Beliefs, Doubts and Learning:
Valuing Macroeconomic Risk

Lars Peter Hansen[*]

Published in the May 2007 Issue of the *American Economic Review*[†]

This essay examines the problem of inference within a rational expectations model from two perspectives: that of an econometrician and that of the economic agents within the model. The assumption of rational expectations has been and remains an important component to quantitative research. It endows economic decision makers with knowledge of the probability law implied by the economic model. As such, it is an equilibrium concept. Imposing rational expectations removed from consideration the need for separately specifying beliefs or subjective components of uncertainty. Thus it simplified model specification and implied an array of testable implications that are different from those considered previously. It reframed policy analysis by questioning the effectiveness of policy levers that induce outcomes that differ systematically from individual beliefs.

I consider two related problems. The first is the problem of an econometrician who follows Muth (1961), Lucas and Prescott (1971), Lucas (1972a), Sargent (1973) and an extensive body of research by adopting an assumption of rational expectations on the part of economic agents. In implementing this approach, researchers abstract from hard statistical questions that pertain to model specification and estimation. The second problem is that of economic decision-makers or investors who must forecast the future to make sophisticated investment decisions. Should we put econometricians and economic agents on comparable footing, or should we endow economic agents with much more refined statistical knowledge?

From an econometric standpoint, the outcome of the rational expectations approach is the availability of extra information about the underlying economic model. This information is reflected in an extensive set of cross-equation restrictions. These restrictions allow an econometrician to extract more precise information about parameters or to refine the specification of exogenous processes for the model builder. To understand the nature of these restrictions, consider a dynamic model in which economic agents must make investment decisions in physical, human or financial capital. The decision to invest is forward-looking because an investment made today has ramifications for the future capital stock. The forward-looking nature of investment induces decision makers to make predictions or forecasts as part of their current period choice of investment. The forward-looking perspective affects equilibrium outcomes including market valuations of capital assets. Rational expectations econometrics presumes that agents know the probabilities determining exogenous shocks as they formulate their choices. This translates to an extensive set of cross-equation restrictions that can be exploited to aid identification and inference.

The cross-equation restrictions broadly conceived are a powerful tool, but to what extent should we as applied researchers rely on it? As applied time series econometricians, we

routinely confront challenging problems in model specification. How do we model stochastic dynamics in the short and long run? What variables are the best forecasters? How do we select among competing models?

A heuristic defense for rational expectations appeals to a Law of Large Numbers and gives agents a wealth of data. This allows, at least as an approximation, for us the model builders to presume investor knowledge of a probability model and its parameters. But statistical inference, estimation and learning can be difficult in practice. In actual decision making, we may be required to learn about moving targets, to make parametric inferences, to compare model performance, or to gauge the importance of long-run components of uncertainty. As the statistical problem that agents confront in our model is made complex, rational expectations' presumed confidence in their knowledge of the probability specification becomes more tenuous. This leads me to ask: (a) how can we burden the investors with some of the specification problems that challenge the econometrician, and (b) when would doing so have important quantitative implications? I confront these questions formally by exploring tools that quantify when learning problems are hard, by examining the Bayesian solution to such problems and by speculating on alternative approaches.

In this essay I use the literature that links macroeconomics and asset pricing as a laboratory for examining the role of expectations and learning. The linkage of macroeconomics and finance is a natural choice for study. Even with a rich array of security markets, macroeconomic risks cannot be diversified away (averaged out across investors) and hence are reflected in equilibrium asset prices. Exposure to such risks must be rewarded by the marketplace. By studying asset pricing, we as model-builders specify the forward-looking beliefs of investors and how they cope with risk and uncertainty. Prior to developing asset pricing applications, we consider some stylized statistical decision and inferential problems that turn out to be informative.

I ask five questions that are pertinent to modeling the linkages between asset pricing and macroeconomics:

1. When is estimation difficult?

2. What are the consequences for the econometrician?

3. What are the consequences for economic agents and for equilibrium outcomes?

4. What are the real time consequences of learning?

5. How is learning altered when decision makers admit that the models are misspecified or simplified?

By answering these questions, we will see how statistical ambiguity alters the predicted risk-return relation, and we will see when learning induces model uncertainty premia that are large when macroeconomic growth is sluggish.

1 Rational Expectations and Econometrics

The cross-equation restrictions are the novel component to rational expectations econometrics. They are derived by assuming investor knowledge of parameters and solving for equilibrium decision rules and prices. I consider two examples of such restrictions from the asset pricing literature, and review some estimation methods designed for estimating models subject to such restrictions. One example is the equilibrium wealth-consumption ratio and the other is a depiction of risk prices.

1.1 Cross-equation restrictions

Consider an environment in which equilibrium consumption evolves as:

$$
\begin{aligned}
c_{t+1} - c_t &= \mu_c + \alpha \cdot z_t + \sigma_c \cdot u_{t+1} \\
z_{t+1} &= A z_t + \sigma_z \cdot u_{t+1},
\end{aligned}
\tag{1}
$$

where c_t is the logarithm of consumption, $\{u_t\}$ is an iid sequence of normally distributed random vectors with mean zero and covariance matrix I and $\{z_t\}$ is process used to forecast consumption growth rates. I take equation (1) as the equilibrium law of motion for consumption.

Following Kreps and Porteus (1978) and Epstein and Zin (1989), I use a model of investor preferences in which the intertemporal composition of risk matters. I will have more to say about such preferences subsequently. As emphasized by Epstein and Zin (1989), such preferences give a convenient way to separate risk and intertemporal substitution. Campbell (1996) and others have used log linear models with such investor preferences to study cross-sectional returns.

1.1.1 Wealth-consumption ratio

Let ρ be the inverse of the intertemporal elasticity of substitution and β be the subjective discount factor. Approximate (around $\rho = 1$):

$$w_t - c_t \approx -\log(1 - \beta) + (1 - \rho)\left[\beta\alpha'(I - \beta A)^{-1}z_t + \mu_v\right]. \tag{2}$$

where w_t is log wealth. The constant term μ_v includes a risk adjustment. A key part of this relation is the solution to a prediction problem:

$$E\left[\sum_{j=1}^{\infty}\beta^j(c_{t+j} - c_{t+j-1} - \mu_c)|z_t\right] = \beta\alpha'(I - \beta A)^{-1}z_t.$$

Formula (2) uses the fact that the preferences I consider are represented with an aggregator that is homogeneous of degree one. As a consequence, Euler's theorem gives a simple relation between the shadow value of the consumption process and the continuation value for that process. This shadow value includes the corresponding risk adjustments. The intertemporal budget constraint says that wealth should equal the value of the consumption process. The formula follows by taking a derivative with respect to ρ.[1]

The restriction across equations (1) and (2) is exemplary of the type of restrictions that typically occur in linear rational expectations models. The matrix A that governs the dynamics of the $\{z_t\}$ process also shows up in the formula for the wealth-consumption ratio, and this is the *cross equation restriction*. Very similar formulas emerge in models of money demand (Saracoglu and Sargent (1978)), quadratic adjustment cost models (Hansen and Sargent (1980)) and in log-linear approximations of present-value models (Campbell and Shiller (1988)).

1.1.2 Shadow risk prices

Assume a unitary elasticity of substitution and a recursive utility risk parameter γ and a discount factor β and the same consumption dynamics. Consider the price of the one-period exposure to the shock vector u_{t+1}. Following the convention in finance, let the price be quoted in terms of the mean reward for being exposed to uncertainty. For Kreps and Porteus (1978) preferences the intertemporal composition of risk matters, and as a

---

[1]See Hansen, Heaton, Lee, and Roussanov (2006a) for a derivation and see Campbell and Shiller (1988) and Restoy and Weil (1998) for closely related log-linear approximations.

consequence the consumption dynamics are reflected in the equilibrium prices, including the one-period risk prices. This linkage has been a focal point of work by Bansal and Yaron (2004) and others. Specifically, the one period price vector is:

$$p = \sigma_c + (\gamma - 1)\left[\sigma_c + \beta\alpha'(I - \beta A)^{-1}\sigma_z\right].$$

Later I will add more detail about the construction of such prices. For now, I simply observe that while this price vector is independent of the state vector $\{z_t\}$, it depends on the vectors σ_c and σ_z along with the A matrix. Again we have cross equation restrictions, but now the coefficients that govern variability also come into play.

Pricing a claim to the next period shock is only one of many prices needed to price a cash flow or a hypothetical claim to future consumption. Indeed risk prices can be computed for all horizons. Moreover, as shown by Hansen, Heaton, and Li (2006b) for log linear models like this one, and more generally by Hansen and Scheinkman (2006), the limit prices are also well defined. In this example the limit price is:

$$\left[\sigma_c + \alpha'(I - A)^{-1}\sigma_z\right] + \left[\beta(\gamma - 1)\alpha'(I - \beta A)^{-1}\sigma_z\right].$$

Cross-equation restrictions again link the consumption dynamics and the risk prices.

For these asset pricing calculations and for some that follow, it is pedagogically easiest to view (1) as the outcome of an endowment economy, as in Lucas (1978). There is a simple production economy interpretation, however. Consider a so-called Ak production economy where output is a linear function of capital and a technology shock. Since consumers have unitary elasticity of intertemporal substitution (logarithmic utility period utility function), it is well known that the wealth-consumption ratio should be constant. The first-difference in consumption reveals the logarithm of the technology shock. The process $\{z_t\}$ is a predictor of the growth rate in the technology. Of course this is a special outcome of this model, driven in part by the unitary elasticity assumption. The setup abstracts from issues related to labor supply, adjustment costs and other potentially important macroeconomic ingredients, but it gives pedagogical simplicity that we will put to good use.[2] In summary, under the simple production-economy interpretation, our exogenous specification of a consumption-endowment process becomes a statement about the technology shock process.

---

[2]Tallarini (2000) considers a production counterpart with labor supply, but without the extra dependence in the growth rate of technology shock and without adjustment costs.

In computing the equilibrium outcomes in both examples, I have appealed to rational expectations by endowing agents with knowledge of parameters. A rational expectations econometrician imposes this knowledge on the part of agents when constructing likelihood functions, but necessarily confronts statistical uncertainty when conducting empirical investigations. Economic agents have a precision that is absent for the econometrician. Whether this distinction is important or not will depend on application, but I will suggest some ways to assess this. Prior to considering such questions, I describe some previous econometric developments that gave economic agents more information in addition to knowledge of parameters that generate underlying stochastic processes.

1.2   Econometrics and limited information

Initial contributions to rational expectations econometrics devised methods that permitted economic agents to observe more data than an econometrician used in an empirical investigation. To understand how such methods work, consider again the implied model of the wealth consumption ratio and ask what happens if the econometrician omits information by omitting components of z_t. Let $\mathcal{H}_t$ denote the history up to date t of data used by the econometrician. Rewrite the representation of the wealth-consumption ratio as:

$$w_t - c_t \approx -\log(1-\beta) + (1-\rho)\left(E\left[\sum_{j=1}^{\infty}\beta^j(c_{t+j}-c_{t+j-1}-\mu_c)|\mathcal{H}_t\right] + \mu_v\right) + e_t.$$

The "error" term e_t captures omitted information. Given that the econometrician solves the prediction problem correctly based on his more limited information set, the term e_t satisfies:

$$E\left[e_t|\mathcal{H}_t\right] = 0$$

and this property implies orthogonality conditions that are exploitable in econometric estimation. Econometric relations often have other unobservable components or measurement errors that give additional components to an error term. Alternative econometric methods were developed for handling estimation in which information available to economic agents is omitted by an econometrician (see Shiller (1972), Hansen and Sargent (1980), Hansen (1982), Cumby, Huizinga, and Obstfeld (1983) and Hayashi and Sims (1983)). A reduced-information counterpart to the rational expectations cross-equation restrictions are present in such estimation.

When the only source of an "error term" is omitted information, then there is another

possible approach. The wealth-consumption ratio may be used to reveal to the econometrician an additional component of the information available to economic agents. See for example Hansen, Roberds, and Sargent (1991) and Hansen and Sargent (1991). This is the econometricians' counterpart to the literature on rational expectations with private information in which prices reveal information to economic agents.

There is related literature on estimating and testing asset pricing restrictions. Asset pricing implications are often represented conveniently as conditional moment restrictions where the conditioning information set is that of economic agents. By applying the Law of Iterated Expectations, an econometrician can in effect use a potentially smaller information set in empirical investigation (see Hansen and Singleton (1982), Hansen and Richard (1987), and others.)

All of these methods exploit the potential information advantage of investors in deducing testable restrictions. The methods work if the information that is omitted can be averaged out over time. These methods lose their reliability, however, when omitted information has has a very low frequency or time invariant component as in the case of infrequent regime shifts.

While this literature is aimed at giving economic agents more information than an econometrician along with knowledge of parameters, in what follows I will explore ways to remove some of this disparity and I will illustrate some tools from statistics that are valuable in quantifying when model selection is difficult.

2   Statistical Precision

Statistical inference is at the core of decision making under uncertainty. According to statistical decision theory, enlightened choices are those based on the data that have been observed. When imposing rational expectations, a researcher must decide with what prior information to endow the decision maker. This specification could have trivial consequences, or it could have consequences of central interest. In this section, I consider a measure of statistical closeness that will be used later in this paper. This measure helps quantify statistical challenges for econometricians as well as economic agents.

Suppose there is some initial uncertainty about the model. This could come from two sources: the econometrician not knowing the model (this is a well known phenomenon in rational expectations econometrics) or the agents themselves not knowing it. Past observations should be informative in model selection for either the econometrician or economic

agent. Bayesian decision theory offers a tractable way to proceed. It gives us an excellent benchmark and starting point for understanding when learning problems are hard.

In a Markov setting, a decision maker observes states or signals, conditioning actions on these observations. Models are statistically close if they are hard to tell apart given an observed history. With a richer history, *i.e.* more data, a decision maker can distinguish between competing models more easily. Rational expectations as an approximation conceives of a limit that is used to justify private agents' commitment to one model. When is this a good approximation? A statistical investigation initiated by Chernoff (1952) gives a way to measure how close probability models are, one to another. It quantifies when statistical discrimination is hard, and what in particular makes learning challenging.

Suppose there is a large data set available that is used prior to a decision to commit to one of two models, say model a or model b. Consider an idealized or simplified decision problem in which one of these models is fully embraced given this historical record without challenge. By a model I mean a full probabilistic specification of a vector of observations Y. Each model provides an alternative probability specification for the data. Thus a model implies a likelihood function, whose logarithms we denote by $\ell(Y|m = \mathsf{a})$ and $\ell(Y|m = \mathsf{b})$ respectively where m is used to denote the model. The difference in these log-likelihoods summarizes the statistical information that is available to tell one model from another given data, but more information is required to determine the threshold for such a decision. For instance, Bayesian and mini-max model selection lead us to a decision rule of the form: choose model a if

$$\ell(Y|m = \mathsf{a}) - \ell(Y|m = \mathsf{b}) \geq \mathbf{d}$$

where $\mathbf{d}$ is some threshold value. What determines the threshold value $\mathbf{d}$? Two things: the losses associated with selecting the wrong model and the prior probabilities. Under symmetric losses and equal prior probabilities for each model the threshold $\mathbf{d}$ is zero. Under symmetric losses, the mini-max solution is to choose $\mathbf{d}$ so that the probability of making a mistake when model a is true is the same as the probability of making a mistake when model b is true. Other choices of loss functions or priors result in other choices of $\mathbf{d}$. As samples become more informative, the mistake probabilities converge to zero either under non-degenerate Bayesian priors or under the mini-max solution.

Limiting arguments can be informative. After all, rational expectations is itself motivated by a limiting calculation, the limit of an infinite number of past observations in which the unknown model is fully revealed. Chernoff's method suggests a refinement of this by asking what happens to mistake probabilities as the sample size of *signals* increases. Cher-

noff studies this question when the data generation is iid, but there are extensions designed to accommodate temporal dependence in Markov environments (see for example Newman and Stuck (1979)). Interestingly, the mistake probabilities eventually decay at a common geometric rate. The decay rate is independent of the precise choice of priors and it is the same for the mini-max solution. I call this rate the Chernoff rate and denote it by ρ.[3]

In an iid environment, Chernoff's analysis leads to the study of the following entity. Let f_a be one probability density and f_b another, both of which are absolutely continuous with respect to a measure η. This absolute continuity is pertinent so that we may form likelihood functions that can be compared. The Chernoff rate for iid data is:

$$\rho = -\log \inf_{0 \leq \alpha \leq 1} E\left(\exp\left[\alpha\ell(Y_i|m = \mathsf{b}) - \alpha\ell(Y_i|m = \mathsf{a})\right] |m = \mathsf{a}\right).$$

This formula is symmetric in the role of the models, as can be verified by interchanging the roles of the two models throughout and by replacing α by $1 - \alpha$. The Chernoff rate is justified by constructing convenient bounds of indicator functions with exponential functions.[4] Chernoff (1952)'s elegant analysis helped to initiate an applied mathematics literature on the theory of large deviations.

The following example is simple but revealing, nevertheless.

Example 2.1. *Suppose that x_t is iid normal. Under model* a *the mean is μ_a and under model* b *the model is μ_b. For both models the covariance matrix is Σ. In addition suppose that model* a *is selected over model* b *if the log-likelihood exceeds a threshold. This selection criterion leads us to compute the difference in the log likelihood:*

$$-\frac{1}{2}\sum_{t=1}^{T}(x_t - \mu_a)'\Sigma^{-1}(x_t - \mu_a) + \frac{1}{2}\sum_{t=1}^{T}(x_t - \mu_b)'\Sigma^{-1}(x_t - \mu_b) =$$
$$-\sum_{t=1}^{T}(x_t)'\Sigma^{-1}(\mu_b - \mu_a) + \frac{T}{2}(\mu_b)'\Sigma^{-1}\mu_b - \frac{T}{2}(\mu_a)'\Sigma^{-1}\mu_a.$$

Notice that the random variable in the second equality is normally distributed under each model. Under model a *the distribution is normal with mean:*

$$\frac{T}{2}\left[-2(\mu_a)'\Sigma^{-1}(\mu_b - \mu_a) + (\mu_b)'\Sigma^{-1}\mu_b - (\mu_a)'\Sigma^{-1}\mu_a\right] = \frac{T}{2}(\mu_a - \mu_b)'\Sigma^{-1}(\mu_a - \mu_b)$$

---

[3]It is often called Chernoff entropy in the statistics literature.

[4]While it is the use of relative likelihood functions that links this optimal statistical decision theory, Chernoff (1952) also explores discrimination based on other *ad hoc* statistics.

and variance equal to twice this number. Under model **b** the mean is the negative of this quantity and the variance remains the same. Thus the detection error probabilities are representable as probabilities that normally distributed random variables exceeds a threshold.

In this simple example the Chernoff rate is:

$$\rho = \frac{1}{8} \left[(\mu_a - \mu_b)' \Sigma^{-1} (\mu_a - \mu_b) \right].$$

This can be inferred directly from properties of the cumulative normal distribution, although the Chernoff (1952) analysis is much more generally applicable. The logarithm of the average probability of making a mistake converges to zero at a rate ρ given by this formula. This representation captures in a formal sense the simple idea that when the population means are close together, they are very hard to distinguish statistically. In this case, the resulting model classification error probabilities converge to zero very slowly, and conversely when the means are far apart.

While the simplicity of this example is revealing, the absence of temporal dependence and nonlinearity is limiting. I will explore a dynamic specification next.

Example 2.2. *Following Hansen and Sargent (2006a) consider two models of consumption: one with a long-run risk component and one without. Model* **a** *is a special case of the consumption dynamics given in (1) and is motivated by the analysis in Bansal and Yaron (2004):*

$$
\begin{aligned}
c_{t+1} - c_t &= .0056 + z_t + .0054 u_{1,t+1} \\
z_{t+1} &= .98 z_t + .00047 u_{2,t+1},
\end{aligned}
\tag{3}
$$

and model **b** *has the same form but with $z_t = 0$ implying that consumption growth rates are i.i.d.*[5]

Are the models **a** and **b** easy to distinguish? The mistake probabilities and their logarithms are given in Figures 1 and 2. These figures quantify the notion that the two models are *close* using an extension to Chernoff (1952)'s calculations. For both models the statistician is presumed not to know the population mean and for model **a** the statistician does

---

[5]The mean growth rate .0056 is the sample mean for post war consumption growth and coefficient on .0054 on $u_{1,t+1}$ is the sample standard deviation. In some of my calculations using continuous-time approximations, simplicity is achieved by assuming a common value for this coefficient for models with and without consumption predictability. The parameter value .0047 is the mode of a very flat likelihood function constructed by fixing the two volatility parameters and the autoregressive parameter for $\{z_t\}$. The data and the likelihood function construction are the same as in Hansen and Sargent (2006a).

Figure 1: **Mistake Probabilities**
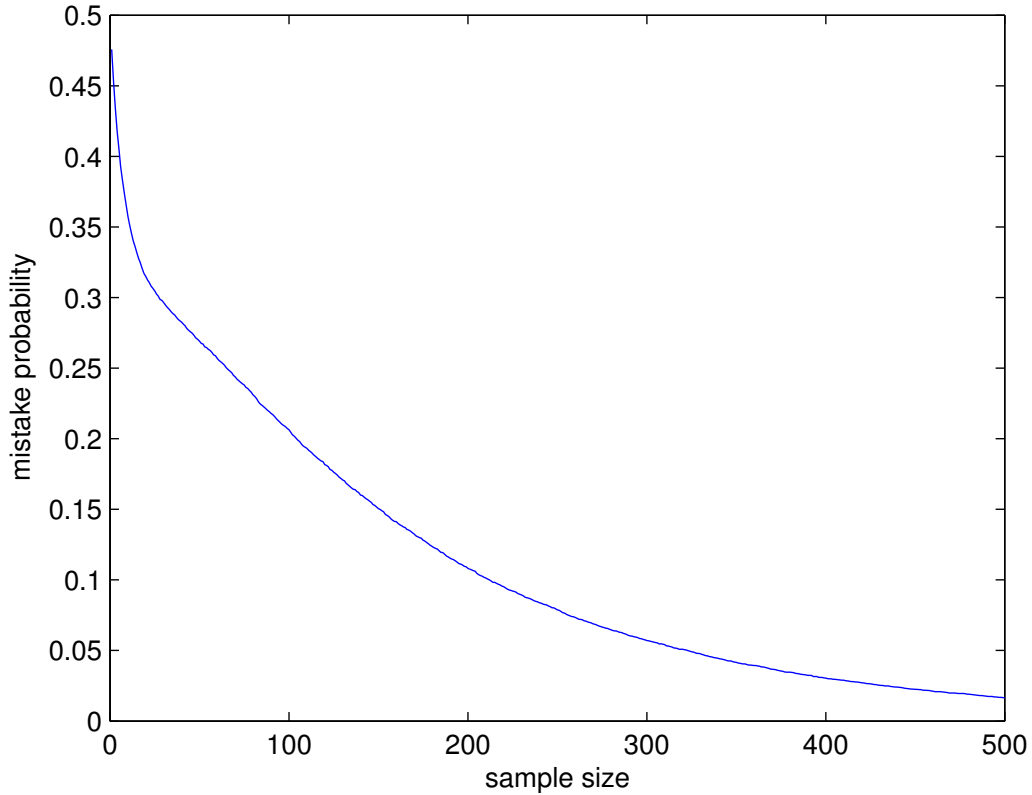
Notes: This figure displays the probability of making a mistake as a function of sample size when choosing between the predictable consumption growth rate model and the i.i.d. model for consumption growth. The probabilities assume a prior probability of one-half for each model. The mistake probabilities are essentially the same if mini-max approach is used in which the thresholds are chosen to equate the model-dependent mistake probabilities. The curve was computed using Monte Carlo simulation. For the predictable consumption growth model, the state $\{z_t\}$ is unobservable and initialized in its stochastic steady state. For both models the prior mean for μ_c is .0056 and the prior standard deviation is .0014.

Figure 2: **Logarithm of Mistake Probabilities**
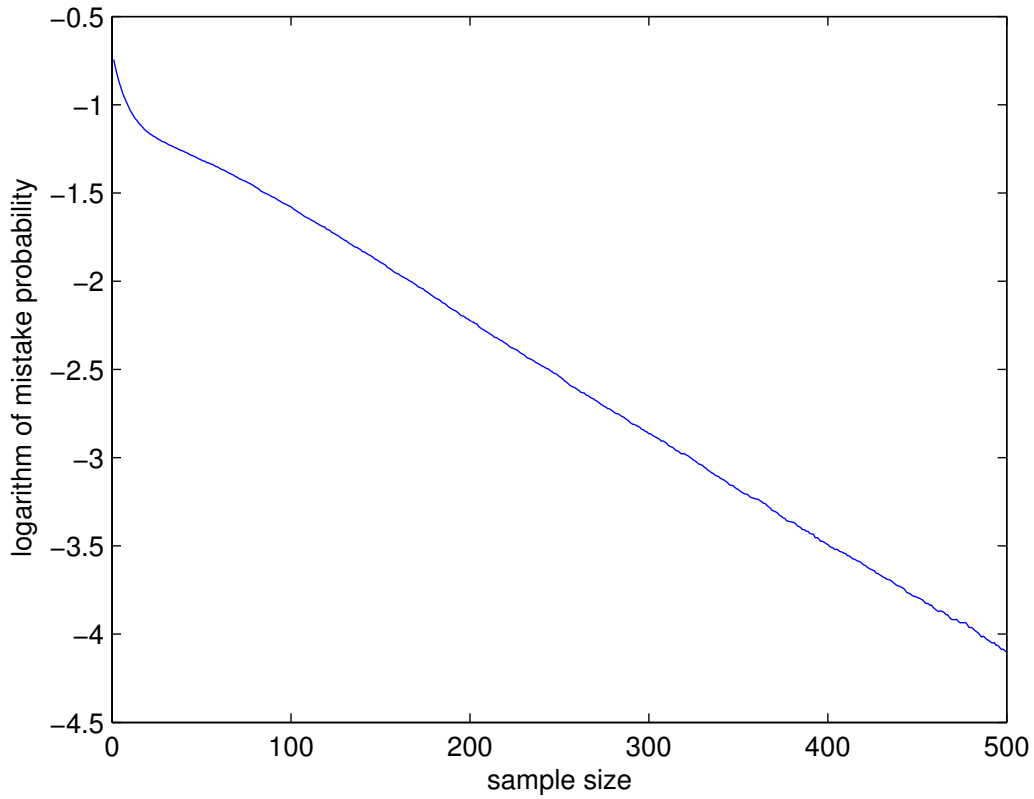
Note: This figure displays the logarithm of the probability of making a mistake as a function of sample size when choosing between the predictable consumption growth model and the i.i.d. model for consumption growth. This curve is the logarithm of the curve in Figure 1.

not know the hidden state. All other parameters are known, arguably simplifying the task of a decision maker. Data on consumption growth rates are used when attempting to the tell the models apart.

From Figure 1 we see that even with a sample size of one hundred (say twenty five years) there is more than a twenty percent chance of making a mistake. Increasing the sample size to two hundred reduces the probability to about ten percent. By sample size five hundred a decision maker can confidently determine the correct model. Taking logarithms, in Figure 2, the growth rate analyzed by Chernoff (1952) and Newman and Stuck (1979) becomes evident. After an initial period of more rapid learning, the logarithm of the probabilities decay approximately linearly. The limiting slope is the Chernoff rate. This is an example in which model selection is difficult for an econometrician, and it is arguably problematic to assume that investors inside a rational expectations model solved it *ex ante*.

Arguably, sophisticated investors know more and process more information. Perhaps this is sufficient for confidence to emerge. There may be other information or other past signals used by economic agents in their decision making. Our simplistic one signal model may dramatically understate prior information. To the contrary, however, the available past history may be limited. For instance, endowing investors with full confidence in model a applied to post war data could be misguided, given the previous era was characterized by higher consumption volatility, two world wars and a depression.

3   Risk Prices and Statistical Ambiguity

In this section I will show that there is an intriguing link between the statistical detection problem we have just described and what is known as a risk price vector in the finance literature. First, I elaborate on the notion of a risk price vector by borrowing some familiar results, and then I develop a link between the Chernoff rate from statistics and the maximal Sharpe ratio. With this link I quantify sensitivity of the measured trade-off between risk and return to small statistical changes in the inputs.

3.1   A Digression on Risk Prices

Risk prices are the compensation for a given risk exposure. They are expressed conveniently as the required mean rewards for confronting the risk. Such prices are the core ingredients in the construction of mean-standard deviation frontiers and are valuable for summarizing

asset pricing implications.

Consider an n-dimensional random vector of the form: $\mu + \Lambda u$ where u is a normally distributed random vector with mean zero and covariance matrix I. The matrix Λ determines the risk exposure to be priced. This random vector has mean μ and covariance matrix $\Sigma = \Lambda\Lambda'$. I price risks that are lognormal and constructed as a function of this random vector:

$$\exp\left(\omega \cdot \mu + \omega'\Lambda u - \frac{1}{2}\omega'\Sigma\omega\right)$$

for alternative choices of the n-dimensional vector ω. The quadratic form in ω is subtracted so that this risk has mean with a logarithm given by $\omega \cdot \mu$.

Let $\exp(r^f)$ be the risk free return. The logarithm of the prices can often be represented as:

$$\log P(\omega) = \omega \cdot \mu - r^f - \omega'\Lambda p$$

for some n-dimensional vector p, where the vector p contains what are typically called the risk prices.

Suppose that the matrix Λ is restricted so that whenever ω is a coordinate vector, a vector with zeros except for one entry which instead contains a one, the risk has a unit price $P(\omega)$ or a zero logarithm of a price. Such an asset payoff is a gross return. Moreover, the payoff associated with any choice of ω with coordinates that sum to one, $i.e.$ $\omega \cdot 1_n = 1$, is also a gross return and hence has a price with logarithms that is zero. Thus, in logarithms the excess return over the risk free return is:

$$\omega \cdot \mu - r^f = \omega'\Lambda p$$

for any ω such that $\omega \cdot 1_n = 1$. The vector p prices the exposure to shock u and is the *risk price vector*. It gives the compensation for risk exposure on the part of investors in terms of logarithms of means.

Such formulas generalize to continuous time economies with Brownian motion risk. The risk prices given in Section 1.1.2 have this form where u is a shock vector at a future date. While the risk prices in that example are constant over time, in Section 7 I will give examples where they vary over time.

3.2 Sharpe Ratios

The familiar Sharpe ratio (Sharpe (1964)) is the ratio of an excess return to its volatility. I consider the logarithm counterpart and maximize by choice of ω:

$$
\begin{aligned}
\max_{\omega, \omega \cdot 1_n = 1} \frac{\omega \cdot \mu - r^f}{\sqrt{\omega' \Sigma \omega}} &= \max_{\omega} \frac{\omega' \Lambda p}{\sqrt{\omega' \Sigma \omega}} \\
&= |p| \\
&= \left[(\mu - 1_n r^f)' \Sigma^{-1} (\mu - 1_n r^f) \right]^{1/2}.
\end{aligned}
$$

The solution measures how steep the risk-return tradeoff is, but it also reveals how large the price vector p should be. A steeper slope of the mean-standard deviation frontier for asset returns imposes a sharper lower bound on $|p|$.

Both risk prices and maximal Sharpe ratios are of interest as diagnostics for asset pricing models. Risk prices give a direct implication when they can be measured accurately, but a weaker challenge is to compare $|p|$ from a model to the empirical solution to (4) for a limited number of assets used in an empirical analysis. Omitting assets will still give a lower bound on $|p|$. Moreover, there are direct extensions that do not require the existence of a risk-free rate and are not premised on log-normality (*e.g.* see Shiller (1982) and Hansen and Jagannathan (1991)). Omitting conditioning information has a well known distortion characterized by Hansen and Richard (1987).[6]

3.3 Statistical Ambiguity

Even if all pertinent risks can be measured by an econometrician, the mean μ is not revealed perfectly to an econometrician or perhaps even to investors. Both perspectives are of interest. I now suggest an approach and answer to the question: Can a small amount of statistical ambiguity explain part of the asset pricing anomalies? Part of what might be attributed to a large risk price p is perhaps small statistical change in the underlying probability model.

Suppose statistical ambiguity leads us to consider an alternative mean μ^*. The change

---

[6]Much has been made of the equity premium puzzle in macroeconomics including, in particular, Mehra and Prescott (1985). For our purposes it is better to explore a more flexible characterization of return heterogeneity as described here. Particular assets with "special" returns can be easily omitted from an empirical analysis. While Treasury bills may contain an additional liquidity premia because of their role as close cash substitutes, an econometrician can compute the maximal Sharpe ratio from other equity returns and alternative risk free benchmarks.

$\mu^* - \mu$ alters the mean-standard deviation tradeoff. Substitute this change into the maximal Sharpe ratio:

$$\left[(\mu^* - \mu + \mu - 1_n r^f)'\Sigma^{-1}(\mu^* - \mu + \mu - 1_n r^f)\right]^{1/2}.$$

Using the Triangle Inequality,

$$\begin{aligned}
\left[(\mu^* - \mu)'\Sigma^{-1}(\mu^* - \mu)\right]^{1/2} &- \left[(\mu - 1_n r^f)'\Sigma^{-1}(\mu - 1_n r^f)\right]^{1/2} \\
&\leq \left[(\mu^* - 1_n r^f)'\Sigma^{-1}(\mu^* - 1_n r^f)\right]^{1/2}.
\end{aligned}$$

This inequality shows that if

$$\left[(\mu^* - \mu)'\Sigma^{-1}(\mu^* - \mu)\right] \tag{4}$$

is sizable and offsets the initial Sharpe ratio, then there is a sizable movement in the Sharpe ratio.

More can be said if I give myself the flexibility to choose the direction of the change. Suppose that I maximize the new Sharpe ratio by choice of μ^* subject to a constraint on (4). With this optimization, the magnitude of the constraint gives the movement in the Sharpe ratio.

Chernoff's formula tells us when (4) can be economically meaningful but statistically small. Squaring (4) and dividing by eight gives the Chernoff rate. This gives a formal link between the statistical discrimination of alternative models and what are referred to *risk prices*. The link between the Chernoff rate and the maximal Sharpe ratio gives an easily quantifiable role for statistical ambiguity either on the part of an econometrician or on the part of investors in the interpretation of the risk-return tradeoff.

Could the maximal Sharpe ratio be equivalent to placing alternative models on the table that are hard to discriminate statistically? Maybe it is too much to ask to have models of risk premia that assume investor knowledge of parameters bear the full brunt of explaining large Sharpe ratios. Statistical uncertainty might well account for a substantial portion of this ratio.

Consider a Chernoff rate of 1% per annum or .25% per quarter. Multiply by eight and take the square root. This gives a increase of about .14 in the maximum Sharpe ratio. Alternatively, a Chernoff rate of .5% per annum gives an increase of 0.1 in the maximum Sharpe ratio. These are sizable movements in the quarterly Sharpe ratio accounting for somewhere between a third and a half of typical empirical measurements.

There are two alternative perspectives on this link. First is measurement uncertainty

faced by an econometrician even when economic agents know the relevant parameters. For instance the risk price model of Section 1.1.2 may be correct, but the econometrician has imperfect measurements. While the Chernoff calculation is suggestive, there are well known ways to account for statistical sampling errors for Sharpe ratios in more flexible ways including, for example, Gibbons, Ross, and Shanken (1989). Alternatively, investors themselves may face this ambiguity which may alter the predicted value of p and hence $|p|$ coming from the economic model. I will have more to say about this in the next section.

The particular formula for the Chernoff rate was produced under very special assumptions, much too special for more serious quantitative work. Means and variances are dependent on conditioning information. Normal distributions may be poor approximations. Anderson, Hansen, and Sargent (2003) build on the work of Newman and Stuck (1979) to develop this link more fully. Under more general circumstances, a distinction must be made between local discrimination rates and global discrimination rates. In continuous time models with a Brownian motion information structure, the local discrimination rate has the same representation based on a normal distributions with common covariances, but this rate can be state dependent. Thus, the link between Sharpe ratios and the local Chernoff rate applies to an important class of asset pricing models. The limiting decay rate is a global rate that averages the local rate in a particular sense.

4 Statistical Challenges

In this section, I revisit model a (see equation (3)) of example 2.2 from two perspectives. I consider results first, from the vantage point of an econometrician and second, from that of investors in an equilibrium valuation model.

4.1 The Struggling Econometrician

An econometrician uses post-war data to estimate parameters that are imputed to investors. I present the statistical evidence available to the econometrician in estimating the model. I construct posterior distributions from alternative priors and focus on two parameters in particular: the autoregressive parameter for the state variable process $\{z_t\}$ and the mean growth rate in consumption. For simplicity, and to anticipate some of the calculations that follow, I fixed the coefficient on $u_{1,t+1}$. I report priors that are not informative (loose priors) and priors that are informative (tight priors). It turns out that there is very little

sample information about the coefficient on $u_{2,t+1}$. As a consequence, I used a informative prior for this coefficient in generating the "loose prior" results, and I fixed this coefficient at .00047 when generating the "tight prior" results.

I depict the priors and posteriors in Figure 3. There is very weak sample information about the autoregressive parameter, and priors are potentially important. There is some evidence favoring coefficients close to unity. Under our rational expectations solutions we took the parameter to be .98, in large part because of our interest in a model with a low frequency component.[7] The posterior distribution for the mean for consumption growth is less sensitive to priors. Without exploiting cross equation restrictions, there is only very weak statistical evidence about the process $\{z_t\}$ which is hidden from the econometrician. Imposing the cross-equation restrictions begs the question of where investors come up with knowledge of the parameters that govern this process.

4.2 The Struggling Investors

The rational expectations solution of imposing parameter values may be too extreme, but for this model it is also problematic to use loose priors. Geweke (2001) and Weitzman (2007) show dramatic asset return sensitivity to such priors in models without consumption predictability. While loose priors are useful in presenting statistical evidence, it is less clear that we should embrace them in models of investor behavior. How to specify meaningful priors for investors becomes an important specification problem when Bayesian learning is incorporated into a rational expectations asset pricing model and in the extensions that I will consider. Learning will be featured in the next two sections, but before incorporating this extra dimension, I want to re-examine the risk prices derived under rational expectations and suggest an alternative interpretation for one of their components.

In Section 1.1.2 I gave the risk price vector for an economy with predictable consumption. Since investors are endowed with preferences for which the intertemporal composition of risk matters, the presence of consumption predictability alters the prices. Recall the one-period risk price vector is

$$p = \sigma_c + (\gamma - 1) \left[\sigma_c + \beta \alpha' (I - \beta A)^{-1} \sigma_z \right].$$

---

[7]Without this focus one might want to examine other aspects of consumption dynamics for which a richer model could be employed. Hansen, Heaton, and Li (2006b) use corporate earnings as a predictor variable and document a low frequency component using a vector autoregression provided that a cointegration restriction is imposed.

Figure 3: **Prior and Posterior Probabilities**
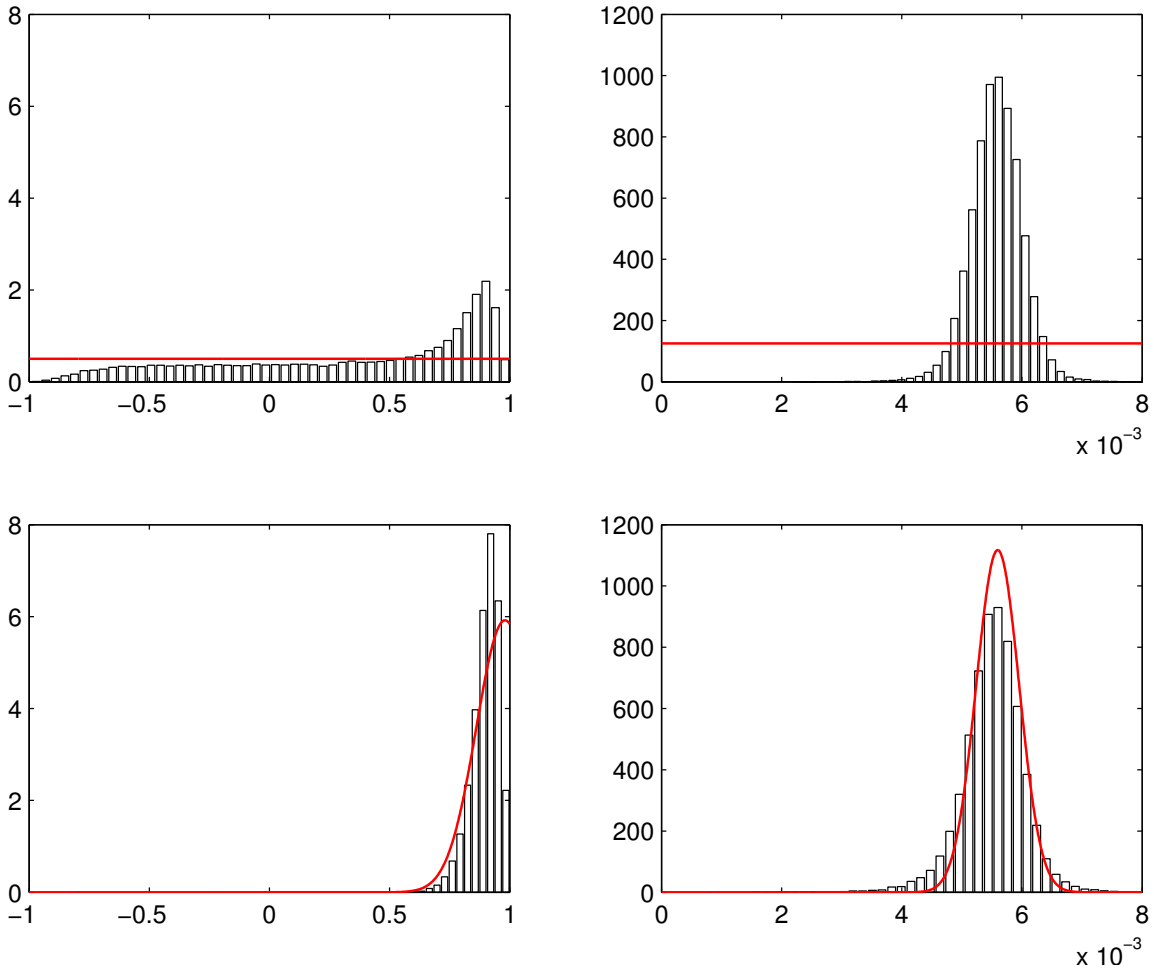
Note: This figure displays the priors (the lines) and the posteriors histograms for two parameters of the model with predictable consumption growth. The left column gives the densities for the autoregressive parameter for the hidden state and the right column the mean growth rate of consumption. The results from the first row were generated using a relatively loose prior including an informative prior on the conditional variance for the hidden state. The prior for the variance is an inverse gamma with shape parameter 10 and scale parameter 1.83×10^{-7}. The implied prior mode for σ_z is .00041. The prior for the AR coefficient is normal conditioned on σ_z with mean 0 and standard deviation $\sigma_z \times 1.41 \times 10^6$ truncated to reside between minus one and one. The prior for μ_c has mean .003 and standard deviation .27. The results from the second row were generated with a informative prior and fixed the conditional standard for the hidden state at .00047. The prior for AR coefficient is normal with mean .98 and standard deviation .12. The prior for μ_c is normal with mean .0056 and standard deviation .00036. The posterior densities were computed using Gibbs sampling with 50,000 draws after ignoring the first five thousand.

19

One way to make risk prices large is to endow investors with large values of the risk aversion parameter γ. While γ is a measure of risk aversion in the recursive utility model, Anderson, Hansen, and Sargent (2003) give a rather different interpretation. They imagine investors treat the model as possibly misspecified and ask what forms of model misspecification investors fear the most. The answer is a mean shift in the shock vector u_{t+1} that is proportional to the final term above

$$\left[\sigma_c + \beta \alpha'(I - \beta A)^{-1}\sigma_z \right].\tag{5}$$

This is deduced from computing the continuation value for the consumption process. Instead of a measure of risk aversion, $\gamma - 1$ is used to quantify an investors' concern about model misspecification.

Is this distortion statistically large? Could investors be tolerating statistical departures of this magnitude because of their concern about model misspecification? Our earlier Chernoff calculations are informative. Even with temporal dependence in the underlying data generating process, the Chernoff discrimination rate is:

$$|\gamma - 1|^2 \frac{|\sigma_c + \beta \alpha'(I - \beta A)^{-1}\sigma_z|^2}{8}.$$

Consider now the parameter values given in first model example 2.2. Then

$$|\gamma - 1|^2 \frac{|\sigma_c + \beta \alpha'(I - \beta A)^{-1}\sigma_z|^2}{8} \approx .000061|\gamma - 1|^2.\tag{6}$$

For instance, when $\gamma = 5$ the implied discrimination rate is just about a half percent per year. This change endows the state variable process $\{z_t\}$ with a mean of $-.002$ and a direct mean decrease in the consumption growth equation of $-.0001$, which is inconsequential. The contribution to $|p|$ measure by the norm of (5) scaled by $\gamma - 1 = 4$ is about .09. While both distortions lower the average growth rate in consumption, only the second one is substantial. Investors make a conservative adjustment to the mean of the shock process $\{u_{2,t}\}$ and hence to the unconditional mean of $\{z_t\}$. This calculation gives a statistical basis for a sizeable model uncertainty premium as a component of p. Similar calculations can be made easily for other values of γ.

While a mean distortion of $-.002$ in the consumption dynamics looks sizable, it is not large relative to sampling uncertainty. The highly persistent process $\{z_t\}$ makes inference

about consumption growth rates difficult.[8] Moreover, my calculation is sensitive to the inputs that are not measured well by an econometrician. Conditioned on .98, the statistical evidence for σ_z is not very sharp. Reducing σ_z by one half only changes the log-likelihood function[9] by .3. Such a change in σ_z reduces the Chernoff rate and the implied mean distortion attributed to the $\{z_t\}$ process by factors in excess of three.

Suppose that investors only use data on aggregate consumption. This presumes a different model for consumption growth rates, but one with the same implied probabilities for the consumption process. This equivalent representation is referred to as the *innovations representation* in the time series literature and is given by:

$$
\begin{aligned}
c_{t+1} - c_t &= .0056 + \bar{z}_t + .0056\bar{u}_{t+1} \\
\bar{z}_{t+1} &= .98\bar{z}_t + .00037\bar{u}_{t+1}
\end{aligned}
$$

where $\{\bar{u}_{t+1}\}$ is a scalar i.i.d. sequence of standard normally distributed random variables. The implied distortions for the consumption growth rate givensay $\gamma = 5$ are very close to those I gave previously, based on observing both consumption growth and its predictor process.

In this subsection I used a link between distorted beliefs and continuation values to reinterpret part of the risk price vector p as reflecting a concern about model misspecification. This is special case of a more general approach called *exponential tilting*, an approach that I will have more to say about in Sections 6 and 7. Investors tilt probabilities, in this case means of shocks, in directions that value functions suggests are most troublesome. While the tilted probabilities in this section are represented as time invariant mean shifts, by considering learning, I will obtain a source of time-variation for the uncertainty premia.

5   Learning

Up until now we have explored econometric concerns and statistical ambiguity without any explicit reference to learning. Our next task is to explore the real time implications of learning on what financial econometricians refer to as risk prices. To explore learning

---

[8]For the persistence and volatility parameters assumed in this model, μ_c is estimated with much less accuracy than that shown in Figure 3. The posteriors reported in this figure assign considerable weight to processes with much less persistence.

[9]As a rough guide, twice the log-likelihood difference is a little more than half the mean of a $\chi^2(1)$ random variable.

in a tractable way, consider what is known in many disciplines as a hidden Markov model (HMM). In what follows we let ξ be a realized value of the signal while s^* denotes the signal, which is a random vector. We make the analogous distinction between realzed values of the state ζ versus the random state vector z. Suppose that the probability density for a signal or observed outcome s^* given a Markov state z is denoted by $f(\cdot|z)$. This density is defined relative to an underlying measure $d\eta(\xi)$ over the space of potential signals $\mathbb{S}$. A realized state is presumed to reside in a space $\mathbb{Z}$ of potential states.

In a HMM the state z is disguised from the decision maker. The vector z could be (a) a discrete indicator of alternative models; (b) an unknown parameter; (c) a hidden state that evolves over time in accordance to a Markov process as in a regime shift model of Wonham (1964), Sclove (1983) and Hamilton (1989). The signal or outcome s^* is observed in the next time period. If z were observed, we would just use f as the density for the next period outcome s^*. Instead inferences must be made about z to deduce the probability distribution for s^*. For simplicity, we consider the case in which learning is passive. That is, actions do not alter the precision of the signals.

5.1 Compound Lottery

To apply recent advances in decision theory, it is advantageous to view the HMM as specifying a compound lottery repeated over time. Suppose for the moment that z is observed. Then for each z, $f(\cdot|z)$ is a lottery over the outcome s^*. When z is not observed, randomness of z makes the probability specification a *compound lottery*. Given a distribution π, we may *reduce* this compound lottery by integrating out over the state space $\mathbb{Z}$:

$$\bar{f}(\xi) = \int f(\xi|\zeta) d\pi(\zeta).$$

This reduction gives a density for s^* that may be used directly in decision-making without knowledge of z. In the applications that interest us, π is a distribution conditioned on a history $\mathcal{H}$ of signals.[10]

5.2 Recursive Implementation

In an environment with repeated signals, the time t distribution, π_t, inherits dependence on calendar time through the past history of signals. Bayes rule tells us how to update this

---

[10]Formally, $\mathcal{H}$ is a sigma algebra of conditioning events generated by current and past signals.

equation in response to a new signal. Repeated applications gives a recursive implementation of Bayes rule.

Consider some special cases:

5.2.1   Case 1: Time Invariant Markov State

Suppose that z is time invariant as in the case of an unknown parameter or an indicator of a model. Let π denote a probability distribution conditioned on a history $\mathcal{H}$, and let π^* denote the updated probability measure given that s^* is observed. Bayes rule gives:

$$\pi^*(d\zeta) = \frac{f(s^*|\zeta)d\pi(\zeta)}{\int f(s^*|\zeta)d\pi(\zeta)}.$$

The signal s^* enters directly into this evolution equation. Applying this formula repeatedly for a sequence of signals generates a sequence of probability distributions $\{\pi_t\}$ for z that reflect the accumulation of information contained in current and past signals.

Since z is time invariant, the constructed state probability distribution $\{\pi_t\}$ is a martingale. Since π_t is a probability distribution, this requires an explanation. If the set of potential states $\mathbb{Z}$ consists of only a finite number of entries, then each of the probabilities is a martingale. More generally, let ϕ be any bounded function of the hidden state z.[11] An example of such a function is the so-called indicator functions that is one on set and zero on its complement. The integral $\int \phi(\zeta)d\pi(\zeta)$ gives the conditional expectation of $\phi(z)$ when $d\pi(\zeta)$ is the conditional distribution for z given the current and past signal history $\mathcal{H}$.

In contrast to π, the distribution π^* incorporates information available in the signal s^*. Then

$$
\begin{aligned}
E\left[\int \phi(\zeta)d\pi^*(\zeta)|\mathcal{H}\right] &= \int \left[\frac{\int f(s^*|\zeta)\phi(\zeta)d\pi(\zeta)}{\int f(s^*|\zeta)d\pi(\zeta)}\right]\left[\int f(s^*|\zeta)\pi(d\zeta)\right]d\eta(s^*) \\
&= \int\int f(s^*|\zeta)\phi(\zeta)d\pi(\zeta)d\eta(s^*) \\
&= \int \phi(\zeta)d\pi(\zeta)
\end{aligned}
\tag{7}
$$

since $\int f(\xi|\zeta)d\eta(\xi) = 1$. This implies the familiar martingale property associated with parameter learning, that the best forecast of $\int \phi(\zeta)\pi_{t+1}(d\zeta)$ given current period information

---

[11]Formally, we also restrict ϕ to be Borel measurable.

is $\int \phi(\zeta)\pi_t(d\zeta)$. Thus given the sequence of probability distributions $\{\pi_t(d\zeta) : t = 0, 1, ...\}$, the sequence random variables $\{\int \phi(\zeta)\pi_t(d\zeta) : t = 0, 1, ...\}$ is a martingale. In fact it is a bounded martingale and it necessarily converges.

By making an invariant z unobservable, we have introduced a strong form of stochastic dependence as reflected by the martingale property. Note, however, that the stochastic structure will become degenerate as the martingale converges. When learning problems are difficult, the convergence will be slow as we have seen in our discussion of Chernoff (1952).

5.2.2   Case 2: Time Varying Markov State

Consider the case in which z is not invariant and its evolution is modeled as a Markov process. The dynamics for this hidden Markov state influence directly the learning dynamics in ways that I will illustrate. Let $T(\zeta^*|\zeta)$ be the transition density of z relative to a measure $\lambda(d\zeta)$ over the hidden states. The measure λ is chosen for convenience depending upon the details of the application. Later I will feature examples in which state space $\mathbb{Z}$ contains a finite set of values and the measure λ just assigns one to each element of this set. Other measures are used when z is continuous.

Our previous calculations extend, except that the updating equation for the z^* posterior distribution π^* must accommodate this evolution. From Bayes rule:

$$
\begin{aligned}
\pi^*(d\zeta^*) &= \frac{\left[\int T\left(\zeta^*|\zeta\right) f\left(s^*|\zeta\right) d\pi\left(\zeta\right)\right] d\lambda(\zeta^*)}{\int \int T(\tilde{\zeta}|\zeta) f(s^*|\zeta) d\lambda(\tilde{\zeta}) d\pi(\zeta)} \\
&= \overline{T}(s^*, \pi)
\end{aligned}
\tag{8}
$$

The distribution π^* evolves from π as a function of the signal s^* in accordance to the function $\overline{T}$. Given the stochastic evolution of the hidden state, we lose the martingale property. The case in which z is invariant, considered previously, is a special case with a degenerate specification of the transition law: $z^* = z$. When the hidden state has a nondegenerate transition law, we lose the martingale property. If the transition law T is stochastically stable (that is, there is unique stationary distribution associated with T), then this asymptotic stability carries over to the evolution of the probability distributions: $\{\pi_t\}$ captured by $\overline{T}$.

5.3 A New Markov Process

It follows from what we have just shown that we can represent this form of learning as a new Markov process. For this new process, the hidden state z is replaced by a distribution over the hidden state. The density for the signal is

$$\bar{f}(\xi|\pi) = \int f(\xi|\zeta)d\pi(\zeta)$$

and π evolves according to $\overline{T}$ given in (8). Thus we may conceive of learning as justifying a Markov process with a "state variable" π.

I derived this learning solution using Bayes rule, but otherwise we did not appeal to a specific decision problem. The hidden state may be hidden to the econometrician or it may be subjective uncertainty in the minds of investors. If the former, its estimation is a problem only for an econometrician. If the latter, both the econometrician and the investor being modeled may aim to integrate it out or reduce the compound lottery. For instance, I could use this learning solution to alter the model of exogenous shock processes such as technology shocks. I simply replace one state variable, z, by another, the distribution π. The recursive solution becomes an input into our rational expectations model with the additional econometric challenge of specifying an initial condition for π, *a priori*.[12] Since π is a distribution, for many state spaces it can be an infinite dimensional state variable; but that is a computational, not a conceptual issue.

Given this change, I may define a rational expectations equilibrium to determine the endogenous state variables such as capital stocks and the endogenous prices. More generally, I could introduce private signals and endogenously determined price signals. This approach to learning is an enrichment of rational expectations to include subjective uncertainty while preserving the essential equilibrium components. The resulting equilibrium model is what Bray and Kreps (1987) call learning within a rational expectations model. After all, Lucas's (1972b) use of rational expectations in a private information economy has agents learning from price signals, so aspects of learning have been central features in rational expectations equilibria from the outset.

There are other ways to introduce learning that push us outside the realm of rational expectations in a more substantial way. For instance, we might pose the learning challenge directly on the price dynamics or the endogenous state variables. This has led to

---

[12]Moreover, since z may be disguised, identification of its dynamics as captured by T may be more challenging.

Bray and Kreps (1987) call this a model of learning about a rational expectations equilibrium. Adaptive control methods or Bayesian methods are applied that fail to impose some of the internal consistency conditions of a rational expectations equilibrium. See Bray (1982), Chen and White (1998), Marcet and Sargent (1989), Sargent (1999) and Evans and Honkapohja (2001) for examples of what has become an impotent literature in macroeconomics. Agents are assumed to apply Bayesian learning methods to misspecified but typically simpler models or they apply adaptive methods that aim to provide a more flexible approximation. The outcome of these misspecificied Bayesian or adaptive learning algorithms are fully embraced as beliefs by the economic agents when making forward-looking decisions. There is typically no acknowledgment of the potential misspecification. The dynamic systems may have limit points, but they may imply weaker consistency requirements than a rational expectations equilibrium.[13] Since this approach to learning does not presume that decision makers fully perceive the uncertainty they confront, the resulting equilibria ignore a potentially important source of uncertainty premia that might show up in prices that clear security markets. The economic agents in such models experience no specification doubts.

5.4   Dynamic Learning in a Regime Shift Model

To illustrate the dynamics of learning, we use a solution first characterized by Wonham (1964) to a filtering problem that economists sometimes refer to as a regime shift model. The model and solution are given most conveniently in continuous time. Consider a signal:

$$ds_t = \kappa \cdot z_t dt + \sigma dB_t$$

where $\{B_t\}$ is standard Brownian motion and $\{z_t\}$ is a hidden state Markov chain with intensity matrix A. The intensity matrix conveniently summarizes the matrix of transition probabilities for the hidden state via the formula $\exp(tA)$ for any positive number t. The realized value of z_t is a coordinate vector. Thus $\kappa \cdot z_t$ selects among the entries in the vector κ in determining the local growth rate for the signal process. This specification is a continuous-time counterpart to the regime switching model of Sclove (1983) and Hamilton (1989). It has been used in asset pricing models by David (1997) and Veronesi (2000). Given this model, we can think of ds_t conditioned on the state z_t as a compound lottery.

---

[13]The weaker equilibrium concept is known as a self-confirming equilibrium. See Sargent (1999) for a discussion.

The Wonham filter gives the solution to *reducing* the compound lotteries while *updating* probabilities based on past data. Since z_t is a coordinate vector, its conditional expectation given the signal history is the vector of hidden state conditional probabilities. As for notation we let $\bar{z}_t = E\left(z_t|\mathcal{H}_t\right)$, which is the vector of hidden state probabilities. Thus, the conditional mean $\bar{z}_t$ contains the vector of state probabilities used to depict π_t. The recursive filtering solution is a stochastic differential equation represented in terms of an alternative standard Brownian motion $\{\bar{B}_t\}$:

$$
\begin{aligned}
ds_t &= \kappa \cdot \bar{z}_t dt + \sigma d\bar{B}_t \\
d\bar{z}_t &= A'\bar{z}_t dt + \Delta(\bar{z})\left(ds_t - \kappa \cdot \bar{z}_t dt\right) \\
\Delta(\bar{z}) &= \frac{1}{\sigma^2}\text{diag}(\bar{z}_t)\left(\kappa - \mathbf{1}_n \kappa \cdot \bar{z}_t\right).
\end{aligned}
$$

The first equation gives the continuous-time counterpart to $\bar{f}$, the density for the signal. The second equation gives the counterpart to $\bar{T}$, the evolution equation for the probabilities. The Brownian motion increment $d\bar{B}_t$ can be inverted from the signal evolution equation.

There are notable features of this solution. First, the matrix A used to model the hidden state dynamics plays a central role in the dynamics for $\bar{z}$. It enters directly into the formula for the local conditional mean, $A'\bar{z}$. Second, the new information contained in the signal evolution is captured by the increment to the Brownian motion $d\bar{B}$, which we express as:

$$
d\bar{B}_t = \kappa \cdot (z_t - \bar{z}_t)dt + dB_t. \tag{9}
$$

This represents the new information encoded in the signal history as a compound lottery. Both Brownian motions B and $\bar{B}$ are standardized (they have unit variance over a time interval of length one). The reduced information does not alter the local accuracy of our forecast of the signal. While this is a special property of continuous-time models with signal noise generated by a Brownian motion, it gives us an informative limiting case. Finally, the vector Δ contains the local (in time) regression coefficient of the hidden state onto the new information in the signal. These coefficients depend on the state probability vector $\bar{z}_t$. When one of the entries of $\bar{z}_t$ is close to unity, the vector Δ is close to zero.

The following example is of some pedagogical interest because it illustrates how varying parameters of this model alters the temporal dependence of the probabilities and the sensitivity of these probabilities to new information.

Example 5.1. *Consider the following two-state example. The two by two intensity matrix*

is parameterized by:

$$A = \begin{bmatrix} -a_1 & a_1 \\ a_2 & -a_2 \end{bmatrix}$$

where $a_1 \geq 0$ and $a_2 \geq 0$. Since probabilities add up to one, it suffices to consider only one of the probabilities, say the probability of being in state one. Substituting from our parameterization of A,

$$d\bar{z}_{1,t} = -(a_1 + a_2)\left(\bar{z}_{1,t} - \frac{a_2}{a_1 + a_2}\right) dt + \bar{z}_{1,t}(1 - \bar{z}_{1,t})\left(\frac{\kappa_1 - \kappa_2}{\sigma}\right) d\bar{B}_t.$$

In this example, the unconditional mean of the probability of being in state one is $a_2/(a_1 + a_2)$. The local volatility of the probability scales with the difference between means relative to the signal volatility. When the difference in the $\kappa's$ is large relative to σ, the probabilities are more responsive to the new information contained in the signals. This responsiveness becomes arbitrarily small if the probability is close to zero or one. If $a_1 = a_2 = 0$, then the probability is a positive martingale. When a_1 and a_2 are both positive, the probability process is asymptotically stationary. Larger values imply more mean reversion in the probabilities.[14]

While I feature the Wonham filter in this essay, there are other well known filtering methods including the Kalman filter, the particle filter and the Zakai equation. There are alternative ways of characterizing the solutions to $\bar{f}$ and $\bar{T}$.

5.5 Real Time Model Detection

I began this essay by considering a model detection problem posed by Chernoff (1952). The stochastic specification of the Wonham filter gave me a way to move across regimes in real time, but it also includes time invariant indicators of models as envisioned by Chernoff. Such indicators are natural limits of low frequency movements in regimes. Learning about low frequencies will be an important component to some of our calculations, and therefore it warrants special consideration.

When time-invariant indicators are included as possibilities, it is no longer fruitful to appeal to a stochastic steady state. While such steady states exist, they are degenerate and the interesting question becomes one of convergence. The rate of convergence is precisely

---

[14]We do not mean to imply that the drift determines the pull of the process towards the center of its distribution. Given that volatility is also state dependent, it also plays a role in pulling the distribution away from the boundaries. When volatility is relatively low, the pull by the drift is more potent.

what Chernoff's analysis allows us to investigate. The inferential problem that is presumed in my application of the Wonham filter includes a model selection problem where the invariant state is a model indicator indexed by an invariant state.

I now use the stochastic structure of the Wonham (1964) filtering model to explore dynamics of model selection.

Example 5.2. *Consider an example with three states. Two states give rise to movements in the growth rate for consumption. Movements between states are random and shift the growth rate in the signal as in example 5.1. The third state is invariant. It cannot be reached from the first two states. Formally, the A matrix is:*

$$A = \begin{bmatrix} -a_1 & a_1 & 0 \\ a_2 & -a_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where $a_1 > 0$ and $a_2 > 0$. There is no possible movement from states one and two to state three or from state three to states one and two. While the third state is invariant, the decision-maker does not know if this third state or regime is the relevant one. Thus, he faces a model selection problem.

Given the existence of a time invariant hidden state, the dynamic extension of the Chernoff (1952) analysis determines the asymptotic discrimination rate between models (states one and two versus state three.) This leads to the study of the asymptotic behavior of the filtering solution when the signal is restricted to spend all of its time in states one and two or when the signal is restricted to spend all of its time in state three. In the former case, the process $\{\bar{z}_{3,t}\}$ will converge to zero eventually at an exponential rate, while in the latter case $\{\bar{z}_{1,t} + \bar{z}_{2,t}\}$ will converge to zero eventually at this same rate.

The local counterpart to Chernoff's discrimination rate is:

$$\frac{\left[\kappa_1 \left(\frac{\bar{z}_{1,t}}{\bar{z}_{1,t} + \bar{z}_{2,t}} \right) + \kappa_2 \left(\frac{\bar{z}_{2,t}}{\bar{z}_{1,t} + \bar{z}_{2,t}} \right) - \kappa_3 \right]^2}{8\sigma^2}.$$

This rate depends on the local mean difference between the two models where the first model is the original two-state model of example 5.1 and the second model has local mean of $\kappa_3 dt$ that is time invariant. Small mean differences across the models relative to the volatility make model discrimination challenging. Since this local rate is time varying, as

29

I argued before, the asymptotic discrimination rate is an average of this local rate with respect to an appropriately defined mixture model. See Newman and Stuck (1979).

In addition to considering learning with time invariant hidden states, I will also explore the implications of recent decision theory that will allow us to feature learning and concerns about model specification but preserve many other useful features of a rational expectations equilibrium.

6   Beliefs and Preferences

Expected utility theory embraces the axiom that compound lotteries should be reduced. If, as suggested previously, we view $f(\xi^*|\zeta)$ and $d\pi(\zeta)$ as a compound or two-step lottery, then the ranking induced by expected utility preferences depends on the *reduced* lottery:

$$\bar{f}(\xi) = \int f(\xi|\zeta)d\pi(\zeta).$$

The integration defines a lottery that does not condition on $z = \zeta$, and compounding is just a way to depict or even restrict lotteries of interest. Similarly, decisions or actions that depend on s^* can be represented as a compound lottery that can be reduced using the density $\bar{f}$. For the example economies that we explore, the use of expected utility theory implies that rational learning has only modest implications for predicted risk premia. This leads me to employ generalizations of this theory that avoid the presumption that compound lotteries should simply be reduced. Kreps and Porteus (1978), Segal (1990) and Klibanoff, Marinacci, and Mukerji (2005) provide alternative decision theories that resist the reduction of compound lotteries. Associated with some of these formulations are alternative beliefs that are tilted in ways that I characterize.

6.1   Irreducible Lotteries

Segal (1990) studies two-stage lotteries and axioms that do not imply reduction. Instead the conditional composition of risk matters. We explore two distinct motivations for why conditioning might matter. First we distinguish the riskiness of s^* conditioned on z from riskiness over the hidden state or time invariant parameter z. Klibanoff, Marinacci, and Mukerji (2005) develop this idea further to distinguish risk or objective uncertainty, captured by the signal density $f(\cdot|z)$, from subjective uncertainty, captured by probability

distribution π over hidden states. They give an axiomatic justification for a convenient representation of preferences.

Let the adjustment for risk conditioned on z be represented by an increasing concave function h:

$$\mathbb{V}(a|z) = h^{-1}\left(\int h[a(\xi)]f(\xi|z)d\eta(\xi)\right) \tag{10}$$

where a is some action or decision expressed as a function of the signal. The h^{-1} transformation is convenient because if the random s^* can be perfectly predicted given z, the right-hand side of (10) gives the state contingent action. Construct a second-stage ranking based on the utility function:

$$\int [g[\mathbb{V}(a|\zeta)]\,d\pi(\zeta) \tag{11}$$

using the strictly increasing concave utility function g. As a special case, if $g = h$

$$\int\int h[a(\xi)]f(\xi|\zeta)d\eta(\xi)d\pi(\zeta) = \int h[a(\xi)]\bar{f}(\xi)d\eta(\xi).$$

Preferences that do not reduce compound lotteries permit h to differ from g. The behavioral responses to the two different forms of risk or uncertainty are allowed to be different. Klibanoff, Marinacci, and Mukerji (2005) defend this as allowing for a smooth version of ambiguity aversion when $g \circ h^{-1}$ is concave.

Following Kreps and Porteus (1978), we may use the same setup to consider a rather different question. Consider two lotteries. One is $a(s^*)$ where z is observed at an intermediate date. Then $\mathbb{V}(a|z)$ can be thought as the conditional utility at this intermediate date and the initial period utility is given by (11). How does this lottery compare to a second lottery with the identical reduced distribution, but all information is revealed at the final date? The second lottery uses the density $\bar{f}$ for s^*. At the intermediate date, no new information is revealed about the lottery and the resulting valuation is:

$$\bar{\mathbb{V}}(a) = h^{-1}\left[\int h[a(\xi)]\bar{f}(\xi)d\eta(\xi)\right].$$

This valuation is not conditional on any information, so at the outset we simply evaluate g at $\bar{\mathbb{V}}(a)$ to obtain the initial period utility. Provided that $g \circ h^{-1}$ is convex, the first lottery is preferred to the second. The converse is true if this function is concave. Knowing z at an intermediate date alters preferences even when the a is only allowed to depend on the signal s^*. In contrast to expected utility preferences, the timing of when uncertainty is

resolved matters.

Thus, there are two rather different motivations for building preferences that depend on more than reduced lotteries: i) wanting to incorporate a formal distinction between risk conditioned on a hidden state versus subjective uncertainty about that state and ii) wanting preferences that are sensitive to when information is revealed even when the (reduced) decision distribution is unaltered. Epstein and Zin (1989) build on this latter motivation by featuring an implied distinction between risk aversion and intertemporal substitution. In the next section I will implement both of these modifications to preferences in dynamic settings. Both can amplify the impact of learning on risk prices.

6.2   Exponential Tilting

For some convenient parameterizations, there are substantially different interpretations of this utility representation that will allow us to explore implications of statistical ambiguity. These different interpretations are implications of a well known result from applied probability:

$$\min_{m \geq 0, Em=1} EmV + \theta E\left(m \log m\right) = -\theta \log E \exp\left(-\frac{1}{\theta}V\right) \tag{12}$$

where V is a random variable that represents the future value of a stochastic process of consumption and m is a random variable used to distort probabilities. The right-hand side of (12) is a special case of:

$$h^{-1}\left(E\left[h(V)\right]\right)$$

where h is minus the negative exponential function parameterized by $\frac{1}{\theta}$: $h(V) = -\exp\left(-\frac{1}{\theta}V\right)$. As featured by Jacobson (1973), Whittle (1981) and others in the control theory literature, the left-hand side of (12) offers a rather different perspective than the apparent risk adjustment made on the right-hand side of (12). The computation:

$$EmV$$

for a positive random m with a mean of one gives an alternative way to form expectations. Formally, the random variable m induces a different probability distribution and the term $\theta E(m \log m)$ is a convex penalty in the distortion m. The left-hand side of (12) explores expectations of V using different probability distributions. By setting the parameter θ arbitrarily large, probability distortions are penalized so severely as to approximate the original expectation EV. Finite values of θ permit consideration of alternative probability

measures subject to penalty. Thus formula (12) gives an explicit link between *robustness* (left-hand side) and *risk sensitivity* (right-hand side), where the latter is modeled using an exponential risk adjustment.

Robustness allows us to endow our decision-maker with an operational form of skepticism about his model. It is implemented by the choice of a tilted or distorted probability measure induced by the minimizing m. The solution is:

$$
m = \frac{\exp\left(-\frac{1}{\theta}V\right)}{E\left[\exp\left(-\frac{1}{\theta}V\right)\right]}
\tag{13}
$$

provided that the denominator is finite. This solution gives what is known as an *exponential tilting* of the original probability. Smaller values of V receive relatively more weight than larger values in the altered probability distribution. The altered distribution is tilted towards states with lower continuation values.

The implementation via a tilted probability turns out to be of considerable value. The minimizing solution is useful for representing uncertainty premia and providing a different perspective on the source of those premia. Previously, I described the potential role for statistical latitude among alternative probability models given data histories. I now have a way to construct these alternative models and to ask how large the resulting statistical discrepancy between the minimized solution and the original benchmark probability model is. [15]

While this representation of preferences using exponential tilting relies on a particular parametric structure, it is mathematically convenient. In what follows I will apply (12) in multiple ways. In dynamic contexts, it is most fruitful to work with continuation values for optimal plans because of the usefulness of Bellman-equation methods. First, I will exploit (12) as applied to future continuation values by either endowing the decision-maker with a concern about the specification of Markov state transition probabilities (left-hand side) or a concern about the intertemporal composition of risk as in Kreps and Porteus (1978), Epstein and Zin (1989) and others (right-hand side). Second, by characterizing the dependence of future continuation values computed a function of a hidden state z, I will use (12) in conjunction with a negative exponential specification of the g function to endow decision makers either with a concern about the specification of the probabilities assigned

---

[15]In a choice problem such as an investment problem, the minimizing solution will differ as alternative choices are considered. It will often be the case that the minimization can be done after maximization of utility without changing the value. Thus a min-max theorem can be invoked. In such circumstances we can still infer a worst case probability distribution by exchanging the order of minimization and maximization.

to the hidden states (left-hand side) or a smooth ambiguity adjustment as in Klibanoff, Marinacci, and Mukerji (2005) (right-hand side).

These ideas are developed more formally in several recent papers. While (12) exploits a particular functional form, Maccheroni, Marinacci, and Rustinchini (2006b) provide an axiomatic justification for a more general version of this penalization formulation given by the left-hand side of (12), where the convex function $m \log m$ is replaced by a more general convex function. Hansen, Sargent, Turmuhambetova, and Williams (2006c) show how the intertemporal counterpart to (12) is related to the max-min expected utility of Gilboa and Schmeidler (1989) by formally interpreting the penalization parameter θ as a Lagrange multiplier on a constraint over a family of probability distributions. Maccheroni, Marinacci, and Rustichini (2006a) explore more general dynamic formulations of preferences based on penalization. Finally, Hansen and Sargent (2006b) use two versions of negative-exponential formulation to address simultaneously two forms of misspecification described previously: (a) misspecification in the underlying Markov law for the hidden states and (b) misspecification of the probabilities assigned to the hidden Markov states.[16]

7 Learning and Uncertainty Premia

Empirical evidence suggests that risk premia move in response to aggregate fluctuations (*e.g.* see Campbell and Cochrane (1999) and Lettau and Ludvigson (2009)). I now explore how learning might contribute to an explanation for this phenomenon. While I will present some highly stylized models, the lessons from this analysis are informative for more ambitious quantitative investigations. The Wonham (1964) filter will be a key input into our characterization.

My characterizations of prices will focus on what is usually termed the "local risk return tradeoff." In continuous-time environments, "local" means instantaneous and the tradeoff answers the question: "how do we compensate investors for risk borne in the immediate future?" I use the term "uncertainty premia" to capture the additional components to pricing that emerge from using the decision theory of Section 6 that in some way or another does not simply reduce compound lotteries. In dynamic economies valuation of cash flow exposure uncertainty is pertinent for all horizons, not just the immediate one. The recursive nature of asset pricing allows us to, in effect, integrate the local consequences into

---

[16]Epstein and Schneider (2003) make similar distinctions while developing other interesting formulations and applications of ambiguity aversion and learning.

implications for longer horizons. There is good reason to suspect that learning can have a more potent impact for valuation over longer horizons. Constructing a model in which learning matters for short term risk analysis is a tall order, but such a model will likely pay off by also implying substantial consequences for risk-return tradeoffs for longer horizons.

For the equilibrium calculation I imitate a device used in the rational expectations literature (see Lucas and Prescott (1971)) by introducing a fictitious social planner. Given a consumption endowment, the role of this planner is to compute value functions and the exponentially slanted probabilities associated with these functions. The sole purpose of this planning problem is to characterize these implied probability distortions. If production were incorporated, then the planner's problem would be more ambitious, but it would still include the computation of these distortions. Behind this solution to the planner's problem is a counterpart to rational expectations equilibrium with decentralized prices. Continuation values are the utility values assigned to consumption processes looking forward, and they will be computed as functions of the Markov state using continuous-time versions of Bellman's equation.

Conveniently, the probability distortion associated with exponential tilting, formula (13), is computed using continuation values. This approach can be viewed as a device for computing risk premia, as a way to generate alternative beliefs, or as a reflection of statistical ambiguity on the part of investors. It is latter interpretation that I feature here.

Following Veronesi (2000), we use the probability model assumed by Wonham (1964) in which the signal is the growth rate in consumption. In this specification, the expected growth rate of consumption has infrequent jumps:

$$dc_t = \kappa \cdot z_t + \sigma dB_t.$$

By solving the filtering problem, I compute a second evolution for consumption that endows investors with less information. In this second specification, the expected growth rate of consumption moves continuously as a function of the probabilities, the $\bar{z}_t$'s. To an econometrician looking only at consumption data, these two specifications are indistinguishable. I will make reference to both information structures and their implications for pricing.

In what follows I compute alternative value functions and probability distortions, beginning with expected utility. My approach in this paper will be derivation by assertion, and the interested reader will have to look elsewhere for formal derivations.

7.1 Continuation Values for Expected Utility

Given the assumption of a unitary elasticity of substitution, we look for continuation values of the form $V_t + c_t$, where V_t depends either on the state vector z_t or the hidden state probabilities $\bar{z}_t$.

Suppose for the moment that z_t is observed as of date t. For a reference point consider discounted expected utility in continuous time. In this case we may represent the continuation value as $V_t = v \cdot z_t + c_t$ where v is an n-dimensional vector of numbers. The vector v satisfies the linear equation

$$0 = -\delta v + Av + \kappa, \tag{14}$$

and hence $v = (\delta I - A)^{-1} \kappa$. The continuation value when z_t is not observed is $V_t = v \cdot \bar{z}_t + c_t$, which may be computed by applying the Law of Iterated Expectations or equivalently by reducing the associated compound lottery.

When the jumps are observed, there are two risk components to price: the Brownian increment dB_t and the jump process $\{z_t\}$. Since the consumption does not jump (only its conditional mean jumps), the local risk price for the jump component is zero. Since the elasticity of substitution is unity, the Brownian motion risk price is σ. In the reduced information economy in which the jump component is not observed, only the increment $d\bar{B}_t$ is priced. Since the coefficients on dB_t and $d\bar{B}_t$ are the same for both information structures, the local risk prices remain the same for this economy. In this sense the introduction of learning within a rational expectations equilibrium is inconsequential for the local risk price vector.[17]

In defense of rational learning, the prices of cash flows over finite time intervals will be sensitive to the information structure, and this sensitivity can be substantial depending on the model specification. In order to generate a model in which learning alters local prices, however, I explore other preferences as described previously.

7.2 Continuation Values and Exponential Tilting

Consider next a modification as in Kreps and Porteus (1978) under the assumption that z can be observed. Let h be the negative exponential function with parameter value θ_f. This

---

[17]Arguably this conclusion takes time separability in preferences too literally in a continuous-time model. Hindy and Huang (1992) and Heaton (1995) argue that locally durability should be an important feature of preferences specified at high frequencies. Nevertheless, I find this local analysis here to be revealing because it shows how to amplify the role of learning.

function is used to adjust future continuation values. In this case we modify Bellman's equation:

$$0 = -\delta v + \kappa - \theta_f \text{diag} \left[\exp \left(\frac{v_i}{\theta_f} \right) \right] \left(A \left[\exp \left(-\frac{v}{\theta_f} \right) \right] \right) - \frac{1}{2\theta_f} \sigma^2 \mathbf{1}_n \qquad (15)$$

where $\mathbf{1}_n$ is an n-dimensional vector of ones. The new terms included in the Bellman equation adjust the continuation values for risk both jump risk and the Brownian motion risk. (See Anderson, Hansen, and Sargent (2003) for a derivation.) As θ_f gets arbitrarily large, this Bellman equation collapses to the equation (14) used for evaluating discounted expected utility.

This Bellman equation is the counterpart to the right-hand side of (12). Associated with the left-hand side is probability distortions induced by exponential tilting. While we will not formally derive this distortion, it is easy to characterize. For this continuous time limit, the exponential tilting has a simple impact on the underlying Brownian motion. A constant drift is added of the form $-\frac{\sigma}{\theta_f}$. The negative of this drift is the uncertainty premia added to the risk premium σ derived for the expected utility model.

Under this distorted probability, consumption evolves according to:

$$dc_t = \kappa \cdot z_t dt - \frac{\sigma^2}{\theta_f} dt + \sigma d\widetilde{B}_t$$

for some standard Brownian increment $d\widetilde{B}_t$. Thus we have subtracted $\frac{\sigma^2}{\theta_f}$ from all of the hypothetical growth states. This constant adjustment is a feature of other models as well, including the discrete time models of Tallarini (2000) and Hansen, Heaton, and Li (2006b).

The transition probabilities for the Markov process are also distorted by the exponential tilting. The transitions to states with the smaller continuation values will be made more probable. The jump risk exposure now has a nonzero uncertainty premia in contrast to the zero risk premium from the expected utility economy.[18]

This gives a continuous-time counterpart to the discussion in Section 4.2. By interpreting the uncertainty premia as reflecting statistical ambiguity on the part of investors, Anderson, Hansen, and Sargent (2003) and Maenhout (2006) argue that the statistical discrimination analysis of Chernoff (1952) suggests how large this uncertainty component could plausibly be. It suggests how much statistical latitude there might plausibly be in distorting the consumption growth rates from the vantage point of skeptical investors,

---

[18]See Liu, Pan, and Wang (2005) for a related example featuring jump risk.

investors whose doubts about their model specification cannot be dismissed easily with statistical evidence.

Alternatively, a rational expectations econometrician calibrating the model could have made a mistake in building a rational expectations model by not endowing agents with lower potential growth rates for consumption. A Chernoff-type calculation based on real data controls the extent to which growth rates could be diminished, but they are set at this new level with full investor confidence.

This model of investor preferences increases the predicted uncertainty premia associated with the Brownian motion increment $d\widetilde{B}_t$, but it does not cause them to be time varying. I now examine another modification to the model which delivers time-varying premia.

7.3 Exponential Tilting and Less Information

Suppose now that the state variable z is not observed. Instead it is disguised requiring that statistical inferences be made using the Wonham (1964) filter. I may not just average the solution to equation (15) over the hidden states to obtain the solution to this problem. Instead, as Kreps and Porteus (1978) and Epstein and Zin (1989) show, the intertemporal composition of risk matters. Thus I must solve a new Bellman equation that includes an alternative risk adjustment to the continuation value

$$0 = -\delta\mathbb{V} + \kappa \cdot \bar{z} + \bar{z}'A\frac{\partial\mathbb{V}}{\partial\bar{z}} + \frac{1}{2}\Delta'\frac{\partial^2\mathbb{V}}{\partial\bar{z}\partial\bar{z}'}\Delta - \frac{1}{2\theta_f}\left[\left|\frac{\partial\mathbb{V}}{\partial\bar{z}}\cdot\Delta\right|^2 + 2\sigma\frac{\partial\mathbb{V}}{\partial\bar{z}}\cdot\Delta + \sigma^2\right] \qquad (16)$$

where the value function is $\mathbb{V}(\bar{z}) + c$. The last term captures the risk adjustment to continuation values necessary for the Kreps and Porteus (1978) recursion (see Duffie and Epstein (1992)).

Again I use a link to robustness to construct an implied change in probability measure. The distortion again adds a drift to the Brownian motion, but now the drift depends on the state probabilities. Three contributions to the uncertainty premia are given in Table 1. While we will not derive this formula, it follows from the analysis in Hansen, Sargent, Turmuhambetova, and Williams (2006c). The first term is the risk adjustment from expected utility theory when the IES is unity. The second term is familiar from our analysis of the model in which the jump component is observed.[19] The impact of learning

---

[19]An astute reader will notice that I have also distorted the dynamics for $\bar{z}$. We retain use of this as a state variable, but we lose its interpretation as the solution to a simple filtering problem.

Table 1: **Risk and Uncertainty Premia**

Exp. Utility IES $= 1$	exponential tilting consumption dynamics	exponential tilting state estimation dynamics
σ	$\frac{\sigma}{\theta_f}$	$\frac{1}{\theta_f}\Delta(\bar{z}) \cdot \frac{\partial \mathbb{V}(\bar{z})}{\partial \bar{z}}$
time invariant	time invariant	time varying

Notes: The value function has functional form: $\mathbb{V}(\bar{z})+c$, σ is the response of consumption to new information, and $\Delta(\bar{z})$ is the vector of responses of the probabilities to new information.

is reflected in the third term, which depends explicitly on $\bar{z}$. The second two terms depend on the derivatives of the value function and the local volatility. They distort the evolution of consumption and the state probabilities from the Wonham filter via exponential tilting. Unfortunately, we lose some pedagogical simplicity because the value function, and hence its derivative, must be computed numerically.

To illustrate this solution, consider an example with two states as in Cagetti, Hansen, Sargent, and Williams (2002).[20] From example 5.1, when there are two hidden states, the vector Δ has:

$$\bar{z}_{1,t}(1 - \bar{z}_{1,t})\left(\frac{\kappa_1 - \kappa_2}{\sigma}\right)$$

as its first entry and the negative of this as its second entry. The scale factor $\bar{z}_{1,t}(1 - \bar{z}_{1,t})$ is close to zero when there is a preponderance of evidence for one or the other states. This term is large when it is hard to tell the two states apart, that is when $\bar{z}_{1,t} = 1 - \bar{z}_{1,t} = 1/2$. The actual uncertainty prices depend on value function derivative as well, but it remains true that uncertainty prices become large when there is ambiguity about the hidden state probabilities, as illustrated Figure 4.

When Cagetti, Hansen, Sargent, and Williams (2002) fit a technology shock model with two regimes using econometric methods, like Hamilton (1989), they found growth rate regimes that moved over what macroeconomists typically refer to as the business cycle. The time series of resulting uncertainty prices are large at dates at which investors do not know which regime they are in: these are dates at which both regime probabilities

---

[20]Cagetti, Hansen, Sargent, and Williams (2002) consider formally a production economy and they do not impose a unitary elasticity of substitution.

Figure 4: **Uncertainty Prices**
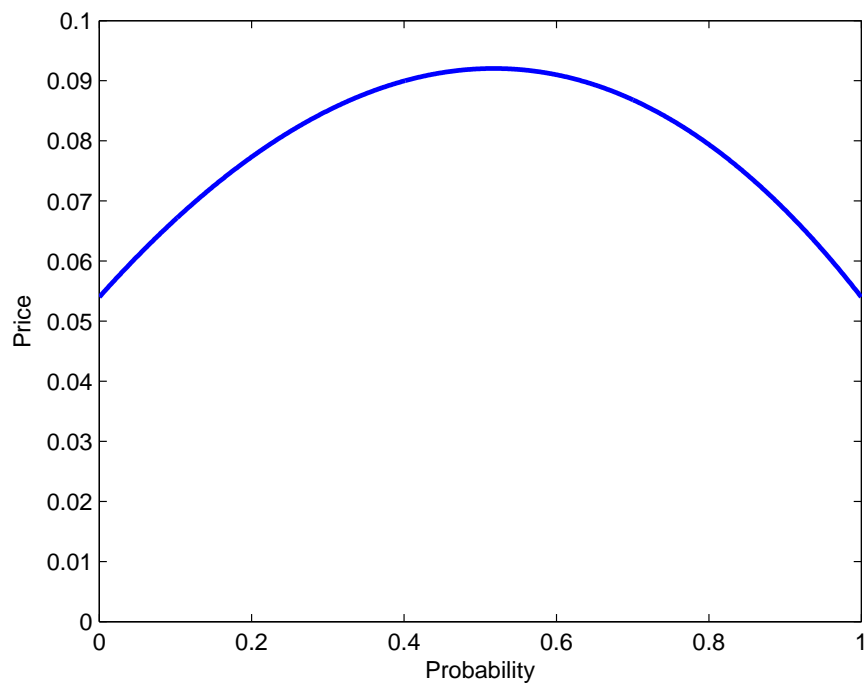
Notes: The uncertainty price is the sum of the second and third components given in Table 1. It is computed for a two-state Markov chain. To produce this curve, I assumed an intensity matrix A from Cagetti, Hansen, Sargent, and Williams (2002) with off-diagonal elements equal to .0736 and .352. The growth rates are $\kappa_1 = .0071$ and $\kappa_2 = .0013$. The volatility of consumption is $\sigma = .0054$. The value function was computed by solving the Bellman equation in Section 7.3 as an ordinary differential equation in $\bar{z}_1$ with boundary conditions at $\bar{z} = (\bar{z}_1, 1 - \bar{z}_1) = (0, 1)$ and $\bar{z} = (\bar{z}_1, 1 - \bar{z}_1) = (1, 0)$. This computation was performed in Matlab using the function bvp4c.

are one-half. Repeated observations of low consumption growth strengthen investor beliefs that they are in the low growth regime, thereby resolving some of the uncertainty and reducing the premium. I imagine that by including more states, in particular more low frequency movements in growth rates, I can modify this outcome so that some repeated observations with low growth *increases* the uncertainty about an underlying growth rate regime.[21]

While this example produces interesting time series variation in local uncertainty premia, it does so by distorting the dynamic evolution equation for the state vector. This includes distorting the component originally constructed as Wonham's (1964) solution to a filtering problem. Investors treat state estimates from the Wonham (1964) filter like any other observable state variable, and they do not distort the current period state probabilities.

7.4   Estimation Ambiguity

I now explore an alternative approach that directly distorts the state probabilities. To feature the role of ambiguity in the assignment of state probabilities, I follow Klibanoff, Marinacci, and Mukerji (2005) and Hansen and Sargent (2006b) by introducing a separate adjustment for ambiguity over the probabilities assigned to states.

Using a continuous-time counterpart to a decision model of Hansen and Sargent (2006b) and decomposition (9), we may obtain a modified version of the Bellman equation (16). To feature the role of hidden states, the fictitious planner modifies the equation by considering the evolution of continuation values prior to the information reduction. Even though the value function depends only on $\bar{z}$ and c, its evolution now depends on the realized hidden states. Hansen and Sargent (2006b) introduce a second parameter, say θ_b, to penalize distortions to the probability vector $\bar{z}$ used by the planner for computing the averages required for a new continuous-time Bellman equation. The resulting solution remains difficult to compute unless the number of states is small.

For my numerical examples, I use a second approach suggested by Hansen and Sargent (2006b). The solution can be easier to compute, which we exploit in solving the four hundred state Markov chain example which follows.[22] Consider again the Kreps and Porteus (1978) recursion conditioned on the hidden state z. Recall that the value function is $v \cdot z + c$

---

[21]Such extensions are worthwhile, but the value function for this model must be solved numerically. This limits the scope of such an analysis.

[22]While computationally simpler, its game theoretic underpinnings are more subtle.

where v is a vector of real numbers. I use the continuation values in conjunction with (12) for $\theta = \theta_b$ to infer Recall that the probability distortion results in an exponential tilting of the probability assignment toward states with the lowest continuation values. That is, let

$$v_i^* = \exp\left(-\frac{v_i}{\theta_b}\right)$$

for some positive value of the parameter θ_b. Large values of θ_b make v_i^*'s close to a constant value of unity. The distorted or exponentially tilted probabilities assigned to hidden state are:

$$\tilde{z}_t^i = \frac{v_i^* \bar{z}_{i,t}}{\sum_i v_i^* \bar{z}_{i,t}}.$$

These tilted probabilities induce a distortion in the expected growth rate for consumption and hence add a component to the uncertainty premia.

Table 2: **Risk and Uncertainty Prices**

Exp. Utility IES $= 1$	exponential tilting consumption dynamics	exponential tilting state estimation
σ	$\frac{\sigma}{\theta_f}$	$\frac{(\bar{z}-\tilde{z})\cdot\kappa}{\sigma}$
time invariant	time invariant	time varying

Notes: The value function as functional form: $\mathbb{V}(z) + c$, σ is the response of consumption to new information, $\bar{z}$ is the vector of probabilities from the Wonham filter, $\tilde{z}$ is the exponentially tilted counterpart and κ is the vector of alternative growth rates.

Three contributions to the uncertainty premia are given in Table 2. The first two are familiar from our previous example economies and the third is unique to this example. The first is a risk premia, and the second term is determined by the continuation values conditioned on z and the parameter θ_f. Both are constant. The third term is unique to this example. Since it depends on the hidden state probabilities and their distortions, this term is time varying. Its magnitude is determined in part by the parameter θ_b used in computing the exponential tilted state probabilities. The third term is unique to this example. From a robustness standpoint the parameter θ_f reflects forward-looking skepticism about assumed dynamics and the parameter θ_b a backward looking skepticism about the constructed state

probabilities.

The time series plots in Figure 5 display the sum of the second and third components of the uncertainty premia, which are the components associated with probability slanting. I construct a Markov chain to approximate four different parameter configurations or sub-models. Formally a submodel is a collection of states for which there is no chance of leaving that collection. I design a four hundred state Markov chain to approximate a model selection problem or estimation problem for investors. One hundred states were used for each of the four sub-models. The corresponding intensity matrix A is block diagonal with four blocks. I construct the first three sub-models by approximating the consumption dynamics given in example (2.2) in which the process $\{z_t\}$ is hidden from the agents. I use three different values of the autoregressive parameter .97, .98 and .99. The corresponding coefficients on the shock $u_{2,t+1}$ (the conditional standard deviations of the hidden state process) were obtained by maximizing the likelihood over this parameter conditioned on the autoregressive parameter and the coefficient .0054 on the shock $u_{1,t+1}$ to the consumption growth rate equation. I use the method of Tauchen and Hussey (1991) to obtain a discrete state approximation for each of these three models. The other one hundred states are all invariant states designed to approximate alternative mean growth rates. I apply a standard quadrature method in constructing the discrete states.[23] For computational purposes and for the computation of uncertainty prices, I take the discrete states literally; but the setup is designed to be similar to an economy studied in great depth by Hansen and Sargent (2006a). In that paper, we used Kalman filtering methods for two alternative models and computed the sequence of posterior probabilities for these models given sample evidence on consumption. Here I use the same data as used in that analysis to solve the filtering problem.

The second term in Table 2 is time invariant. By changing θ_f, I alter the level of the uncertainty premia. This is reflected in Figure 5. The two lower curves were computed for $\theta_f = .10$ and upper curves for $\theta_f = .05$. A smaller value of θ_f implies less penalization in the investors' search over alternative probability distributions. The third term in Table 2 induces time series variation in the uncertainty premia while having little impact on the level. The smooth curves in Figure 5 are computed for $\theta_b = 24$ and the more volatile curves for $\theta_b = 6$. By construction, the time series trajectories are similar to those reported in the more comprehensive analysis by Hansen and Sargent (2006a) except that I have introduced

---

[23]To form an (approximate) intensity matrix for the continuous-time Markov chain, I subtracted an identity matrix from the discrete-time transition probability matrix.

Figure 5: **Time Series of Uncertainty Prices**
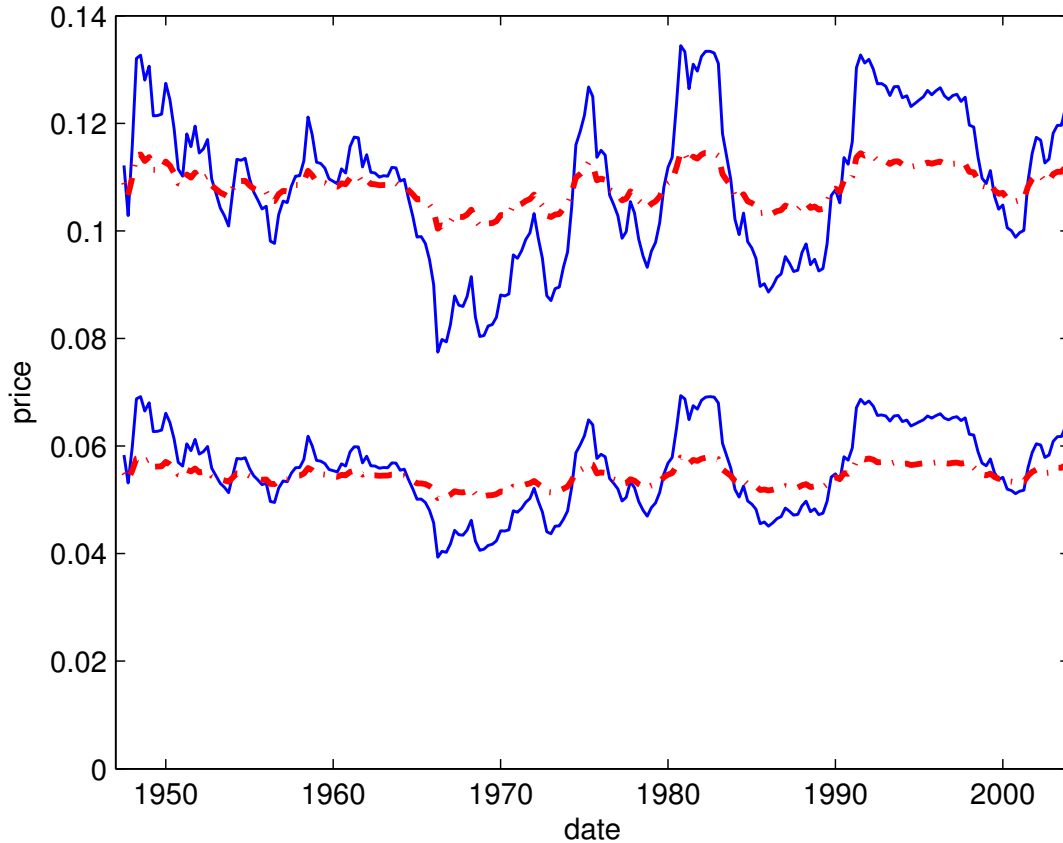
Notes: The uncertainty prices are the sum of the second and third components given in Table 2. The Markov chain has four hundred states, one hundred for each of four sub-models. The four sub-models are i) AR1 with an AR coefficient of .97, a shock standard deviation of .00058 and an unconditional mean of .0056; ii) AR1 with an AR coefficient of .98, a shock standard deviation of .00047 and unconditional mean of .0056; iii) AR1 with an AR coefficient of .99, a shock coefficient of .00024 and an unconditional mean of .0056; iv) i.i.d. model with prior on the mean given by .0056 and a prior standard deviation of .00036. The $-$ curve was computed assuming that $\theta_b = 24$, and $-\cdot$ curve was computed assuming that $\theta_b = 6$ is. The upper two plots were computed assuming that $\theta_f = .05$, and the lower two plots were computed assuming that $\theta_f = .1$.

additional models to approximate the problem of estimating the parameters governing the dynamics of $\{z_t\}$.

I find it convenient to think of the first three sub-models as three different parameter specifications of predictability in consumption growth. By including all three sub-models in the analysis I have approximated an estimation problem. The fourth sub-model is different because consumption growth rates are not predictable. As a consequence it implies less long-run uncertainty. The signal history of post-war consumption growth does not allow investors either to confirm or reject this fourth sub-model with full confidence. Probabilities are tilted away from this sub-model based on the continuation values. A string of relative high or relatively low consumption growth rates both give evidence for consumption predictability. The relatively high growth rates induce less tilting towards the sub-models with predictability in consumption because if consumption is predictable it should remain high, at least temporarily. In contrast, relatively low growth rates in consumption induce more tilting towards the sub-models with predictable consumption growth and this in turn gives a larger uncertainty premia.

8 Extensions

There are very special ingredients in my example economies. They were designed in part to magnify the impact of learning on uncertainty prices. On the other hand, there are empirical limitations to these economies that can be anticipated from previous literature.

My example economies have arguably withheld too much information from economic agents. For instance, multiple signals make learning more informative, and it remains valuable to explore implications that allow for an econometrician to understate the knowledge of economic agents. Learning within a rational expectations equilibrium already adds to the econometrician's challenge by requiring an initial specification or prior for the beliefs about hidden Markov states, parameters, or model indicators. The decision theory that we explored avoids the reduction of compound lotteries and thus prevents direct application of the Law of Iterated Expectations as commonly used in rational expectations econometrics to deduce robust implications. While econometric analysis may be more challenging, it is a challenge with potentially valuable payoffs.

I chose not to feature models in which there is conditional volatility present in the evolution for consumption. I did this to show how learning can induce time variation in uncertainty prices without an additional exogenous source of variation. Low frequency

45

volatility movements, however, are a potentially important additional ingredient.[24]

Similarly, to simplify the characterization of value functions and probability distortions I restricted the IES (intertemporal elasticity of substitution) to be one. For models that seek to understand better wealth and aggregate stock price dynamics, this restriction is problematic because it implies constant wealth consumption ratios. On the other hand, approximating around an economy with IES=1 can be a useful characterization device as I illustrated in Section 1.

My focus on one-period (in discrete time) or local (in continuous-time) uncertainty prices made it more difficult for learning to matter. If learning matters for short-horizon valuation, then its impact should be more potent for longer horizons. Recent asset pricing literature has focused on the role of long run uncertainty on cash flow valuation. (For example, see Campbell and Vuolteenaho (2004), Bansal, Dittmar, and Lundblad (2005), Santos and Veronesi (2005), Hansen, Heaton, and Li (2006b) and Croce, Lettau, and Ludvigson (2006).) Since statistical measurements for long-horizons are known to be fragile, formally incorporating learning into such analysis is an obvious but important extension.

My models imposed homogeneity on investors. This allowed me to compute a single tilted probability model and simplified my analysis. While introducing heterogeneity among investors will complicate model solution, it has intriguing possibilities. The investors will slant probabilities in different directions giving rise of a form of *ex post* heterogeneity in beliefs. There is much more to be done.

9   Conclusion

The cross-equation restrictions used in rational expectations econometrics get much of their empirical power by endowing agents with more precise information than econometricians. This includes information on endowments, cash flows and technology shocks. The rational expectations agents have done a lot of un-modeled work before the econometrician steps in. In this paper I have explored ways to close this informational gap by giving economic agents some skepticism about the models they use. I showed how investor concerns about statistical ambiguity are reflected in equilibrium prices. In our example economies I avoided endowing economic agents with full confidence in probability models that are demonstrably hard to estimate. By introducing *learning within an equilibrium*, I showed

---

[24]Weitzman (2007) has recently shown that for some priors on volatility, learning can be particularly challenging and consequently can have a big impact on the predicted asset returns.

how learning is reflected in the dynamic evolution of local uncertainty prices. These uncertainty premia reflect investors' doubts about their probability models. Learning about very low frequency events including the primitive model specification can lead to uncertainty premia that are large when macroeconomic growth is sluggish. This changes the structure of cross-equation restrictions, but not necessarily their potency. While there are other possible interpretations for the equilibrium outcomes I displayed, including changing beliefs or embracing preferences that decompose risks in alternative ways, I find the relation to statistical ambiguity to be the most appealing.

There are analogous questions regarding the role of uncertainty in the exploration of hypothetical government interventions. The models I used drew a distinction between risks conditioned on a hidden model specification or a hidden state, and uncertainty about that specification or hidden state. If this distinction is important in understanding evidence from security market data, then use of this evidence in the analysis of stochastic interventions will require a careful accounting of the probability structure of the policy intervention. What skepticism will economic agents have about the alternative probability structure and what role will learning play in validating or altering beliefs? While such distinctions are not typical in the formal analysis of policy changes, perhaps they should become part of the normative vocabulary as argued in the context of monetary policy by Milton Friedman. Rational expectations models have been demonstrably successful in featuring the role of credibility in policy making, but there is scope to explore further the role of beliefs, doubts and learning in a formal way.

References

Anderson, Evan W., Lars Peter Hansen, and Thomas J. Sargent. 2003. A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection. *Journal of the European Economic Association* 1:69–123.

Bansal, Ravi and Amir Yaron. 2004. Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. *Journal of Finance* 59:1481–1509.

Bansal, Ravi, Robert F. Dittmar, and Christian T. Lundblad. 2005. Consumption, Dividends, and the Cross-Section of Equity Returns. *Journal of Finance* 60:1639–1672.

Bray, Margaret. 1982. Learning, Estimation and the Stability of Rational Expectations. *Journal of Economic Theory* 26:318–339.

Bray, Margaret and David M. Kreps. 1987. Rational Learning and Rational Expectations. In *Arrow and the Accent of Modern Economic Theory*, edited by G. R. Feiwel, 597–625. New York: New York University Press.

Cagetti, Marco, Lars Peter Hansen, Thomas J. Sargent, and Noah Williams. 2002. Robustness and Pricing with Uncertain Growth. *Review of Financial Studies* 15:363–404.

Campbell, John. 1996. Understanding Risk and Return. *Journal of Political Economy* 104:298–345.

Campbell, John and Robert Shiller. 1988. Stock Prices, Earnings, and Expected Dividends. *Journal of Finance* 43:661–676.

Campbell, John Y. and John Cochrane. 1999. Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy* 107:205–251.

Campbell, John Y. and Tuomo Vuolteenaho. 2004. Bad Beta, Good Beta. *American Economic Review* 94:1249–1275.

Chen, Xiaohong and Halbert White. 1998. Nonparametric Adaptive Learning with Feedback. *Journal of Economic Theory* 82:190–222.

Chernoff, Herman. 1952. A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations. *Annals of Mathematical Statistics* 23:493–507.

Croce, Mariano M., Martin Lettau, and Sydney Ludvigson. 2006. Investor Information, Long-Run Risk, and the Duration of Risky Assets. New York University.

Cumby, Robert, John Huizinga, and Maurice Obstfeld. 1983. Two-step Two-stage Least Squares Estimation in Models with Rational Expectations. *Journal of Econometrics* 21:333–355.

David, Alexander. 1997. Fluctuating Confidence in Stock Markets: Implications for Returns and Volatility. *Journal of Financial and Quantitative Analysis* 32:457–462.

Duffie, Darrell and Larry G Epstein. 1992. Stochastic Differential Utility. *Econometrica* 60 (2):353–94.

Epstein, Larry and Martin Schneider. 2003. IID: Independently and Indistinguishably Distributed. *Journal of Economic Theory* 113:32–50.

Epstein, Larry and Stanley Zin. 1989. Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. *Econometrica* 57:937–969.

Evans, George W. and Seppo Honkapohja. 2001. *Learning and Expectations in Macroeconomics.* Princeton, New Jersey: Princeton University Press.

Geweke, John. 2001. A Note on Some Limitations of CRRA Utility. *Economics Letters* 71:341–345.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken. 1989. A Test of Efficiency of a Given Portfolio. *Econometrica* 57:1121 – 1152.

Gilboa, Itzhak and David Schmeidler. 1989. Maxmin Expected Utility with a Non-Unique Prior. *Journal of Mathematical Economics* 18:141–153.

Hamilton, James D. 1989. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* 57:357–384.

Hansen, Lars Peter. 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50:1029–1054.

Hansen, Lars Peter and Ravi Jagannathan. 1991. Implications of Security Market Data for Models of Dynamic Economies. *Journal of Political Economy* 99.

Hansen, Lars Peter and Scott F. Richard. 1987. The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica* 55.

Hansen, Lars Peter and Thomas J. Sargent. 1980. Formulating and Estimating Dynamic Linear Rational Expectations Moddels. *Journal of Economic Dynamics and Control* 2:7–46.

———. 1991. Time Series Implications of Present-Value-Budget Balance and Martingale Models of Consumption and Taxes. In *Rational Expectations Econometrics*, edited by Lars Peter Hansen and Thomas J. Sargent, 121–161. Westview Press.

———. 2006a. Fragile Beliefs and Price of Model Uncertainty. University of Chicago and New York University.

———. 2006b. Recursive Robust Estimation and Control Without Commitment. *Journal of Economic Theory* forthcoming.

Hansen, Lars Peter and Jose Scheinkman. 2006. Long Run Risk: An Operator Approach. University of Chicago and Princeton University.

Hansen, Lars Peter and Kenneth J. Singleton. 1982. Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. *Econometrica* 50:1269–1299.

Hansen, Lars Peter, William Roberds, and Thomas J. Sargent. 1991. Exact Liner Rational Expectations Models. In *Rational Expectations Econometrics: Specification and Estimation*, edited by Lars Peter Hansen and Thomas J. Sargent, 45–76. Westview Press.

Hansen, Lars Peter, John Heaton, Junghoon Lee, and Nikolai Roussanov. 2006a. Risk Aversion and Intertemporal Substitution. In *Handbook of Econometrics*, edited by James J. Heckman. Elsevier. Forthcoming.

Hansen, Lars Peter, John C. Heaton, and Nan Li. 2006b. Consumption Strikes Back?: Measuring Long Run Risk. University of Chicago.

Hansen, Lars Peter, Thomas J. Sargent, Gauhar A. Turmuhambetova, and Noah Williams. 2006c. Robust Control and Model Misspecification. *Journal of Economic Theory* 128:45–90.

Hayashi, Fumio and Christopher Sims. 1983. Nearly Efficient Estimation of Time-Series Models with Predetermined Variables. *Econometrica* 51.

Heaton, John C. 1995. An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications. *Econometrica* 63:681–717.

Hindy, Ayman and Chi-Fu Huang. 1992. Intertemporal Preferences for Uncertain Consumption: A Continuous Time Approach. *Econometrica* 60:781–801.

Jacobson, David H. 1973. Optimal Stochastic Linear Systems with Exponential Performance Criteria and Their Relation to Deterministic Differential Games. *IEEE Transactions for Automatic Control* AC-18:1124–131.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji. 2005. A Smooth Model of Decision Making Under Uncertainty. *Econometrica* 73:1849–1892.

Kreps, David M. and Evan L. Porteus. 1978. Temporal Resolution of Uncertainty and Dynamic Choice. *Econometrica* 46:185–200.

Lettau, Martin and Sydney Ludvigson. 2009. Measuring and Modeling Variation in the Risk-Return Tradeoff. In *Handbook of Financial Econometrics*, edited by Yacine Ait-Sahalia and Lars Peter Hansen, 617–690. North Holland.

Liu, Jun, Jun Pan, and Tan Wang. 2005. An Equilibrium Model of Rare-Event Premia and Its Implication for Option Smirks. *Review of Financial Studies* 18:131–164.

Lucas, Robert E. 1972a. Econometric Testing of the Natural Rate Hypothesis. In *The Econometrics of Price Determination*, edited by O. Eckstein, 50–59. Washington, D.C.: Board of Governors of the Federal Reserve Board.

———. 1972b. Expectations and Neutrality of Money. *Journal of Economic Theory* 4:103–124.

———. 1978. Asset Prices in an Exchange Economy. *Econometrica* 46:1429–1445.

Lucas, Robert E. and Edward C. Prescott. 1971. Investment Under Uncertainty. *Econometrica* 39:659–681.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini. 2006a. Dynamic Variational Preferences. *Journal of Economic Theory* 128:4–44.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustinchini. 2006b. Ambiguity Aversion, Robustness, and the Variational Representation of Preferences. *Econometrica* 74:1147–1498.

Maenhout, Pascal. 2006. Robust Portfolio Rules and Detection-Error Probabilities for a Mean-Reverting Risk Premium. *Journal of Economic Theory* 128:136–163.

Marcet, Albert and Thomas J. Sargent. 1989. Convergence of Least Squares Learning Mechanisms in Self-referential Linear Stochastic Models. *Journal of Economic Theory* 48:337–368.

Mehra, Rajnish and Edward C. Prescott. 1985. The Equity Premium: A puzzle. *Journal of Monetary Economics* 15:145–161.

Muth, John H. 1961. Rational Expectations and the Theory of Price Movements. *Econometrica* 29.

Newman, Charles M. and Barton W. Stuck. 1979. Chernoff Bounds for Discriminating Between Two Markov Processes. *Stochastics* 2:139–153.

Restoy, Fernando and Philippe Weil. 1998. Approximate Equilibrium Asset Prices. NBER Working Paper 6611.

Santos, Jesus and Pietro Veronesi. 2005. Cash-Flow Risk, Discount Risk, and the Value Premium. NBER Working Papers 11816, National Bureau of Economic Research.

Saracoglu, Rusdu and Thomas J. Sargent. 1978. Seasonality and Portfolio Balance under Rational Expectations. *Journal of Monetary Economics* 4:435–458.

Sargent, Thomas J. 1973. Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment. *Brookings Papers in Economic Activity* 2.

———. 1999. *The Conquest of American Inflation.* Princeton, New Jersey: Princeton University Press.

Sclove, Stanley L. 1983. Time-Series Segmentation: A Method and a Model. *Information Science* 29:7–25.

Segal, Uzi. 1990. Two-Stage Lotteries Without the Reduction Axiom. *Econometrica* 58:349–377.

Sharpe, William. 1964. Capital Asset Prices: A Theory of Market Equilibrium. *Journal of Finance* 19.

Shiller, Robert. 1972. *Rational Expectations and the Structure of Interest Rates*. Ph.D. thesis, M.I.T.

———. 1982. Consumption, Asset Markets and Macroeconomic Fluctuations. *Carnegie-Rochester Conference Series on Public Policy* 17:203–238.

Tallarini, Thomas D. 2000. Risk-sensitive real business cycles. *Journal of Monetary Economics* 45 (3):507–532.

Tauchen, George and Robert Hussey. 1991. Quadrature-Based Method for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models. *Econometrica* 59:371–396.

Veronesi, Pietro. 2000. How Does Information Quality Affect Stock Returns. *Journal of Finance* 55:907–837.

Weitzman, Martin. 2007. Prior Sensitive Expectations and Asset Return Puzzles. *American Economic Review* Forthcoming.

Whittle, Peter. 1981. Risk Sensitive Linear Quadratic Gaussian Control. *Advances in Applied Probability* 13:764–777.

Wonham, W. M. 1964. Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering. *SIAM Journal on Control* 2:347–368.