t = \sqrt{z_t(1 - \gamma)}\sigma_{v,t}dW_t^* \]
\[ d\tilde{W}_t = \sqrt{z_t(1 - \gamma)}\tilde{\sigma}_{v,t}d\tilde{W}_t^* \]

where \( \{(W_t^*, \tilde{W}_t^*): t \geq 0\} \) is a multivariate Brownian motion. As a consequence, the distorted evolution is:

\[ dx_t = Ax_tdt + (1 - \gamma)B(B'U_v + H)z_tdt + \sqrt{z_tBdW_t^*} \]
\[ dz_t = \tilde{A}(z_t - \mu_z)dt + (1 - \gamma)\tilde{B}(\tilde{B}U_v + \tilde{H})z_tdt + \sqrt{z_t\tilde{B}d\tilde{W}_t^*} \]
\[ d\log C_t = G'x_tdt + \mu_vdt + (1 - \gamma)H'(B'U_v + H)z_tdt + \sqrt{z_tH'dW_t^* + \sqrt{z_tHd\tilde{W}_t^*}} \]

(26)

Let \( \tilde{\xi}_{v,t} \) denote the resulting distorted drift for the derivative. Then rewrite equation (25) as:

\[ \tilde{\xi}_{v,t} = \frac{\delta(c_t - v_t1)^2}{2} + \delta Dv_t^1 \]

which can be solved forward as:

\[ Dv_t^1 = -\delta \int_0^\infty \exp(-\delta u)E^*[(c_{t+u} - v_{t+u1})^2|x_t, z_t]du. \]

\( Dv_t^1 \) is a linear/quadratic function of the composite Markov state \((x, z)\). See appendix A.2.

### 5.3.4 Stochastic discount factor

Let \( s_t^\rho \) be the logarithm of the continuous time stochastic discount factor for parameter \( \rho \). This stochastic discount factor process encodes discounting for all horizons from the vantage point of time zero. Specifically \( \exp(s_t^\rho) \) is discount factor over horizon \( t \) and \( \exp(s_{t+\tau}^\rho - s_t^\rho) \) is the discount factor for horizon \( t \) from the vantage point of date \( \tau \). Then

\[ ds_t^\rho = -\delta dt - \rho dc_t + (\rho - \gamma) \left[ dv_t^\rho - \xi_t^\rho dt - z_t \left( \frac{\rho - \gamma}{2} \right) \left( \sigma_{v,t}^0 \cdot \sigma_{v,t}^0 + \tilde{\sigma}_{v,t}^0 \cdot \tilde{\sigma}_{v,t}^0 \right) dt \right] \]
\[ = -\delta dt - \rho dc_t \]
\[ + (\rho - \gamma) \left[ \sqrt{z_t}\sigma_{v,t}^0dW_t + \sqrt{z_t}\tilde{\sigma}_{v,t}^0d\tilde{W}_t - z_t \left( \frac{\rho - \gamma}{2} \right) \left( \sigma_{v,t}^0 \cdot \sigma_{v,t}^0 + \tilde{\sigma}_{v,t}^0 \cdot \tilde{\sigma}_{v,t}^0 \right) dt \right]. \]

Differentiating we find that the \( \rho \) derivative process \( \{D_s_t: t \geq 0\} \) evolves as

\[ dD_{s_t} = -dc_t + \left[ \sqrt{z_t}\sigma_{v,t}^1dW_t + \sqrt{z_t}\tilde{\sigma}_{v,t}^1d\tilde{W}_t - z_t (1 - \gamma) \left( \sigma_{v,t}^1 \cdot \sigma_{v,t}^0 + \tilde{\sigma}_{v,t}^1 \cdot \tilde{\sigma}_{v,t}^0 \right) dt \right] \]
\[ + (1 - \gamma) \left[ \sqrt{z_t}D\sigma_{v,t}dW_t + \sqrt{z_t}\tilde{D}\sigma_{v,t}d\tilde{W}_t - z_t (1 - \gamma) \left( D\sigma_{v,t} \cdot \sigma_{v,t}^0 + D\tilde{\sigma}_{v,t} \cdot \tilde{\sigma}_{v,t}^0 \right) dt \right]. \]

Thus the \( \rho \) approximation is

\[ s_t^\rho \approx s_t^1 + (\rho - 1)Ds_t \]

with the following contributions to the stochastic evolution of the approximation:
5.3.5 Risk Prices

Of particular interest is the recursive utility adjustment to the Brownian motion risk prices. The $\rho$ approximations are given by the negatives of the values reported in (b) and (c):

i) $\sqrt{z_t}(\rho H' + \sqrt{z_t}(\gamma - \rho)\sigma_{v,t}^1) + \sqrt{z_t}(\rho - 1)(\gamma - 1)D\sigma_{v,t}$ - risk prices for exposure to $dW_t$ risk;

ii) $\sqrt{z_t}(\bar{\rho} H' + \sqrt{z_t}(\gamma - \rho)\bar{\sigma}_{v,t}^1) + \sqrt{z_t}(\rho - 1)(\gamma - 1)D\bar{\sigma}_{v,t}$ - risk prices for exposure to $d\bar{W}_t$ risk.

These prices are quoted in terms of required mean compensation for the corresponding risk exposure. The first vector is the mean compensation for exposure to $dW_t$ and the second vector is the mean compensation for exposure to $d\bar{W}_t$.

The risk premia earned by an asset thus consist of a covariance with consumption innovations (multiplied by the intertemporal substitution parameter) and components representing covariance with innovations in the continuation value (weighted by a combination of intertemporal substitution and risk aversion parameters). This characterization is closely related to the two-factor model derived by Duffie and Epstein (1992a), where the second risk term is the covariance with the total market portfolio.

Consider the special case in which $\bar{H}$ is zero. Then under the Breeden model, the volatility shock $d\bar{W}_t$ has zero price. Under the forward-looking recursive utility model, this shock is priced. For instance, for large $\gamma$ and $\rho$ close to one, the contribution is approximately $\sqrt{z_t}(\gamma - 1)\bar{B}\bar{U}_v$. The recursive utility also amplifies the risk prices for $dW_t$ risk exposure. For large $\gamma$ and $\rho$ close to one the prices are approximately $\sqrt{z_t}(\gamma - 1)(H' + U'_v\bar{B})$, which is the continuous time counterpart to the discounted impulse response function for consumption growth rates. When the importance of volatility becomes arbitrarily small ($\bar{B}$ declines to zero), the volatility state ceases to vary and collapses to $\mu_z$. The predictability in consumption continues to amplify risk prices but the prices cease to vary over time.

Again we consider two specifications. The first is a continuous time version of Bansal and Yaron (2004). In contrast with our discrete time example, but consistent with Bansal and Yaron (2004), we introduce stochastic volatility.

\[
\begin{align*}
dc_t &= .0015dt + x_tdt + \sqrt{z_t}.0078dW_{1,t} \\
dx_t &= -.021x_tdt + \sqrt{z_t}.00034dW_{2,t} \\
dz_t &= -.013(z_t - 1)dt + \sqrt{z_t}.038d\bar{W}_t
\end{align*}
\]
By construction the volatility process \( \{ z_t \} \) has a unit mean.

In the Bansal and Yaron (2004) model, risk premia fluctuate. We use a Feller square root process for conditional variances while Bansal and Yaron (2004) used first-order autoregression with normal errors. In our specification, the stationary distribution for conditional variances is in the gamma family and for their specification the distribution is in the normal family. We report the two densities in figure 14. Our square root specification is by design analytically tractable and it formally restricts variances to be positive.\(^9\) Thus it is more convenient for our purposes to work with a square root process. The two densities are quite similar, and both presume that there are considerable long run fluctuations in volatility.

While we expect \( \gamma \) to have direct impact on risk prices, it is useful to quantify the role of \( \rho \) because changing intertemporal substitution parameter will alter risk prices. To quantify this effect, consider the first order combined expansion in \( \rho \) and \( \gamma \) around the values \( \rho = 1 \) and \( \gamma = 1 \) is:\(^{10}\)

\[
\sqrt{z_t} \left[ H - (\rho - 1)B'U_v + (\gamma - 1)(B'U_v + H) \right] = \sqrt{z_t} \left( \begin{bmatrix} 2.70 \\ 0 \end{bmatrix} - (\rho - 1) \begin{bmatrix} 0 \\ 5.12 \end{bmatrix} + (\gamma - 1) \begin{bmatrix} 2.70 \\ 5.12 \end{bmatrix} \right). 
\]

While Bansal and Yaron (2004) use monthly time units, we have rescaled the time units to annual and we have further multiplied prices by one hundred so that the value units are in expected rates of return expressed as percentages.

In contrasting the contributions of \( \rho \) and \( \gamma \), note that while increases in \( \gamma \) amplify both risk prices, increases in \( \rho \) reduce the risk price for the shock to the growth rate in consumption. It is the recursive utility adjustment induced by persistence in the growth rate to consumption that makes the risk price of exposure to \( dW_t^2 \) different from zero. In this Bansal and Yaron (2004) specification, the risk price of \( dW_t^2 \) exposure is double that of \( dW_t^1 \). As we will see, the recursive utility contribution is much more challenging to measure reliably.

For pedagogical convenience, we have featured the first-order term in \( \gamma \), in fact this is not critical. The higher-order term allows us to explore non-local changes in the parameter \( \gamma \). For instance, as we change \( \gamma \) to be five and then ten, the first-order expansions in \( \rho \) evaluated at \( x_t = 0 \) and \( z_t = 1 \) are:

\[
\gamma = 5: \quad \sqrt{z_t} \left( \begin{bmatrix} 13.5 \\ 20.5 \end{bmatrix} - (\rho - 1) \begin{bmatrix} 0 \\ 5.9 \end{bmatrix} \right)
\]

\[
\gamma = 10: \quad \sqrt{z_t} \left( \begin{bmatrix} 27.0 \\ 46.1 \end{bmatrix} - (\rho - 1) \begin{bmatrix} 0 \\ 5.3 \end{bmatrix} \right)
\]

\(^9\)Negative variances are very unlikely for the parameter values used by Bansal and Yaron (2004). Moreover, in the unlikely event that zero is reached in a continuous time version of their model, one could impose a reflecting barrier.

\(^{10}\)This expansion illustrates a point made by Garcia et al. (2006) that when \( \rho \) is small, \( \gamma \) underestimates the contribution of risk aversion and conversely when \( \rho \) is large.
Figure 14: --- depicts the stationary density of $z$: gamma(18.0, 0.056). -. depicts the normal density with the same mean 1 and the same standard deviation .236 for comparison.
The $\rho$ derivatives change as we alter $\gamma$, but not dramatically so.

Consider next the price of exposure to volatility risk. For model (28), $\bar{H} = 0$ and the magnitude of $\bar{U}_v$ depends explicitly on the choice of $\gamma$. In the local to unity expansion of $\gamma$ and $\rho$, level term and the coefficients on both $\rho - 1$ and $\gamma - 1$ are zero suggesting that volatility risk premia are relatively small. When we increase $\gamma$ we obtain the following first-order expansions in $\rho$ evaluated at $z_t = 1$ and $x_t = 0$:

$\gamma = 5 : \sqrt{z_t}[-2.0 + (\rho - 1) 0.7]$

$\gamma = 10 : \sqrt{z_t}[-10.3 + (\rho - 1) 1.1]$

The level terms in the risk prices are negative for the volatility shock. While increases in consumption are valued increases in consumption volatility is not. There is apparently substantial nonlinearity in how these level terms increase in $\gamma$. Doubling $\gamma$ from five to ten leads to a five fold increase in the magnitude of the volatility risk price.

Consider next the continuous time counterpart to our second specification. In this specification there is no stochastic volatility. The first order expansion in $\rho$ and $\gamma$ around the values $\rho = 1$ and $\gamma = 1$ is:

$$
[H - (\rho - 1)B'U_v + (\gamma - 1)(B'U_v + H)]
= \left(\begin{array}{c}
.96 \\
0
\end{array}\right) - (\rho - 1) \left(\begin{array}{c}
.79 \\
1.01
\end{array}\right) + (\gamma - 1) \left(\begin{array}{c}
1.75 \\
1.01
\end{array}\right).
$$

Again the coefficient on $\rho - 1$ is negative while the coefficient on $\gamma - 1$ is positive increasing $\rho$ diminishes the risk prices. The magnitude of the $\rho$ derivative for pricing the shock to corporate earnings is larger than the shock to consumption, but the reverse is true for the consumption shock. As we change $\gamma$ to five and then 10, we find that

$\gamma = 5 : \begin{bmatrix} 7.95 \\ 4.04 \end{bmatrix} - (\rho - 1) \begin{bmatrix} 1.08 \\ 1.63 \end{bmatrix}$

$\gamma = 10 : \begin{bmatrix} 16.69 \\ 9.09 \end{bmatrix} - (\rho - 1) \begin{bmatrix} 1.43 \\ 2.36 \end{bmatrix}$

so the $\rho$ derivatives get larger in magnitude for larger values of $\gamma$.

Overall the risk prices are smaller for the second specification than for the first one. Bansal and Yaron (2004) intended to match data going back to 1929 including the pre-war period whereas Hansen et al. (2005) used estimates obtained with post-war data. There is much less consumption volatility in this latter sample.

5.3.6 Riskfree rate

Consider next the instantaneous risk-free rate. For an arbitrary $\rho$, this is given by limit:

$$
r^\rho_{f,t} = \lim_{\epsilon \downarrow 0} -\log E [\exp(s_{t+\epsilon} - s_t) | \mathcal{F}_t] = \delta + \rho G'x_t + \rho \mu_v - \frac{\rho^2 z_t}{2} \left(H'H + \bar{H}^2\right)
$$

51
\begin{equation}
+z_t \rho (\rho - \gamma) (H' \cdot \sigma_{1,t}^2 + H^* \hat{\sigma}_{1,t}^2) .
\end{equation}

The last term on the right-hand side gives the contribution for recursive utility and depends in part on the discrepancy between \( \rho \) and \( \gamma \).

In particular, when \( \rho = 1 \)
\[
\begin{align*}
r_{f,t}^1 &= \delta + G'x_t + \mu_c - \frac{z_t}{2} (H'H + \bar{H}^2) + z_t (1 - \gamma) (H' \cdot \sigma_{1,t}^1 + \bar{H}\sigma_{1,t}^1).
\end{align*}
\]

The \( \rho \) derivative of the risk-free rate is:
\[
Dr_{f,t} = G'x_t + \mu_c + z_t \left[ -H' + (2 - \gamma) \bar{\sigma}_{1,t}^1 + (1 - \gamma) D\sigma_{v,t} \right] \cdot H' \]
\[+ z_t \left[ -\bar{H} + (2 - \gamma) \bar{\sigma}_{1,t}^1 + (1 - \gamma) D\bar{\sigma}_{v,t} \right] \bar{H}.
\]

The approximation is:
\[
r_{f,t}^0 = r_{f,t}^1 + (\rho - 1) Dr_{f,t}.
\]

While this expression is a bit tedious, it is informative to contrast the local to unity contributions of \( \rho \) to those of \( \gamma \). At \( \gamma = 1 \), \( \bar{\sigma}_{v,t} = 0 \) and thus the local approximation is:
\[
\begin{align*}
\delta &+ G'x_t + \mu_c - \frac{z_t}{2} (H'H + \bar{H}^2) + \\
+ (\rho - 1) \left[ G'x_t + \mu_c - z_t (H'H + \bar{H}^2) + z_t H' \cdot \sigma_{1,t}^1 + z_t \bar{H}\sigma_{1,t}^1 \right] \\
+ (\gamma - 1) z_t \left( -H' \cdot \sigma_{1,t}^1 - \bar{H}\sigma_{1,t}^1 \right).
\end{align*}
\]

Importantly, the term multiplying \((\gamma - 1)\) does not include \( G'x_t + \mu_c - z_t (H'H + \bar{H}^2) \). In particular, the conditional mean in the growth rate of consumption, as reflected in \( \mu_c + G'x_t \) contributes only to the \( \rho \) derivative. Increases in \( \rho \) will unambiguously increase \( \rho \mu_c \) making the interest rate larger. This can be offset to some extent by shrinking \( \delta \) but only up to where \( \delta = 0 \). This tension is a version of Weil (1989)'s risk free rate puzzle.

The term:
\[
(\rho - \gamma) z_t (H' \cdot \sigma_{1,t}^1 + \bar{H}\hat{\sigma}_{1,t}^1)
\]
has the interpretation of changing probability measures by adding drift \((\rho - \gamma) z_t \sigma_{1,t}^1\) and \((\rho - \gamma) z_t \hat{\sigma}_{1,t}^1\) to the respective Brownian motions \( dW_t \) and \( d\bar{W}_t \). Changing \( \rho \) or \( \gamma \) will, of course, alter this term, but
\[
z_t (H' \cdot \sigma_{1,t}^1 + \bar{H}\hat{\sigma}_{1,t}^1)
\]
is typically smaller than the mean growth rate of consumption.\footnote{This term is 0.07 (in annualized percent) in the Bansal and Yaron (2004) model, which is small relative to the 1.8 percent growth rate in consumption when evaluated at \( z = 1 \). In the Hansen et al. (2005) model this term is 0.02 percent which is small relative to a per capita consumption growth rate of 2.9 percent. The remaining term from consumption volatility \( z_t (H'H + \bar{H}^2) \) at \( z = 1 \) is also small, 0.07 in the Bansal and Yaron (2004) model and 0.01 in the Hansen et al. (2005) model.} More generally, these risk-free rate approximations give a formal sense in which changes in \( \gamma \) have a much more modest impact on the instantaneous interest rate than changes in \( \rho \) and allows us to consider a wide range of values of \( \gamma \).
5.3.7 Cash flow returns

As we have seen, the local evolution of the stochastic discount factor implies a vector of local risk prices. Next we explore cash-flow counterparts, including a limiting notion of an expected rate of return that compensates for exposure to cash flow risk.

Consider a cash flow that can be represented as

\[ D_t = G_t f(X_t)D_0 \]

where \( G_t \) is a stochastic growth process initialized to be one at date zero, \( D_0 \) is an initial condition and \( f(X_t) \) is a transient component and the process \( X \) evolves as a Markov process. For instance, the Markov process \( X \) could consist of \((x, z)\) with evolution equation (23).

Define the expected rate of return to a cash flow as:

\[
\frac{1}{t} \log E \left[ G_t f(X_t) | \mathcal{F}_0 \right] - \frac{1}{t} \log E \left[ S_t G_t f(X_t) | \mathcal{F}_0 \right].
\]

Let the gross return to a holding a cash flow over a unit horizon be:

\[
\log E \left( S_{t-1} G_t f(X_t) | \mathcal{F}_1 \right) - \log E \left( S_t G_t f(X_t) | \mathcal{F}_0 \right)
\]

An equity is a portfolio of claims to such returns. Both of these returns typically have well defined limits as \( t \to \infty \) and these limits will remain invariant over a class of functions \( f \) used to define transient components to cash flows. As emphasized by Hansen et al. (1995) and Lettau and Wachter (2006), the intertemporal composition is of interest.

As featured by Hansen et al. (2005), Hansen (2006), we can construct long run counterpart to risk prices by considering the long run excess returns for alternative \( G \) specified by martingales that feature the components of cash flow risk. To be concrete, suppose that:

\[
d \log G_t = -\frac{1}{2} (K' \bar{K} + \bar{K}' K) z_t + \sqrt{z_t} K' dW_t + \sqrt{z_t} \bar{K} d\bar{W}_t.
\]

This specification allows us to focus on the growth rate risk exposure as parameterized by \( K \) and \( \bar{K} \). For instance, \( K \) and \( \bar{K} \) can be vectors of zeros except on one entry in which there is a nonzero entry used to feature this specific risk exposure.

Then the logarithm of the limiting cash flow return is:

\[
\lim_{t \to \infty} \left( \frac{1}{t} \log E [G_t f(X_t) | \mathcal{F}_0] - \frac{1}{t} \log E [S_t G_t f(X_t) | \mathcal{F}_0] \right) = \eta - \nu.
\]

The derivative of \( \eta - \nu \) with respect to \( K \) and \( \bar{K} \) gives the long run cash flow counterpart to a local risk price. Using the method of Hansen and Scheikman (2006), the family of functions \( f \) for which these limits remain invariant can be formally characterized. For such functions \( f \), the cash flow contribution \( f(X_t) \) can be viewed as transient from the vantage point of long run risk prices.
Following Hansen et al. (2005), Hansen and Scheikman (2006) and Hansen (2006), we characterize these limits by solving so called *principal eigenfunction problems*:

\[
\lim_{t \to 0} E \left[ G_t e(X_t) \big| X_0 = X \right] = \eta \tilde{e}(X)
\]

\[
\lim_{t \to 0} E \left[ S_t G_t \hat{e}(X_t) \big| X_0 = X \right] = \nu \hat{e}(X)
\]

Finally the logarithm of the limiting holding period return is:

\[
\lim_{t \to \infty} \left[ \log E \left( S_t-1 G_t f(X_t) \big| F_t \right) - \log E \left( S_t G_t f(X_t) \big| F_0 \right) \right] = -\nu + \log \hat{e}(X_1) - \log \hat{e}(X_0) + \log G_1.
\]

This latter return has three components, a) an eigenvalue component, b) a pure cash flow component and c) an eigenfunction component. The choice of the transient component \( f(X_t) \), typically does not contribute to the value. The valuation implicit in the stochastic discount factor is reflected in both \(-\nu \) and \( \log \hat{e}(X_1) - \log \hat{e}(X_0) \), but of course not in the cash flow component \( \log G_1 \). In contrast to the log-linear statistical decompositions of Campbell and Shiller (1988a), the decompositions we just described require an explicit valuation model reflected in a specification of the stochastic discount factor.

Consider first the Bansal and Yaron (2004) model. The risk prices computed as derivative of long-run return with respect to \( K \) depends on the values of \( K \). As the baseline values of \( K \), we use the risk exposure of the consumption and the state variable. At these baseline values, we obtain the following long run risk prices for \( \rho = 1 \) as we increase \( \gamma \):\(^{12}\)

\[
\begin{bmatrix}
 2.70 \\
 5.62 \\
\end{bmatrix}
\begin{bmatrix}
 13.87 \\
 26.85 \\
\end{bmatrix}
\begin{bmatrix}
 30.30 \\
 58.33 \\
\end{bmatrix}
\]

\[ \gamma = 1 \hspace{1cm} \gamma = 5 \hspace{1cm} \gamma = 10 \]

where \( \beta = .998 \) is assumed as in Bansal and Yaron (2004). The prices are close to linear in \( \gamma \) but there is nonlinear contribution caused by stochastic volatility, which makes the risk prices more than proportional to \( \gamma \). Although the second shock has no immediate impact on consumption and hence a zero local risk price, it has long lasting impact on the stochastic discount factor by altering the predicted growth rate in consumption. As expected in figure 4, it turns out that the long run risk price for this shock is bigger than that for consumption shock.

Consider next the Hansen et al. (2005) model. For this model, the risk prices computed as derivatives of long run return with respect \( K \) of are insensitive to the baseline choice of \( K \). In other words the component prices are constant as shown by Hansen et al.

\[^{12}\text{The prices are slightly decreasing in } K. \text{ At 10 times baseline values of } K, \text{ they are}
\[
\begin{bmatrix}
 2.69 \\
 5.61 \\
\end{bmatrix}
\begin{bmatrix}
 13.54 \\
 26.80 \\
\end{bmatrix}
\begin{bmatrix}
 28.66 \\
 58.04 \\
\end{bmatrix}
\]

\[ \gamma = 1 \hspace{1cm} \gamma = 5 \hspace{1cm} \gamma = 10 \]

54
(2005). For this model we report the long run prices for $\rho = 1$ for three different values of $\gamma$:

\[
\begin{bmatrix}
1.77 \\
1.06 \\
\end{bmatrix}
\begin{bmatrix}
8.76 \\
5.10 \\
\end{bmatrix}
\begin{bmatrix}
17.50 \\
10.15 \\
\end{bmatrix}
\]

$\gamma = 1$ $\gamma = 5$ $\gamma = 10$

The prices are linear and approximately proportional to $\gamma$ and are computed assuming that $\beta = 0.99^{1/4}$ as in Hansen et al. (2005). Even when $\gamma$ and $\rho$ are one, the long run cash flow risk price is positive for the shock to corporate earnings. While the corporate earnings shock is normalized to no immediate impact on consumption, it will have a long run impact and hence this will show up in the equilibrium risk prices.

We report the derivatives of long-run risk price with respect to $\rho$ for both specifications in figure 15. Recall that these derivative were negative for the local prices. As is evident from this figure, for the Bansal and Yaron (2004) model the derivative is positive for low and high values of $\gamma$ for the shock to growth rate in consumption. The derivative is negative for a range of intermediate values.

This differences between the derivatives for long run and local prices are due to the predictability of consumption. With the predictability of consumption, the permanent response of consumption and hence, the permanent response of stochastic discount factor to a shock are more than their contemporary responses. This additional contribution makes the long run risk price and its derivative with respect to $\rho$ larger than their local counterparts. Figure 16 shows this point: long run considerations shift up risk prices and the corresponding $\rho$ derivative.\(^{13}\)

\(^{13}\)Because of stochastic volatility, long run considerations tilt the risk price and its derivative along with shifting them.
Figure 15: The top panel is Bansal-Yaron model: — depicts $\rho$ derivative of long run risk price of exposure to consumption shock. It is calculated by dividing the difference between $\rho$ derivatives of long-run return at $K = [0 \ 0]'$ and $K = [0.0078 \ 0]'$ (risk exposure of $c_t$) by .0078. It is approximation to the cross derivative of long run return with respect to $K$ and $\rho$, that is, $\rho$ derivative of long run risk price. The –. curve depicts $\rho$ derivative of long run risk price of exposure to predicted consumption shock. It is calculated by dividing the difference between $\rho$ derivatives of long-run return at $K = [0 \ 0]'$ and $K = [0 \ 0.00034]'$ (risk exposure of $x_t$) by .000034. The bottom panel is Hansen-Heaton-Li model: — depicts $\rho$ derivative of long run risk price of exposure to consumption shock and –. depicts $\rho$ derivative of the long run risk price of the exposure to corporate earnings. For this model the risk prices, the derivatives with respect to the individual entries of $K$, are constant.
Figure 16: Risk price (top panel) and its derivative (bottom panel) with respect to $\rho$ for the shock to growth rate in consumption in Bansal-Yaron model: — depicts long run risk price and $\rho$ derivative; —. depicts local counterparts. Both levels and derivatives are evaluated at $\rho = 1$. 

57
6 Information about Risk Aversion from Multiple Returns

In the previous section we examined how risk aversion and intertemporal substitution affect predicted risk premia. We now examine predictions for risk aversion using information from the returns to the test assets described in section 4.2. Because of the substantial differences in average returns we will be driven to large levels of risk aversion. For these parameter values, variation in $\rho$ around one has little effect. For this reason and for tractability we assume that $\rho = 1$. For similar reasons Campbell (1996) also considers the case where $\rho$ is close to one and shows that a cross-section of returns can be used to identify $\gamma$.

Returns to our test portfolios are known to have differential predictive power for consumption as shown in the work of Parker and Julliard (2005). To the cointegrated model of consumption and corporate earnings of Hansen et al. (2005) we add the log price dividend ratio and the log dividend growth for each of the five portfolios. To avoid substantial parameter proliferation we estimate each system portfolio by portfolio.

Returning to the discrete time, log-linear setting of section 5.1, the excess return to an asset is determined by the covariance between shocks to the return and shocks to current and future consumption. As in section 4 the return to security $j$ has a moving-average representation given by:

$$r_{t+1}^j = \rho_j^j(L)w_{t+1} + \mu_{t+1}^j.$$ 

Hence the on impact effect of the shock vector $w_{t+1}$ on return $j$ is given by the vector $\rho(0)$.

Under recursive utility risk premia are determined by the exposure of both consumption and the continuation value to shocks. When the intertemporal elasticity of substitution is assumed to be one, shocks to the log continuation value are given by the discounted impulse responses of log consumption to the shocks. These discounted responses are given by the vector:

$$\Theta(\beta) \equiv H + \beta B'(I - \beta A)^{-1}G .$$

Hence we can write the risk premium for security $j$ as:

$$E(r_{t+1}^j|\mathcal{F}_t) - r_{t+1}^f = -\frac{|\rho_j^j(0)|^2}{2} + [H + (\gamma - 1)\Theta(\beta)] \cdot \rho_j^j(0) .$$

(31)

Risk aversion can have a large impact on risk premia if consumption is predictable so that $\Theta(\beta)$ is significant and if innovations to discounted future consumption covary with shocks to returns. This covariance is captured by the term $\Theta(\beta) \cdot \rho(0)$.

As an initial proxy for this covariance we calculate the covariance between returns at time $t + 1$ and $c(t + \tau) - c(t)$ conditional on being at the mean of the state variable and for different values of $\tau$. This calculation ignores discounting through $\beta$ and truncates the effects at a finite horizon. The results of this calculation are reported in figure 17.
for each of the five book-to-market portfolios. The calculation is done using the point estimates from the VAR for each portfolio.

For small values of $\tau$ there is relatively little heterogeneity in the conditional covariance between consumption and portfolio returns. The risk exposure in consumption over the short-term is not a plausible explanation for differences in observed average returns as reported in table 1. Notice, however, that as $\tau$ increases there are pronounced differences in the covariances. For example the covariance between long-run consumption and returns is much higher for portfolio 5 than it is for portfolio 1. Further when $\tau = 40$ the estimated covariances follow the order of the observed average returns. Portfolio 1 has the lowest average return and lowest covariance with consumption. Portfolio 5 has the highest average return and highest covariance.

Figure 18 displays the estimated value of $\Theta(\beta) \cdot \rho_j(0)$ for each security and alternative values of $\beta$. As in figure 17 there are substantial differences in the estimated level of risk exposure across the portfolios as $\beta$ approaches 1.

An implied level of the risk aversion parameter $\gamma$ can be constructed using the esti-
Figure 18: Conditional Covariance Between Returns and $\Theta(\beta)w_{t+1}$
Table 4: Estimates of $\gamma$ for different values of $\beta$, based on (32)

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90</td>
<td>318.1</td>
</tr>
<tr>
<td>0.91</td>
<td>252.0</td>
</tr>
<tr>
<td>0.92</td>
<td>199.4</td>
</tr>
<tr>
<td>0.93</td>
<td>157.0</td>
</tr>
<tr>
<td>0.94</td>
<td>122.7</td>
</tr>
<tr>
<td>0.95</td>
<td>94.9</td>
</tr>
<tr>
<td>0.96</td>
<td>72.2</td>
</tr>
<tr>
<td>0.97</td>
<td>53.6</td>
</tr>
<tr>
<td>0.98</td>
<td>38.5</td>
</tr>
<tr>
<td>0.99</td>
<td>26.1</td>
</tr>
<tr>
<td>1.00</td>
<td>16.1</td>
</tr>
</tbody>
</table>

mates reported in figure 18. To do this consider the difference between (31) for $j = 5$ and $j = 1$ yields:

$$E(r^5_{t+1}|\mathcal{F}_t) - E(r^1_{t+1}|\mathcal{F}_t) = -\frac{\rho^5(0)^2}{2} + \frac{\rho^1(0)^2}{2} + [H + (\gamma - 1)\Theta(\beta)] \cdot (\rho^5(0) - \rho^1(0)).$$

Hence

$$\gamma = \frac{E(r^5_{t+1}|\mathcal{F}_t) - E(r^1_{t+1}|\mathcal{F}_t) + \frac{\rho^5(0)^2}{2} - \frac{\rho^1(0)^2}{2} - (H - \Theta(\beta)) \cdot (\rho^5(0) - \rho^1(0))}{\Theta(\beta) \cdot (\rho^5(0) - \rho^1(0))}. \quad (32)$$

Using the estimated mean returns reported in table 1 and the estimates of $\rho^j(0)$ from each VAR system we construct estimates of $\gamma$ for different values of $\beta$. These are given in table 6. When $\beta$ is small the estimated value of $\gamma$ is quite large. Notice, however that as $\beta$ approaches 1, the two returns have substantially different risk exposures which is reflected in a much smaller estimate of $\gamma$.

The estimates reported in table 6 both ignore sampling uncertainty and are based on estimation that treats each portfolio independently. We repeat the estimation of the VAR except now we consider a six variable system where the dividend growth and price-dividend ratios of portfolio 1 and 5 are included along with $c_t - c_{t-1}$ and $e_t - c_t$.

Further we use the Bayesian simulation technique outlined in appendix B to determine the posterior distribution of the parameters of the VAR systems. For each simulation we infer a value of $\gamma$ using (32).

In our first set of simulations we ignore the estimation in the mean returns. The quantiles from the posterior distribution of $\theta$ are reported in table 6 where inference about $\gamma$ is done conditional on a fixed value of $\beta$. Notice that even when $\beta$ is equal to 1 and sampling error in the means is ignored, there is substantial uncertainty in the estimates of $\gamma$. 
When $\rho = 1$ the wealth-consumption ratio is constant and innovations in consumption could be measured by innovations to wealth. Since the return on the aggregate wealth portfolio is not observable a proxy is necessary. A common procedure is to use the return to an aggregate stock index. One justification for this procedure is to assume that the missing components have returns that are proportional to the stock return as in Campbell (1996).\textsuperscript{14}

We repeat the empirical strategy above but assume that the growth rate in consumption is proportional to return on the market portfolio discussed in appendix D. Figure 19 displays the conditional covariance between the test asset returns and the implied values of $\Theta(\beta) w_{t+1}$ for different values of $\beta$. In this case we fit a VAR with 5 lags to the log market return, the log price-dividend ratio for the market along with the log dividend growth and price-dividend ratio for each portfolio.

In this case the implied ordering of risk across the portfolios is consistent with the observed average returns only when $\beta$ is large enough. When $\beta$ is small the implied values of $\gamma$ are negative. For values of $\beta$ large enough the differences in the covariances between portfolios 1 and 5 imply that portfolio 5 should have a larger return than portfolio 1. Essentially the differential in the return to portfolios 5 and 1, the “return to value” is able to forecast the market return. As in the work of Campbell and Vuolteenaho (2004) the CES model with the market return as a proxy for consumption growth implies that there should be a premium for value over growth: the “value premium.”

\begin{table}[h]
\centering
\begin{tabular}{lcccccc}
\hline
Quantile: & 0.10 & 0.25 & 0.50 & 0.75 & 0.90 \\
\hline
$\beta = 0.98$ & -134.66 & 44.47 & 76.59 & 135.94 & 279.83 \\
$\beta = 0.99$ & -58.71 & 34.53 & 57.76 & 99.48 & 194.87 \\
$\beta = 1$ & -14.41 & 20.72 & 37.37 & 63.84 & 119.84 \\
\hline
\end{tabular}
\caption{Quantiles for $\gamma$, mean returns fixed, 5 lags}
\end{table}

\textsuperscript{14}Lustig and Nieuwerburgh (2006) infer the return to non-traded human capital by using the link between consumption and unobserved wealth implied by several different assumptions about preferences.
Conditional Covariance between returns and $\Theta(\beta)w_{t+1}$

Figure 19: Covariance between shocks to portfolio returns and accumulated shocks to future consumption growth, $\Theta(\beta)w_{t+1}$ for different values of $\beta$
7 GMM estimation of stochastic discount factor models

For a given financial data set, multiple stochastic discount factors typically exist. Only when the econometrician uses a complete set of security market payoffs will there be a unique discount factor. Either an \textit{ad hoc} identification method is used to construct a discount factor, or an explicit economic model is posed that produces this random variable. Alternative economic models imply alternative measurements of a stochastic discount factor including measurements that depend on unknown parameters. Rational expectations comes into play through the use of historical time series data to test relation (2). See Hansen and Singleton (1982) and Hansen et al. (1995). Macroeconomics and finance are integrated through the use of dynamic macroeconomic equilibrium models to produce candidate discount factors.

7.1 Identification

As we have seen, pricing restrictions are typically formulated as conditional moment restrictions. For the purposes of this discussion, we rewrite equation (2):

\[ E(S_{t,t+1}a_{t+1} | F_t) = \pi_t(a_{t+1}) \]  

where \(a_{t+1}\) is the one period gross payoff to holding an asset. It is a state-contingent claim to the numeraire consumption good at date \(t + 1\). Suppose an econometrician observes a vector of asset payoffs: \(x_{t+1}\), a corresponding price vector \(q_t\) and a vector of conditioning variables \(z_t\) that are measurable with respect to \(F_t\). Moreover, the price vector must be a Borel measurable function of \(z_t\). The vector \(q_t\) might well be degenerate and consist of zeros and ones when the payoffs are returns and/or excess returns. An implication of (33) is that

\[ E(S_{t,t+1}x_{t+1} | z_t) = q_t \]  

Suppose for the moment that \(S_{t,t+1}\) is represented as a nonparametric function of a \(k\)-dimensional vector of variables \(y_{t+1}\). That is:

\[ S_{t,t+1} = f(y_{t+1}) \]

for some Borel measurable function \(f\) mapping \(\mathbb{R}^k \rightarrow \mathbb{R}\). Can \(f\) be identified? Suppose that we can construct a function \(h\) such that \(h\) satisfies:

\[ E[h(y_{t+1})x_{t+1} | z_t] = 0. \]  

Then clearly \(f\) cannot be distinguished from \(f + rh\) for any real number \(r\). Thus nonparametric identification depends on whether or not there is a nontrivial solution to (35).
Consider the following problematic examples. If \( y_{t+1} \) includes \( x_{t+1} \) and \( z_t \), then many solutions exist to (35). For any Borel measurable function \( g \), run a population regression of \( g(y_{t+1}) \) onto \( x_{t+1} \) conditioned on \( z_t \) and let \( h(y_{t+1}) \) be the regression residual:

\[
h(y_{t+1}) = g(y_{t+1}) - E[g(y_{t+1})x_{t+1}'|z_t] (E[x_{t+1}x_{t+1}'|z_t]^{-1} x_{t+1})
\]

By construction, this \( h \) satisfies (35).

Suppose that we do not impose exclusion restrictions. Instead suppose the vector \( y_{t+1} \) includes \( x_{t+1} \) and \( z_t \). Stochastic discount factors from explicit economic models are often restricted to be positive. A positive stochastic discount factor can be used to extend the pricing to include derivative claims on the primitive securities without introducing arbitrage.\(^{15}\) Our construction so far ignores this positivity restriction. As an alternative, we may impose it. Identification remains problematic in this case, there are variety ways to construct discount factors.

As shown by Hansen and Jagannathan (1991) and Hansen et al. (1995), the solution to the optimization problem:

\[
\max_{\alpha} -E [(\max \{-x_{t+1} \cdot \alpha(z_t), 0\})^2 |z_t] - 2\alpha(z_t) \cdot q_t
\]

(36)
gives a nonnegative function of \( x_{t+1} \) and \( z_t \) that solves the pricing equation where \( \alpha \) is a function of \( z_t \). From the solution \( \alpha^* \) to this concave problem, we may construct a solution to (34) by

\[
S_{t,t+1} = \max \{-x_{t+1} \cdot \alpha^*(z_t), 0\}.
\]

This is the nonnegative solution that minimizes the second moment. Formally optimization problem (36) is the conjugate to an optimization problem that seeks to find a nonnegative stochastic discount factor that prices the securities correctly whose second moment is as small as possible. Hansen and Jagannathan (1991) were interested in such problems as a device to restrict the set of admissible stochastic discount factors.\(^{16}\) As demonstrated by Luttmer (1996), convex constraints on portfolios can be incorporated by restricting the choice of \( \alpha \). In contrast to Hansen and Jagannathan (1991), Luttmer (1996) and Hansen et al. (1995), we have posed this problem conditionally. We say more about this distinction in the next subsection.

Another extraction choice follows Bansal and Lehmann (1997) and Cochrane (1992) by solving:

\[
\min_{\alpha} -E (\log [-\alpha(z_t) \cdot x_{t+1}] |z_t) - \alpha(z_t) \cdot q_t
\]

Provided this problem has a solution \( \alpha^* \), then

\[
S_{t,t+1} = -\frac{1}{\alpha^*(z_t) \cdot x_{t+1}}
\]

\(^{15}\)On the other hand, stochastic discount factors that are negative with positive probability can price incomplete collections of payoffs without inducing arbitrage opportunities.

\(^{16}\)While this solution need not be strictly positive with probability one, it is nevertheless useful in restricting the family of strictly positive stochastic discount factors.
is a strictly positive solution to (34). This particular solution gives an upper bound on $E[\log S_{t,t+1}|z_t]$. In this case the optimization problem is conjugate to one that seeks to maximize the expected logarithm among the family of stochastic discount factors that price correctly the vector $x_{t+1}$ of asset payoffs.

A variety of other constructions are also possible each of which is an extremal point among the family of stochastic discount factors. Conjugate problems can be constructed for obtaining bounds on convex functions of stochastic discount factors (as in the case of second moments) or concave functions (as in the case of logarithms). As an alternative, Snow (1991) considers bounding other than second moments and Stutzer (1996) constructs discount factors that limit the relative entropy of the implied risk neutral probabilities vis a vis the objective probability distribution.

Thus one empirical strategy is to give up on identification and characterize the family of solutions to equation (34). While this can be useful way to generate model diagnostics, its outcome for actual pricing can be very limited because the economic inputs are so weak. Alternatively, additional restrictions can be imposed, for example parametric restrictions or shape restrictions. Motivated by asset pricing models that exhibit habit formation Chen and Ludvigson (2004) specify a stochastic discount factor as a semiparametric function of current and lagged consumption. They use sieve minimum distance estimation in order to identify the shape of this function. In what follows we will focus on parametric restrictions. We consider estimation with parametric restrictions, say $S_{t,t+1} = f(y_{t+1}, \beta)$ for $\beta$ contained in a parameter space $\mathbb{P}$, a subset of $\mathbb{R}^k$, by fitting the conditional distribution of $x_{t+1}$ and $y_{t+1}$ conditioned on $z_t$. (As a warning to the reader, we have recycled the $\beta$ notation. While $\beta$ is now a vector of unknown parameter, $\exp(-\delta)$ is reserved for the subjective rate of discount. Also we will use the notation $\alpha$ for a different purpose than in section 2.)

### 7.2 Conditioning information

Gallant et al. (1990) fit conditional distributions parameterized in a flexible way to deduce conditional bounds on stochastic discount factors. Relatively, Wang (2003) and Roussanov (2005) propose ways of imposing conditional moment restrictions nonparametrically using kernel methods. An alternative is to convert the conditional moment restriction into an unconditional moment restriction by applying the Law of Iterated Expectations:

$$E[f(y_{t+1}, \beta)x_{t+1} - q_t] = 0.$$ 

A concern might be the loss of information induced by the conditioning down. As shown by Hansen and Singleton (1982) and Hansen and Richard (1987), this loss can be reduced by expanding the array of assets. For instance consider any vector of conditioning variables $h(z_t)$ with the same dimension as $x_{t+1}$. Then $x_{t+1} \cdot h(z_t)$ should

---

17Cochrane and Hansen (1992) show how to use such estimates to decompose the unconditional volatility of stochastic discount factors into on average conditional variability and unconditional variability in conditional means.
have a price \( h(z_t) \cdot q_t \). Thus it is straightforward to increase the number of asset payoffs and prices by forming synthetic securities with payoffs \( h(z_t) \cdot x_{t+1} \) and prices \( q_t \cdot h(z_t) \) through scaling by variables in the conditioning information set of investors.

If we perform such a construction for all possible functions of \( z_t \), that is if we verify that

\[
E[f(y_{t+1}, \beta)h(z_{t})'x_{t+1} - h(z_t)'q_t] = 0
\]

for any bounded Borel measurable vector of functions \( h \), then it is necessarily true that

\[
E[f(y_{t+1}, \beta)x_{t+1} - q_t|z_t] = 0
\]

This, however, replaces a finite number of conditional moment restrictions with an infinite number of unconditional moment restrictions. It suggests, however, a way to approximate the information available in the conditional moment restrictions through the use of unconditional moment restrictions.

For future reference, let \( X_{t+1} \) be the entire vector payoffs including the ones constructed by the econometrician and let \( Q_t \) be the corresponding price vector. The corresponding unconditional moment restriction is:

\[
E[f(y_{t+1}, \beta)X_{t+1} - Q_t] = 0.\tag{37}
\]

### 7.3 GMM estimation

In this discussion we work with the \( \ell \)-period extension of (37).

\[
E[f_{\ell}(y_{t+\ell}, \beta)X_{t+\ell} - Q_t] = 0.\tag{38}
\]

The most direct motivation for this is that the data used in the investigation are asset payoffs with a \( \ell \)-period horizon: \( f_{\ell}(y_{t+\ell}, \beta) \). If purchased at date \( t \), their payoff is at date \( t + \ell \).\(^{18}\) Then \( f_{\ell}(y_{t+\ell}, \beta) \) is the \( \ell \)-period stochastic discount factor. For instance, consider example 3.2. Then

\[
f_{\ell}(y_{t+\ell}, \beta) = \exp(-\delta) \left( \frac{C_{t+\ell}}{C_t} \right)^{-\gamma}
\]

where \( \beta = (\delta, \gamma) \).

Construct the function:

\[
\phi_t(\beta) = f_{\ell}(y_{t+\ell}, \beta)X_{t+\ell} - Q_t,
\]

of the unknown parameter vector \( \beta \). The pricing model implies unconditional moment restriction:

\[
E[f_{\ell}(y_{t+\ell}, \beta)X_{t+\ell} - Q_t] = 0.\tag{39}
\]

\(^{18}\)Considerations of aggregation over time leads some researchers to very similar econometric considerations, but only as an approximation. See Hall (1988) and Hansen and Singleton (1996). For a more ambitious attempt to address this issue via numerical simulation see Heaton (1995).
Using this as motivation, construct

$$\psi_T(b) = \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_t(b) \right]' W_T(b) \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_t(b) \right]$$

where the weighting matrix $W_t$ is adjusted to allow for the moving-average structure in error terms:

$$W_T(b) = \left[ Cov_T^0(b) + \sum_{j=1}^{T-1} (Cov_T^j(b) + Cov_T^j(b)') \right]^{-1}$$

where

$$Cov_T^j(b) = \frac{1}{T} \sum_{t=j+1}^{T} \phi_t(b)\phi_{t-j}(b)'$$

Then the so-called continuous updating GMM estimator (CU) suggested by Hansen et al. (1996) is given by

$$b_T = \arg \min_{b \in P} \psi_T(b),$$

although there are well known two-step and iterated alternatives. Hansen et al. (1996) give some comparisons of the approaches.

By construction, the GMM criterion function has close ties to the chi-square distribution. In particular when $b = \beta$, then

$$\psi_T(\beta) \Rightarrow \chi^2(n)$$

where $n$ is the number of moment conditions. As emphasized by Hansen et al. (1996), this by itself gives a way to conduct inferences about the unknown parameter vector. Construct the set of $b$’s for which $\psi_T(b)$ is less than a threshold value where the threshold value is obtained from the chi-square distribution.\(^{19}\) Stock and Wright (2000) show formally that such a method accommodates a form of weak identification and leads to robust inference. Alternatively,

$$\psi_T(\beta) - \min_{b \in P} \psi_T(b) \Rightarrow \chi^2(n - k)$$

where $k$ is number of free parameters. The minimized objective function is itself distributed as a chi-square as shown in Sargan (1958) for the linear simultaneous equations model and by Hansen (1982) for the more general GMM estimation environment. Moreover,

$$\psi_T(\beta) = \left[ \psi_T(\beta) - \min_{b \in P} \psi_T(b) \right] + \left[ \min_{b \in P} \psi_T(b) \right]$$

\(^{19}\)Stock and Wright (2000) relate this method to an inversion of the Anderson and Rubin (1949) statistic when specialized to the linear simultaneous equations model.
gives a decomposition of $\psi_T(\beta)$ into two components that are asymptotically independent and each have limiting chi-square distributions.

The limiting chi-square distribution for (42) presumes the local identification condition that matrix

$$E \left[ \frac{\partial \phi_t}{\partial \beta} \bigg|_{b=\beta} \right]$$

has full rank $k$. When the partial derivative matrix has reduced rank or when one considers a sequence of experiments with limiting singularity, as in the work of Stock and Wright (2000), the limiting chi-square distribution given in (42) is no longer valid. Limit approximation (41) remains valid, however. Kleibergen (2005) suggests an alternative approach to using this latter approximation to conduct inferences. To test a candidate value of $\beta$, he constructs a test based directly on the first derivative of the CU-GMM objective function. The limiting distribution has a convenient characterization and leads to an alternative chi-square distribution with degrees of freedom equal to the number of free parameters instead of the number of moment conditions. Interestingly, the test does not require the local identification condition.20 As discussed in Kleibergen (2005) this approach can be applied to testing restrictions and constructing confidence intervals. Also it can be used to produce an alternative decomposition of (43) that can help to distinguish parameter values for which first-order conditions are approximately satisfied but the underlying moment conditions are not satisfied.

7.4 GMM System Estimation

As we have seen, the stochastic discount factor formulation often leads to directly to a set of estimation equations, but these are estimation equations for a partially identified model. As an alternative, we add in the remaining components of the model and proceed with a system estimation. One stab at this is given in Hansen and Singleton (1996). The log linear, conditional counterpart to (39) in the case of the power utility model is

$$E \left[ -\gamma (\log C_{t+\ell} - \log C_t) \mathbf{1}_m + \log x_{t+\ell} | z_t \right] + \omega - \log q_t = 0$$

where $\mathbf{1}_m$ is an $m$-dimensional vector of ones and $\omega$ is an $m$-dimensional vector of constants introduced to compensate for taking logarithms and to capture the subjective rate of discount $\delta$. Here we are abstracting from conditional heteroskedasticity. For simplicity, lets suppose that $q_t$ is a vector of ones and hence its logarithm is a vector of zeros.

System (44) gives $m \ell$-period forecasting equations in $m + 1$ variables, the $m$ components of $\log x_{t+\ell}$ and $\log C_{t+\ell} - \log C_t$. Following Hansen and Singleton (1996) we could

20 It requires use of an alternative weighting matrix, one estimates the spectral density at frequency zero without exploiting the martingale structure implicit in multi-period conditional moment restrictions. For instance, $W_T(b)$ given in formula (40) can be replaced by the weighting matrix estimator of Newey and West (1987). While such an estimator tolerates much more general forms of temporal dependence, its rate of convergence is slower than that of (40). On the other hand, the spectral density estimators are, by construction, positive semidefinite in finite samples.
append an additional forecasting equation and estimate the full system as an \( m + 1 \) dimensional system of \( \ell \)-period forecasting equations. The reduced form is a system of forecasting equations for \( \log x_{t+1} \) and \( \log C_{t+\ell} - \log C_t \) conditioned on \( z_t \):

\[
\begin{bmatrix}
\log C_{t+\ell} - \log C_t \\
\log x_{t+\ell}
\end{bmatrix} = \Pi z_t + \varpi + w_{t+\ell}
\]

where

\[
E(w_{t+\ell} \otimes z_t) = 0.
\]

Then under restriction (44), the matrix \( \Pi \) satisfies:

\[
[-\gamma 1_m \quad I_m] \Pi = 0_m
\]

where \( 1_m \) is an \( m \)-dimensional vector of ones, \( I_m \) is an \( m \)-dimensional identity matrix and \( 0_m \) is an \( m \)-dimensional vector of zeros.

Notice that the (44) also implies the conditional moment restriction:

\[
E(\left[ -\gamma 1_m \quad I_m \right] w_{t+\ell} | z_t) = 0.
\]

Hansen and Singleton (1996) show that even if you impose the stronger condition that \( E(w_{t+\ell} | z_t) = 0 \).

in estimation, this does not distort the asymptotic inferences for the curvature parameter \( \gamma \). This means that the reduced-form equation can be estimated as a system GMM estimation, a weighting matrix constructed that does not require a prior or simultaneous estimation of \( \gamma \). Estimates of \( \gamma \) can be constructed as a restricted reduced-form system.

Hansen and Singleton (1996) produce inferences in the analogous ways as for the CUGMM estimator by constructing confidence sets from a GMM objective function by concentrating the all but the parameters of interest.

Notice that if \( E[\phi_t(b)] = 0 \) then it is also true that \( E[\Phi(\beta)\phi_t(\beta)] = 0 \) where \( \Phi \) is a function that maps elements of parameter space \( \mathbb{P} \) into nonsingular matrices. Thus we may use \( \phi_t(b) \) in constructing GMM estimators or \( \Phi(b)\phi_t(b) \). For instance in the log-linear power utility model just considered we might divide the moment conditions by \( \frac{1}{\gamma} \) and instead estimate \( \frac{1}{\gamma} \). Both this restricted reduced form method and the CUE method yield an estimator that is invariant to transformations of this type. The same estimator of the original parameter will be obtained, as is the case in maximum likelihood estimation. This invariance property is not shared by other methods such a two-step methods where a weighting matrix is constructed from an initial consistent estimator. Specifically, it is not satisfied by two-stage least squares when the structural equation to be estimated is over-identified.
7.5 Inference by Simulation

The shape of GMM objective, beyond just derivative calculations with respect to parameters, is informative. For low dimensional problems or problems with sufficient linearity, we can depict this function, its level sets, its behavior as we vary one parameter while minimizing out others. For nonlinear problems, an alternative convenient method is to follow Chernozhukov and Hong (2003) by constructing:

$$\varphi_T(b) \propto \exp\left[-\frac{1}{2}\psi_T(b)\right]$$

over the set $\mathbb{P}$ provided that this set is a compact subset of $\mathbb{R}^k$ with positive Lebesgue measure.\(^{21}\) The right-hand side function is scaled so that

$$\int_{\mathbb{P}} \varphi_T(b) db = 1$$

although there will be no need to compute this scaling factor analytically. The choice of the compact parameter space will be potentially important in applications.

Armed with this construction, we may now use MCMC (Markov chain Monte Carlo) methods to summarize properties of the function $\varphi_T$ and hence of $\psi_T$. Appendix D illustrates how to implement MCMC approach. MCMC methods are widely used in making Bayesian inferences, but also can be applied to this problem even though we will use a transform CU-GMM criterion function instead of a likelihood function.\(^{22}\) We use the MCMC approach as a way to systematically represent the shape of the GMM objective function via random parameter searches, but we will not attempt to give a Bayesian interpretation of this exercise.

Since $\varphi_T(b)$ may be treated mathematically as a density, we may infer “marginals” for individual components of the parameter vector averaging out the remaining components. This integration step is in contrast to practice of concentration that we discussed earlier, producing an objective over a single component of the parameter vector by minimizing the GMM objective over the remaining component for each hypothetical value of the single component. Using the random search embedded in MCMC, approximate level sets can also be inferred.\(^{23}\) Thus this approach can be used fruitfully in characterizing the behavior of the GMM objective function and offers an attractive alternative to minimization and computing derivatives at minimized values.

\(^{21}\)If $\mathbb{P}$ is not compact, then the objective could be scaled by a weighting function that has finite measure over $\mathbb{P}$.

\(^{22}\)To make this link, view the function $-\psi_T$ as the log-likelihood and $\varphi_T$ as the posterior density associated with a uniform prior over the parameter space.

\(^{23}\)Chernozhukov and Hong (2003) justify estimators of the parameter based on averaging or computing medians instead of minimizing the GMM objective.
7.6 Estimation under Misspecification

A feature of the weighting matrix $W_T$ in GMM is that it rewards parameter configurations that imply a large asymptotic covariance matrix. A parameter configuration might look good simply because it is hard to estimate, it is hard to reject statistically. A model specified at the level of a set of moments conditions is in reality only partially specified. Even if we knew the true parameters, we would not know the full time series evolution. If we did, we could form a likelihood function. When combined with a prior distribution over the parameters, we could compute the corresponding posterior distribution; and when combined with a loss function we could produce a parameter estimator that solves a Bayesian decision problem. The GMM estimation is meant to side step the specification of the full model, but at a cost of distancing the inferences from Bayesian methods.

Another way to address this issue is to repose the estimation problem by introducing model misspecification. Instead of aiming to satisfy the moment conditions, suppose we wish to get close to such a specification. This requires a formal statement of what is meant by close, and this choice will alter the population value of the objective. For instance, consider the mean square error objective of minimizing:

$$E \left( \left[ f_\ell(y_{t+\ell}, b) - S_{t,t+\ell} \right]^2 \right)$$

by choice of $S_{t,t+\ell}$ subject to:

$$E \left[ S_{t,t+\ell} X_{t+\ell} - Q_t \right] = 0.$$  

Since the space of stochastic discount factors $S_{t,t+\ell}$ that satisfies this moment restriction can be infinite dimensional, it is most convenient to work with the conjugate problem, which will need to be solved for each value of $b$. For fixed $b$ the conjugate problem is a finite-dimensional concave optimization problem. In this case of mean square approximation of the parameterized model to an admissible stochastic discount factor $S_{t,t+\ell}$, we follow Hansen et al. (1995) and Hansen and Jagannathan (1997) by using the conjugates problems:

$$\min_{b \in P} \max_{\alpha} E \left[ f_\ell(y_{t+\ell}, b)^2 - \left[ f_\ell(y_{t+\ell}, b) - \alpha \cdot X_{t+\ell} \right]^2 - 2\alpha'Q_t \right]$$

or

$$\min_{b \in P} \max_{\alpha} E \left[ f_\ell(y_{t+\ell}, b)^2 - \left[ \max \left\{ f_\ell(y_{t+\ell}, b) - \alpha \cdot X_{t+\ell}, 0 \right\} \right]^2 - 2\alpha'Q_t \right]$$

where in both cases the inner problem is concave in $\alpha$. The second conjugate problem is derived by restricting $S_{t,t+\ell}$ to be nonnegative while the first problem ignores this restriction.

In the case of problem (46), the inner maximization problem is solved by:

$$\alpha^*(b) = \left[ E (X_{t+\ell} X_{t+\ell}') \right]^{-1} E \left[ f_\ell(y_{t+\ell}, b) X_{t+\ell} - Q_t \right]$$

provided that $E (X_{t+\ell} X_{t+\ell}')$ is nonsingular. The concentrated objective function for problem (46) expressed as a function of $b$ is:

$$E \left[ f_\ell(y_{t+\ell}, b) X_{t+\ell} - Q_t \right]' \left[ E (X_{t+\ell} X_{t+\ell}') \right]^{-1} E \left[ f_\ell(y_{t+\ell}, b) X_{t+\ell} - Q_t \right],$$

72
which is the population GMM objective function evaluated using:

$$[E (X_{t+\ell} X_{t+\ell}')]^{-1}$$

as a weighting matrix. Importantly, this matrix does not depend on $b$. There is no reward for imprecision in estimation.

Alternatively, inner part of problem (47) (optimization over $\alpha$) does not have such a convenient analytical solution nor does it provide a simple link to GMM estimation, but it is constructed by restricting the admissible stochastic discount factors to be nonnegative. Specifically, the inner problem provides a solution to stochastic discount fact that satisfies the pricing restrictions of the form:

$$\max \{f_{\ell}(y_{t+\ell}, b) - \alpha^* \cdot X_{t+\ell}, 0\}.$$

The term $\alpha^* \cdot X_{t+\ell}$ is a correction term for misspecification, but is limited so that the resulting stochastic discount factor remains nonnegative.

The sample counterparts to problems (46) and (47) are saddle-point versions of $M$-estimation problems from the statistics literature instead of GMM estimation problems. In the sample counterpart problems, the sample average objective function is minimized instead of the population objective function.

Hansen and Jagannathan (1997) show that these two problems can be re-interpreted as ones in which the parameters are chosen to minimize pricing errors over alternative families of payoffs, where pricing errors are measured relative to the square root of the second moment of the payoffs. As a consequence, it is informative to characterize either:

$$\max_{\alpha} E \left( f_{\ell}(y_{t+\ell}, b)^2 - [f_{\ell}(y_{t+\ell}, b) - \alpha \cdot X_{t+\ell}]^2 - 2\alpha' Q_t \right),$$

$$\max_{\alpha} E \left( f_{\ell}(y_{t+\ell}, b)^2 - [\max \{f_{\ell}(y_{t+\ell}, b) - \alpha \cdot X_{t+\ell}, 0\}]^2 - 2\alpha' Q_t \right)$$

as a function of $b$ to assess model performance for alternative parameter values. Of course other measures of discrepancy between the modeled stochastic discount factor $f_{\ell}(y_{t+\ell}, b)$ and the stochastic discount factors $S_{t,t+\ell}$ that satisfy pricing restrictions can be employed. Provided the objective is convex in the stochastic discount factor $S_{t,t+\ell}$, we will be led to a conjugate problem that is concave in $\alpha$, the Lagrange multiplier on the pricing equation.

While we have formulated these as unconditional problems, there are obvious conditional counterparts that use $x_{t+\ell}$ in place $X_{t+\ell}$, $q_t$ in place of $Q_t$, and condition on $z_t$. Then while $\alpha$ is a function of $z_t$, the problem can be solved separately for each $z_t$.

### 7.7 Intertemporal Elasticity Estimation

Consider first estimation that features a specific assets and other payoffs constructed via scaling. We use the power utility specification and make no attempt to separate risk aversion and intertemporal substitution. Arguably, this is designed to feature estimation
of the intertemporal substitution elasticity because by focusing on time series data about a single return, the estimation is not confronting evidence about risk prices. In our first order expansion of the risk free rate, we saw the impact of both $\rho$ and $\gamma$ on returns. Arguably the impact of changes in $\rho$ might be more potent than changes in $\gamma$, and subsequently we will consider multiple returns and the resulting information about $\gamma$. Specifically, we will freely estimate $\rho$ with a single return in this subsection and then estimate $\gamma$ for fixed alternative values of $\rho$ when we study multiple returns in the section 7.9.

### 7.7.1 Treasury Bills

Let $x_{t+1}$ be the quarterly return to holding Treasury bills, which has price one by construction. In addition to this return we construct two additional payoffs scaling by consumption ratio between dates $t$ and $t-1$, $C_t/C_{t-1}$ and the date $t$ Treasury bill. Thus there were a total of three moment conditions. Nominal Treasury bill returns were converted to real returns using the consumption deflator. We used per-capita consumption.

To facilitate the discussion of inference based on the CU-GMM criterion functions, in figure 20 we report plots of the concentrated criterion function constructed by minimizing with respect to $\delta$ holding $\rho$ fixed over a range of values. We also report the values of the discount rate $\delta$ that minimize the criterion concentrated over $\rho$. The criterion function is minimized at large values of $\rho$ if we do not restrict $\delta$. When we restrict $\delta > 0$, this restriction binds for modest values of $\rho$ and there is notable curvature in the objective function to the right of $\rho = 0.5$. On the other hand, the criterion is very large even at the minimized parameter values. Apparently, it is not possible to satisfy all three moment conditions, even if we allow for sampling uncertainty.

In figure 21 we construct the payoffs differently. We lag the consumption growth factor and return to Treasury bills one period to remove the effect of overlapping information induced by time aggregation. We also set $\ell = 2$ when constructing the weighting matrix. The shape of the objective (with $\delta$ concentrated out), is very similar to that of figure 20 except that it is shifted down. While reduction in the objective function is to be expected because the conditioning information is less potent, the objective function is still quite large. The nonnegativity restriction remains important for inducing curvature to the right of $\rho = 0.5$.

Other researchers have argued that the study of interest rate Euler equation is fertile territory for weak instrument asymptotics, or more generally for weak formulations of identification. While the evidence for predictability in consumption growth is weak, risk free rates are highly predictable. This is potentially powerful identifying information, suggesting perhaps that the intertemporal elasticity of consumption is very small, $\rho$ is

---

24 Stock and Wright (2000) consider setups in which the expected derivative matrix of the moment conditions drifts to a singular matrix. For the log linear version of the Euler equation, we might ask that the projection of consumption growth onto $z_t$ drifts to zero. If the projection of the Treasury bill onto $z_t$ does not also drift to zero then the coefficient of interest, $\rho$ must drift, changing the nature of the large sample embedding. See Hansen and Singleton (1983) for a related discussion.
Figure 20: Continuously-updated GMM criterion function for the Treasury bill Euler equation: for $\ell = 1$. The top panel depicts the objective function with and without the constraint that $\delta = 0$. The bottom panel gives the associated values of $\delta$ obtained by minimizing the GMM objective for each value of $\rho$. The parameter $\delta$ is expressed as percent per annum.
large. Given the observed consumption growth, a large value of $\rho$ requires a negative subjective rate of discount. Unfortunately, as we have seen this simple argument for large values of $\rho$ ignores restrictions on $\delta$ and the overall statistical evidence against the model. Considerations of weak identification are more germane for the study of value-weighted returns.

### 7.7.2 Market Return

Next we let $x_{t+1}$ be the value-weighted return. We form two additional payoffs by using consumption growth between date $t - 1$ and $t$ along with the date $t$ dividend price ratio. The results are depicted in figure 22. The objective function is lower than for Treasury bills. Again the imposition of a nonnegativity constraint is inducing curvature in the objective function, in this case to the right of $\rho = 3.5$. For market returns there is considerably less evidence against the model, but also a very limited statistical evidence about $\rho$.

The results when the scaling variable is shifted back one time period are given in figure 23. Again the shape is similar, and the objective functions is a bit lower.

### 7.8 CES Preferences and the Wealth Return

While the CES parameterized version of the recursive utility model gives a leading example of a stochastic discount factor model, as we have seen the stochastic discount factors depend on continuation values. We have already explored constructions of these values and their use in empirical investigation. Typically, the computation of continuation values requires a complete specification of the consumption dynamics. In this section we have abstracted from that complication. As emphasized by Epstein and Zin (1989b), an appropriately constructed measure of the wealth return can be used in place of continuation values as we now verify.

Pricing the next period wealth is equivalent to imputing the shadow price to the next period continuation value. Thus we are led to compute:

$$
\frac{E[V_{t+1}MV_{t+1}|F_t]}{MC_t} = \left[ \frac{\exp(-\delta)}{1 - \exp(-\delta)} \right] E \left[ (V_{t+1})^{1-\gamma}|F_t \right] \left[ R(V_{t+1}|F_t) \right]^{\gamma-\rho} (C_t)^\rho
$$

where

$$
R(V_{t+1}|F_t) = \left( E \left[ (V_{t+1})^{1-\gamma}|F_t \right] \right)^{1-\gamma}.
$$

---

25The chi-square critical values for two degrees of freedom are 6.0 for probability value of .05 and 9.2 for a probability value of .01. Since the nonnegativity constraint on $\delta$ sometimes binds the chi-square critical values for three degrees of freedom also give a useful reference point. They are 7.8 for probability .05 and 11.3 for probability .01.
Figure 21: Continuously-updated GMM criterion function for the Treasury bill Euler equation: for $\ell = 2$. The top panel depicts the objective function with and without the constraint that $\delta = 0$. The bottom panel gives the associated values of $\delta$ obtained by minimizing the GMM objective for each value of $\rho$. The parameter $\delta$ is expressed as percent per annum.
Figure 22: Continuously-updated GMM criterion function for the market return Euler equation: $\ell = 1$. The top panel depicts the objective function with and without the constraint that $\delta = 0$. The bottom panel gives the associated values of $\delta$ obtained by minimizing the GMM objective for each value of $\rho$. The parameter $\delta$ is expressed as percent per annum.
Figure 23: Continuously-updated GMM criterion function for the market return Euler equation: $\ell = 2$. The top panel depicts the objective function with and without the constraint that $\delta = 0$. The bottom panel gives the associated values of $\delta$ obtained by minimizing the GMM objective for each value of $\rho$. The parameter $\delta$ is expressed as percent per annum.
Thus the return on wealth is given by:

\[ R_{t+1}^w = \exp(\delta) \left( \frac{C_{t+1}}{C_t} \right)^{\rho} \left[ \frac{V_{t+1}}{R(\mathcal{F}_t)} \right]^{1-\rho}. \]

Recall that our previous empirical calculations presumed that \( \gamma = \rho \). If we mistakenly impose this restriction, then the Euler equation error is:

\[ \exp(-\delta) \left( \frac{C_{t+1}}{C_t} \right)^{-\rho} R_{t+1}^w = \left[ \frac{V_{t+1}}{R(\mathcal{F}_t)} \right]^{1-\rho}. \]

Suppose that the continuation value is conditionally normally distributed with variance \(|\sigma_{v,t}|^2\). While this will typically not be case, it can be justified by taking continuous time limits along the lines we have discussed previously. Then the conditional expectation for this misspecified model is:

\[ \exp \left[ (1-\rho) \frac{(\gamma - \rho)}{2} |\sigma_{v,t}|^2 \right] \]

This distortion can be bigger or less than unity depending on whether or not \( \gamma \) is less than or greater than \( \rho \). To the extent that correction is almost constant, it can be absorbed into the subjective rate of discount. Thus GMM estimation with this form of misspecification at the very least alters the restriction imposed on the (potentially distorted) subjective discount rate. Recall that the subjective rate of discount can be an important source of identifying information.

The case of \( \gamma = 1 \) gives an interesting benchmark. In this special case the log-linear version of the Euler equation holds with:

\[ -\delta - \rho [\log C_{t+1} - \log C_t] + \log R_{t+1}^w = (1-\rho) (\log V_{t+1} - E [\log V_{t+1}|\mathcal{F}_t]). \]

(See Epstein and Zin (1989b) for an original reference.) In this special case it is not necessary to use the constant term to even approximately correct for volatility in either consumption or the return to wealth. The constant term captures the true subjective rate of discount for investors. Large values of \( \rho \) (small values of \( \frac{1}{\rho} \)) are ruled out by the positive growth rate in per-capita consumption. More generally, studies like those of Hansen and Singleton (1996), and Yogo (2004) report inferences that apparently tolerate large values of \( \rho \), but they ignore restrictions on the constant term. This additional information can be very informative as we have illustrated.\footnote{On the other hand, the notion of using single returns to identify \( \rho \) independently of \( \gamma \) is typically compromised. The value of \( \gamma \) determines in part what the distortion is in the subjective rate of discount induced by omitting continuation values from the analysis.}
7.9 Multiple Assets and Markov Chain Monte Carlo

When $\rho \neq 1$, we may invert the relation between continuation values and the return on the wealth portfolio as suggested by Epstein and Zin (1989b):

$$
\frac{V_{t+1}}{R(V_{t+1}|F_t)} = \left[\exp(-\delta)R^w_{t+1}\right]^{\frac{1}{1-\rho}} \left(\frac{C_{t+1}}{C_t}\right)^{\frac{\rho}{1-\rho}}.
$$

Thus an alternative stochastic discount factor is:

$$
S_{t,t+1} = \exp(-\delta) \left(\frac{C_{t+1}}{C_t}\right)^{-\rho} \left[\frac{V_{t+1}}{R(V_{t+1}|F_t)}\right]^\rho \left(\frac{C_{t+1}}{C_t}\right)^{\frac{\rho(\gamma-1)}{1-\rho}} R^w_{t+1}^{\frac{\rho}{\gamma}}.
$$

The Euler equation for a vector $X_{t+1}$ of asset payoffs with corresponding price vector $Q_t$ is

$$
E \left(\left[\exp(-\delta)\right]^{\frac{\rho(\gamma-1)}{1-\rho}} \left(\frac{C_{t+1}}{C_t}\right)^{\frac{\rho(\gamma-1)}{1-\rho}} (R^w_{t+1})^{\frac{\rho}{\gamma}} X_{t+1} - Q_t z_t\right) = 1
$$

where $R^w_{t+1}$ is the return on the total wealth portfolio.

In the empirical analysis that follows, we follow Epstein and Zin (1989b) by using the market return as a proxy for the return on the wealth portfolio. Since the market return omits important components to investor wealth, there are well known defect in this approach that we will not explore here. Also, we impose some severe restrictions on $\rho$ as a device to illustrate the information available for identifying $\gamma$ and $\delta$. Freely estimating $\rho$ is problematic because of the poor behavior of the CU-GMM objective in the vicinity of $\rho = 1$. This poor behavior is a consequence of our using an empirical proxy for the return on the wealth portfolio in constructing the stochastic discount factor.

We apply the MCMC simulation method described previously to estimate $\gamma$ and $\delta$ for alternative choices of $\rho$. This gives us a convenient way to summarize the shape of the CU-GMM criterion function through the use of simulation instead of local approximation. A consequence of our stochastic discount factor construction is that the market portfolio cannot be used as one of the test assets and $\rho = 1$ cannot be entertained. Instead we use the “value minus growth” excess return constructed using the portfolios sorted on book-to-market equity, together with Treasury bill return, in order to identify the preference parameters. The scaling factor for the Treasury bill return are the same ones we used previously, the consumption growth factor between $t-1$ and $t$ and the time $t$ Treasury bill return. The value-growth excess return is scaled by the consumption growth factor and the date $t$ value-growth excess return. Thus we use six moments conditions in estimation.

In our estimation we use two different values of $\rho$, $\rho = .5$ and $\rho = 1.5$ and estimate $\gamma$ and $\delta$ subject to the constraints that $0 \leq \delta \leq 5$ and $0 \leq \gamma \leq 10$ where $\delta$ scaled by 400 so that it is expressed as a percent per annum. The resulting histograms are reported in figures 24 and 25. When $\rho = .5$, the histogram for $\delta$ is very much tilted toward zero,
and the histogram for $\gamma$ is very much tilted towards ten. The parameter space bounds play an important role in these calculations, but it is straightforward to impose other bounds. When $\rho = 1.5$, the histogram for $\gamma$ is centered around 3.5, but the histogram for $\delta$ is very much tilted towards the upper bound of five. Increasing the upper bound on $\delta$ causes the $\gamma$ distribution to shift to the right. Thus our chosen upper bound on $\delta$ induces a modest estimation of $\gamma$. The lowest CU-GMM objective encountered in the random search is 9.8 for $\rho = 0.5$ and 21.7 for $\rho = 1.5$ suggesting that there is considerably less evidence against the specification with a lower value of $\rho$.  

The CU-GMM criterion function has the property that the parameter configurations that induce considerable sampling uncertainty in the moment conditions are tolerated because the weighting matrix is the inverse of the sample covariance matrix. For instance, large values of $\gamma$ may induce large pricing errors but nevertheless be tolerated. To explore this possibility, we compute the implied specification errors using the weighting matrix described previously. This weighting matrix is invariant to the parameters and instead comes from a best least squares fit of a misspecified model. The outcome of this calculation is depicted in figure 26 for $\rho = 0.5$ and in figure 27 for $\rho = 1.5$. When $\rho = 0.5$, the lower bound of zero on $\delta$ binds, and the specification errors become large for large values of $\gamma$. When $\rho = 1.5$, the upper bound of zero binds for large values of $\gamma$ which in turn leads to large specification errors. For both figures the implied value of $\delta$ when $\gamma$ is near one becomes enormous to offset the fact that the subjective discount factor is being raised to a very small number.

---

27As a point of reference, the critical values for the chi-square distribution with 4 degrees of freedom are 9.5 for a probability value of .05 and 13.3 for a probability value of .01. Given the important role of the constraints on parameters, the chi-square distribution with five degrees of freedom gives an alternative interesting benchmark. The critical values are 11.1 for a probability value of .05 and 15.1 for a probability value of .01.
Figure 24: MCMC with the continuously-updated GMM criterion function: $\rho = .5$. The histograms are scaled to integrate to one. The parameter $\delta$ is restricted to be in the interval $[0, 5]$ expressed as an annualized percent, and the parameter $\gamma$ is restricted to be in the interval $[0, 10]$. The smallest CU-GMM objective encountered in the random search was 9.8.
Figure 25: MCMC with continuously-updated GMM criterion function: $\rho = 1.5$. The histograms are scaled to integrate to one. The parameter $\delta$ is restricted to be in the interval $[0, 5]$ expressed as an annualized percent, and the parameter $\gamma$ is restricted to be in the interval $[0, 10]$. The smallest CU-GMM criterion function value that was encountered in the random search is 21.7.
Figure 26: Specification errors: $\rho = .5$. The top panel gives the specification error as a function of $\gamma$ when the value of $\delta$ is chosen to minimize the pricing error objective. This pricing error is expressed as the mean-square distance from the misspecified stochastic discount factor to the closest random variable that prices on average the vector of assets. Alternatively, it is the maximal average pricing error per mean-square unit of payoff norm. The bottom panel gives the minimizing choices of $\delta$ for each value of $\gamma$. 
Figure 27: Specification errors: $\rho = 1.5$. The top panel gives the specification error as a function of $\gamma$ when the value of $\delta$ is chosen to minimize the pricing error objective. This pricing error is expressed as the mean-square distance from the misspecified stochastic discount factor to the closest random variable that prices on average the vector of assets. Alternatively, it is the maximal average pricing error per mean-square unit of payoff norm. The bottom panel gives the minimizing choices of $\delta$ for each value of $\gamma$. 
8 Conclusions

Our chapter explores the role of intertemporal substitution and risk aversion in asset pricing. We feature the CES recursive utility model, but of course other asset pricing models warrant comparable consideration. Parameters extracted from other sources, including micro or experimental evidence can be inputs into an analysis of the asset pricing implications of models. For example, Malloy et al. (2005) uses evidence from household level data to explore macroeconomic risk. Even with known preference parameters, measurements of macroeconomic risk exposures are required for quantitative prediction. Since the intertemporal composition of risk can play a central role in asset valuation, this puts an extra premium on the measurement of long-run components to risk. We have not embarked on a comprehensive survey of the empirical literature, but we now explore some of the challenges.

The parameter governing the intertemporal elasticity of substitution is the key linking consumption and wealth. For this link we find it useful to feature the role of continuation values. Since the CES aggregator is homogeneous of degree one, these continuation values encode the shadow values of wealth. In effect the continuation values appropriately scaled give us one side of the intertemporal budget constraint and direct measures of wealth is the other side. There is a return counterpart to this link that has been featured in some portions of the asset pricing literature, but the return based formulations typically omit information, in particular information linking current responses of consumption and wealth.

As we have illustrated following the work Lettau and Ludvigson (2001), use of consumption and financial wealth leads to a macroeconomic version of Shiller (1981)’s excess sensitivity puzzle. There is substantial variability in financial wealth that is not reflected in aggregate consumption. This opens up a variety of measurement challenges that have been explored in the asset pricing literature. For example, financial wealth omits any contribution of labor income (see Campbell (1996) and Jagannathan and Wang (1996) for studies of implications for pricing returns), but the remaining challenge is how to measure and credibly price the corresponding labor income risk exposure. Related to this, Lustig and Nieuwerburgh (2006) explore the required stochastic properties of the omitted components of wealth that are required to repair the model implications.

The use of aggregate nondurable consumption might also be too narrow. For this reason, many studies expand the definition of consumption and refine the preference assumptions when examining both the cross section and time series of asset returns. For example, Piazzesi et al. (2006) consider a separate role for housing, Yogo (2006) and Pakos (2006) examine the importance of consumer durables, and Uhlig (2006) considers leisure. Including these other components of consumption may also prove fruitful for our understanding of the wealth-consumption link. Further as emphasized by Uhlig (2006) these components are also germane to the evaluation of risk embedded in continuation values.

In this chapter we have been guilty of pushing the representative consumer model
too hard. As an alternative to broadening the measure of wealth, we might focus on narrowing the definition of the marginal investor. Heaton and Lucas (2000) and others explore important aspects of investor heterogeneity, participation, market segmentation and limited risk sharing. Others, including Alvarez and Jermann (2000) and Lustig (2004) consider models in which there are important changes over time in the marginal investor participating in market. These changes induce an extra component to risk prices. All of these models provide an alternative valuable frameworks for measurement. They do not, however, remove from consideration the modeling and measurement questions explored in this chapter.

Claims made in the empirical literature that intertemporal substitution can be inferred from the study of single asset returns such as Treasury bills or the risk free rate require qualification.\(^{28}\) They ignore potentially important information that is often buried in the constant terms of log-linear estimation. We have seen how this additional information can rule out small values of the intertemporal substitution parameter (large values of \(\rho\)). The crude counterpart to this that abstracts from uncertainty can be seen by setting the subjective rate of discount to zero and comparing the growth rate of consumption to that of the average logarithm of returns. Excessively large values of our parameter \(\rho\) are inconsistent with the observed relation between means. While suggestive, this simple imitation of macro calibration is not formally correct in this context. As we have seen, the risk aversion parameter also comes into play. Separation can only be achieved as an approximation that abstracts from potentially important sample information.\(^{29}\)

GMM inferences that explore shapes of the objective function through concentration or simulation are often the most revealing, even if they fail to achieve the simplified aims of Murray (2005). While the continuously-updated-GMM estimation has some advantages in terms of reliable inference, it can also reward parameter configurations that make the implied moment conditions hard to estimate. Thus naive use of such methods can lead to what turn out to be uninteresting successes. It is valuable to accompany such estimation with explorations of implied pricing errors or other assessments of potential misspecification.

Consumption-based models with long-run risk components pose interesting statistical challenges because they feature macroeconomic risk exposure over long horizons. Macroeconomic growth rate risk is reflected in continuation values, and continuation values contribute to risk prices defined both locally and in long run. These prices along with cash-flow and return risk exposure determine the heterogeneity in asset prices. Investor risk preference are thus encoded in the predicted asset prices and expected returns. We have illustrated why this source of identifying information about investor risk preferences presents challenges for reliable measurement. Here we have illustrated this using VAR methods to assess such estimates. For more general specifications nonlinear solution methods and estimation methods will come into play.


\(^{29}\)Even in the power utility model with stochastic consumption, risk free rates are sometimes plausible with very large value of \(\rho\) as revealed by the volatility correction in a log-normal approximation.
The incorporation of more formal macroeconomics promises to aid to our understanding of sources of long run risk. Work by Fisher (2006), Mulligan (2001) and others is suggestive of such links. Both use production-based macroeconomic models. Fisher focuses on long run potency of alternative sources of technology shocks. Mulligan (2001) considers consumption - physical return linkages as an alternative to the study of financial returns. Although stochastic volatility in consumption can potentially have long-run effects as well, this additional source of risk should ultimately have its source in shocks to technology and other economic fundamentals. Exploring these features in more fully specified models and focusing on long-run components hold promise for aiding our understanding of asset price heterogeneity.
A Additional formulas for the Kreps-Porteus model

A.1 Discrete time

Recall that \( v_t - c_t = U_v \cdot x_t + \mu_v \) where the formulas for \( U_v \) and \( \mu_v \) are given in (21).

Write:

\[
(v_t^1 - c_t)^2 = x_t' \Lambda x_t + 2 \lambda' x_t + \ell.
\]

We look for a solution for the derivative of the form:

\[
Dv_t^1 = - \left( \frac{1}{2} X_t' \Omega X_t + X_t \cdot \omega + \frac{w}{2} \right)
\]

where

\[
\Omega = \frac{(1 - \beta)}{\beta} \Lambda + \beta A' \Omega A
\]

\[
\omega = \frac{(1 - \beta)}{\beta} \lambda + \beta (1 - \gamma) A' \Omega B (H + B' U_v) + \beta A' \omega
\]

\[
w = \frac{(1 - \beta)}{\beta} \ell + \beta (1 - \gamma)^2 (H + B' U_v)' B' \Omega B (H + B' U_v) \\
+ 2 \beta (1 - \gamma) \omega' B (H + B' U_v) + \beta \text{trace}(B' \Omega B) + \beta w.
\]

The first equation in (49) is a Sylvester equation and is easily solved. Given \( \Omega \), the solution for \( \omega \) is:

\[
\omega = (I - \beta A)^{-1} \left( \frac{1 - \beta}{\beta} \lambda + \beta (1 - \gamma) A' \Omega B (H + B' U_v) \right),
\]

and given \( \omega \), the solution for \( w \) is obtained similarly by solving the third equation of (49).

Next we produce a formula for \( Ds_{t+1,t} \) based on equation (20). From our previous calculations

\[
-(c_{t+1} - c_t) + [v_{t+1}^1 - Q_t(v_{t+1})] = U_v' B w_{t+1} - G' x_t - \mu_c - \frac{1 - \gamma}{2} |U_v' B + H|^2
\]

Using our formulas for \( Dv_{t+1} \) for the distorted conditional expectation:

\[
Dv_{t+1}^1 - E^* (Dv_{t+1}^1 | F_t) = -\frac{1}{2} (B w_{t+1}^*)' \Omega B w_{t+1}^* + \frac{1}{2} \text{trace}(B' \Omega B) \\
- (B w_{t+1}^*)' \Omega [A x_t + (1 - \gamma) B (H + B' U_v)] - \omega' B w_{t+1}^*.
\]

Substituting for \( w_{t+1}^* \) from the relation

\[
w_{t+1} = w_{t+1}^* + (1 - \gamma) [H + \beta B' (I - A' \beta)^{-1} G]
\]

we may implement formula (20) via,

\[
Ds_{t+1,t}^1 = \frac{1}{2} w_{t+1}' \Theta_0 w_{t+1} + w_{t+1}' \Theta_1 x_t + \theta_0 + \theta_1 x_t + \theta_2 w_{t+1}
\]

by constructing the coefficients \( \Theta_0, \Theta_1, \theta_0, \theta_1, \theta_2 \).
A.2 Continuous time

In what follows, we derive the equations implied by (27) that can be used to compute the derivative of the value function in practice. Many readers may choose to skip this part.

To construct the solution, form the state vector:

\[ X_t = \begin{bmatrix} x_t \\ z_t \end{bmatrix} \]

and write composite state evolution (26) as:

\[ dX_t = \tilde{A}X_t dt + \tilde{F}dt + \sqrt{z_t}\tilde{B}_1 dW^*_t + \sqrt{z_t}\tilde{B}_2 d\bar{W}^*_t, \]

and write

\[ (U_v \cdot x + \tilde{U}_v z + \mu_v)^2 = X'\Lambda X + 2\lambda'X + \ell. \]

Look for a derivative expressed as:

\[ Dv^1_t = -\left( \frac{1}{2}X'_t\Omega X_t + X_t \cdot \omega + \frac{w}{2} \right) \]

Substituting into equation (27), \( \Omega \) solves:

\[ -\delta \Lambda + \delta \Omega = \tilde{A}'\Omega + \Omega\tilde{A}; \]

\( \omega \) solves:

\[ -\delta \lambda + \delta \omega = \Omega \tilde{F} + \tilde{A}'\omega + \left[ \frac{1}{2} \text{trace} (\Omega \tilde{B}_1 \tilde{B}_1') + \frac{1}{2} \text{trace} (\Omega \tilde{B}_2 \tilde{B}_2') \right]; \]

and \( w \) solves:

\[ -\delta \ell + \delta w = 2\omega \cdot \tilde{F}. \]

These three equations should be solved in sequence.

Given this solution we may compute the shock exposure vector for the derivative as follows:

\[ \begin{bmatrix} D\sigma'_{v,t} \\ D\bar{\sigma}'_{v,t} \end{bmatrix}' = -[\Omega X_t + \omega]' \begin{bmatrix} \tilde{B}_1 & \tilde{B}_2 \end{bmatrix} = -[\Omega X_t + \omega]' \begin{bmatrix} B & 0 \\ 0 & \tilde{B} \end{bmatrix}. \]

Using these formulas, the risk prices are:

i) \( dW_t: \sqrt{z_t}\rho H' + \sqrt{z_t(\gamma - \rho)}(B'U_v + H)' - \sqrt{z_t}(\rho - 1)(\gamma - 1)[\Omega X_t + \omega]' \tilde{B}_1; \)

ii) \( d\bar{W}_t: \sqrt{z_t}\rho \bar{H} + \sqrt{z_t(\gamma - \rho)}(\tilde{B}U_v + \bar{H}) - \sqrt{z_t}(\rho - 1)(\gamma - 1)[\Omega X_t + \omega]' \tilde{B}_2. \)
B Bayesian Confidence Intervals

Consider the VAR:

\[ A(L) y_t + C_0 + C_1 t = w_t \]

where \( y_{t+1} \) is \( d \)-dimensional. The matrix \( A(0) = A_0 \) is lower triangular. We base inferences on systems that can be estimated equation-by-equation. The \( w_t \) is assumed to be normal with mean zero and covariance matrix \( I \). We follow Sims and Zha (1999) and Zha (1999) by considering a uniform prior on the coefficients and we follow Zha (1999) by exploiting the recursive structure of our models.

Write a typical equation as:

\[ \alpha z_t + \gamma \cdot x_t = v_t \]

where \( v_t \) is distributed as a standard normal, \( x_t \) is a vector of variables that are uncorrelated with \( v_t \), but \( z_t \) is correlated with \( v_t \). This equation can be transformed to a simple regression equation of \( z_t \) onto \( x_t \) with regression coefficients \( \beta = -\frac{1}{\alpha} \gamma \) and regression error variance \( \sigma^2 = \frac{1}{\alpha^2} \). Imposing a uniform prior over \( (\alpha, \gamma) \) does not imply a uniform prior over the regression coefficients, however.

The piece of the likelihood for sample of \( T \) observations pertinent for this equation has the familiar form:

\[
\ell_T \propto |\alpha|^T \exp \left[-\frac{T}{2} \sum_{t=1}^{T} \left( \alpha z_t + x_t \cdot \gamma \right)^2 \right]
\]

Consider first the posterior distribution of \( \gamma \) given \( \alpha \). Using familiar calculations e.g. see Box and Tiao (1973), it follows that

\[ \gamma \sim \text{Normal}(-\alpha b_T, V_T) \]

where \( b_T \) is the least squares estimate obtained by regressing \( z_t \) onto \( x_t \), and

\[ V_T = \left( \sum_{t=1}^{T} x_t x_t' \right)^{-1}. \]

The marginal posterior for \( \alpha \) has a density that is proportional to

\[ |\alpha|^T \exp \left(-\frac{\alpha^2 T s_T}{2} \right) \]

where \( s_T \) is the least squares residual variance

\[ s_T = \frac{1}{T} \sum_{t=1}^{T} (z_t - x_t \cdot b_T)^2. \]

This is just a special case of a formula Theorem 2 of Zha (1999).
It is convenient to use the distribution for $\nu = \alpha^2 TsT$. By the change-of-variables formula the density for $\nu$ is proportional to:

$$\nu^{\frac{T-1}{2}} \exp \left( -\frac{\nu}{2} \right),$$

which is the chi-square density with $T + 1$ degrees of freedom.

We simulate the joint posterior by first simulating $\nu$ using the chi-square distribution, then constructing $\alpha$ up to sign, and finally simulating $\gamma$ conditioned on $\alpha$ according to a normal distribution.

Given the recursive nature of our model, we may follow Zha (1999) by building the joint posterior for all parameters across all equations as a corresponding product. This requires that we include the appropriate contemporary variables on the right-hand side of the equation to ensure that $w_{t+1}$ has the identity as the covariance matrix. In effect we have divided the coefficients of the VAR into blocks that have independent posteriors given the data. We construct posterior confidence intervals for the objects that interest us as nonlinear functions of the VAR coefficients.

C MCMC

The MCMC simulations follow a version of the standard Metropolis-Hastings algorithm (see Chernozhukov and Hong (2003)). Let the parameter combination corresponding to the $i$’th draw be $b^{(i)} = [\delta^{(i)}, \gamma^{(i)}]$ (since we hold $\rho$ constant in these simulations, we omit reference to it here). Then

1. draw $b^{(0)}$ from the prior distribution (uniform on $A$)
2. draw $\zeta$ from the conditional distribution $q(\zeta | b^{(i)})$
3. with probability $\inf \left( \frac{\exp(-\psi_T(b^{(i+1)}))q(b^{(i)}|\zeta)}{\exp(-\psi_T(b^{(i)}))q(\zeta|b^{(i)})}, 1 \right)$ update $b^{(i+1)} = \zeta$; otherwise keep $b^{(i+1)} = b^{(i)}$

A typical choice of transition density is Gaussian, which results in a Markov chain that is a random walk. We are interested in constraining the parameter space to a compact set. Therefore an adjustment needs to be made for truncating the distribution. Specifically, let $\phi$ be the bivariate normal density centered around zero with cdf $\Phi$. Then

$$q(x|y) = \frac{\phi(x-y)}{\Pr(x \in A)},$$

where $x = y + z$, $z \sim \Phi$,

which can be computed straightforwardly. In simulations, the truncation is accomplished by discarding the values of $\zeta$ that fall outside of $A$. A choice needs to be made regarding the dispersion of $\phi$. Too large a variance would generate too many truncations and thus
result in slow convergence while too low a value would produce a very slowly-moving random walk that might fail to visit substantial regions of the parameter space and also lead to slow convergence. We set the standard deviations of $\phi$ for both parameters equal to their respective ranges divided by 50. The reported results are based on simulations with 1,000,000 draws.

D Data Description

Data: population is from NIPA Table 2.1. Risk-free rate is the 3-month Treasury Bill rate obtained from CRSP Fama Risk Free Rates files.

Book-to-Market Portfolios: Returns to value weighted portfolios of stocks listed on NASDAQ, AMEX and NYSE. Stocks sorted by book-to-market value of equity. Construction of these portfolio returns is detailed in Hansen et al. (2005).

Consumption: Aggregate US consumption of nondurables and services as reported in the National Income and Product Accounts of the United States. Seasonally adjusted and converted to real units using the implicit price deflators for nondurables and services. Quarterly from 1947 to 2006.


Dividends: Constructed from the portfolio returns “with” and “without” dividends. Seasonality removed by taking a moving average. Construction of this series is detailed in Hansen et al. (2005).

Market Return: Value weighted return to holding stocks listed on NASDAQ, AMEX and NYSE. Constructed from CRSP data base. Quarterly from 1947 to 2006.


Risk Free Rate: Three-month Treasury Bill return from CRSP. Quarterly from 1947 to 2006.

Wages and Salaries: Wages and salary disbursement from the National Income and Product Accounts of the United States. Seasonally adjusted and converted to real units using the implicit price deflators for nondurables and services. Quarterly from 1947 to 2005.
Wealth: Total financial assets of the United States personal sector less Total liabilities as reported in table L.10 of the Flow of Funds Accounts of the United States. Quarterly from 1952 to 2005.
References


