Effects of Learning the Long-Run Asset Pricing Model

JOB MARKET PAPER

Francisco Vazquez-Grande *

The University of Chicago Booth School of Business

December 1, 2009

Abstract

This paper documents a significant increase of risk-prices in the presence of learning. I solve a model with long-run risk where both, the level and persistence of expected consumption growth are unobserved. I introduce a new methodology to quantify the effects of learning about parameter uncertainty and latent variables. The maximum Sharpe ratios increase from .07 in the benchmark case without learning to .45 in the learning economy. In my model, the representative consumer chooses state variables that are sufficient statistics of the learning problems and, conditional on her information set, forms posterior distributions of the states and future consumption growth. To reduce the complexity of optimization, I present a novel numerical approach that approximates the agent's continuation-value by nesting the solutions of problems with different information sets.

*e-mail: vazquez-grande@chicagobooth.edu. I thank, Lars Hansen, Ruey Tsay, Pietro Veronesi, Ken Judd, Hedibert Lopes, Ralph Koijen, Pedro Gete, Denis Chaves, and Jose Luis Fillat for their insight, encouragement, and support. Credit is shared; I am solely responsible for errors and omissions.

1 Introduction

How do macroeconomic variables and their evolution affect prices of financial claims? This is a fundamental question of asset pricing. To answer it, economists introduce agents with well defined preferences and model the economic environment. In these models agents are assumed to allocate resources optimally which leads to equilibrium prices that depend on the state of the economy.

The traditional approach assumes agents to have precise information about the dynamics of the economy. Not only do these agents know the model, but also all the parameters, and they observe all the macro quantities.

This paper deviates from the classical approach by allowing agents to be uncertain about the parameters of the model and also by entertaining the possibility that some macroeconomic variables are unobservable. I let agents update their beliefs about the unknown quantities as new observations become available; this is the so-called learning process.

More precisely, I consider a representative agent with recursive preferences (see Kreps and Porteus 1978) and an economy that evolves according to the one-channel long-run model of Bansal and Yaron (2004). Based on this, I focus on the effect that learning in the presence of parameter uncertainty and latent variables has on risk prices.

I choose the long-run model for two reasons. First, recursive preferences disentangle the effects of risk aversion and intertemporal substitution for asset prices. This separation, as noted by Ai (2007), allows me to choose preference parameters that deliver an increase in the prices of risk in the presence of learning. Also Bakshi and Skoulakis (2009) use power utility that links risk-aversion and intertemporal substitution, and show a modest increase of risk-prices when the agent learns about volatility.

Second, Bansal and Yaron consider a predictable process for expected consumption growth that cannot be measured. They also report that the persistence parameter has a significant impact on expected returns. Therefore, this model is an ideal tool to test the effect of learning, as it is reasonable to assume that the agent faces the same uncertainty, concerning these two variables as does the econometrician.

I present results for four information endowments. Firstly, the benchmark information set, full-information, refers to an economy in which the agents know the model, can observe its parameters and expected consumption growth, and where the only source of uncertainty comes from the macroeconomic shocks. Secondly, similar to the model of Ai (2007), expected consumption growth is unobserved and filtered out from the economic signals. Thirdly, expected consumption growth is observed, but the agent learns its persistence as observations become available (Ghosh 2008 deals with a similar problem, where the persistence parameter is allowed to take two values). Finally, the agent does not observe expected consumption growth or its persistence and uses signals to calculate posterior distributions of both quantities. The solution to the last case represents the main contribution of my work.

In order to compare the effect that information has for the risk-return trade-off of consumption claims, I compare maximum Sharpe ratios across economies. I focus on Sharpe ratios, the excess expected return per unit of standard deviation, because, as highlighted by Hansen and Sargent (2009), different risks are priced under different information sets, and therefore risk prices are not directly comparable across information structures.

To illustrate the effect of learning I document a sizeable increase in maximum Sharpe ratios: from .07 in the full-information case to an average of .45 when, both the level and persistence of expected consumption growth are unknown.

In this respect I argue that the significant impact that learning has on risk prices in my model is due to two factors, (i) recursive preferences, and (ii) the compound effect of dealing with two unobserved quantities. I show that if the agent is uncertain about a single quantity, the effect of learning is greatly reduced. This is consistent with Ai (2007) and Ghosh (2008).

As a methodological contribution, this paper also proposes how to parameterize the agent's continuation-value when she solves a learning problem. Instead of using a discretized parameter space, as it is common in this literature (e.g., Veronesi 1999; Veronesi 2000; Pakoš 2008; Ghosh 2008), I propose to keep track of the conditional distribution of hidden processes and unknown parameters, summarizing these distributions by sufficient statistics. This method allows me to compute pricing implications, in the presence of parameter uncertainty and filtering problems. The evolution of the sufficient statistics represents the dynamic behavior of the conditional distribution. This idea is inspired by recent work in particle learning from Carvalho et al. (2009), that shows the benefits of using sufficient statistics in sequential learning problems.

Additionally, this paper develops a new numerical solution for problems with learning about parameter uncertainty. I nest the solutions of problems with different information sets reducing the dimensionality of the optimization and requiring consistency across information structures. In this regard, I develop the technique in the spirit of dynamic programming with boundary conditions, allowing the volatility of the posterior distribution of quantities of interest to act as a time-scale. That is, as this volatility tends to zero, the finer and coarse information sets converge, and I impose convergence on the solutions.

This paper is organized as follows. In section 2, I introduce the preferences and describe the economic dynamics and information structure. Section 3 deals with the solution of the model and outlines the numerical technique. In section 4, I describe the data, and outline the MCMC estimation procedure. Section 5 presents the results and section 6 concludes.

2 The Model

2.1 Preferences

I consider an endowment economy in discrete time. The representative agent has recursive preferences, introduced by Kreps and Porteus (1978) and applied by Epstein and Zin (1989) and Weil (1989), over the stochastic consumption path C_t. The agent's continuation-value at time t, V_t, satisfies the following recursion:

$$V_t = \begin{cases} \left((1-\beta)C_t^{1-1/\psi} + \beta\left(\mathbf{R}_t V_{t+1}\right)^{1-1/\psi}\right)^{\frac{1}{1-1/\psi}}, & \psi \neq 1 \\ \exp\left((1-\beta)\log C_t + \beta \log \mathbf{R}_t V_{t+1}\right), & \psi = 1, \end{cases} \tag{1}$$

where $\psi > 0$ is the intertemporal elasticity of substitution (IES), $\beta \in (0, 1)$ is the time-preference parameter, and $\mathbf{R}_t$ is the risk-adjusted conditional expectation operator defined as

$$\mathbf{R}_t V_{t+1} = \begin{cases} \left(\mathbf{E}\left[V_{t+1}^{1-\gamma} | \mathcal{F}_t\right]\right)^{\frac{1}{1-\gamma}}, & \gamma \neq 1 \\ \mathbf{E}\left[V_{t+1} | \mathcal{F}_t\right], & \gamma = 1, \end{cases} \tag{2}$$

where $\gamma > 0$ is the risk-aversion parameter and $\mathcal{F}_t$ denotes the information set of the representative agent at time t.

Following Hansen et al. (2007), the implied stochastic discount factor has the following form:

$$S_{t,t+1} = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-1/\psi} \left[\frac{V_{t+1}}{\mathbf{R}_t V_{t+1}}\right]^{1/\psi - \gamma}. \tag{3}$$

2.2 Economy Dynamics

I specify processes for the persistent component of expected consumption growth x and the economic signals s,

$$x_{t+1} = \rho x_t + \sigma_x \varepsilon_{t+1} \tag{4}$$

$$s_{t+1} = M + A x_t + B \varepsilon_{t+1}, \tag{5}$$

where ε_{t+1} is a vector of standard normal shocks, and consumption growth is stored in the first component of s,

$$\Delta c_{t+1} \equiv s^1_{t+1}.$$

The univariate process x comes from the long-run risk literature (Bansal and Yaron 2004; Hansen, Heaton, and Li 2008) and drives the movements in the conditional mean of consumption growth as well as affects the conditional mean of the rest of the signals.

2.3 Information Structure

I describe below the different information endowments given to the representative agent. I specify the state variables used under each information structure, and provide the distribution of future signals (and therefore consumption growth) and the evolution equations for each of the state variables, conditional on the information set at time t. I define the state variables to be sufficient statistics of the unknown quantities, inspired by the ideas of particle learning from Carvalho et al. (2009).

2.3.1 Case 1: Full-Information

I consider a benchmark case where the agent observes all history of the variables of interest and the parameters of the model are known. In this case, as x is observed, the agent does not need to use the signal vector to filter it, and the economic dynamics reduce to the one-channel long-run risk model from Bansal and Yaron (2004).

From equation (5), the distribution of future signals conditional on the information set at time t is

$$s_{t+1}|\mathcal{F}_t \sim N(M + Ax_t, BB').$$

In this case, the only state variable is x, and equation (4) describes its evolution.

2.3.2 Case 2: Learning ρ

The agent does not observe the persistence parameter ρ. I assume the x process, and the rest of the parameters of the model are observable. The agent enters each period with a prior distribution of ρ given her information set. She uses the prior to construct predictive distributions of the economic signals. Also, as a new observation of x becomes available, the agent combines this information with her prior to form a posterior distribution of the persistence parameter. This combination defines the evolution equations of the state variables. She uses the posterior as her prior next period, completing the learning process.

Proceeding this way, I consider a truncated normal prior distribution of ρ conditional on the information set of the agent at time t. I choose a truncated distribution to ensure that ρ remains within $(-1,1)$:

$$\rho|\mathcal{F}_t \sim N(\rho_t, \sigma\sigma'\pi_t^r) \text{ and } \rho \in (-1,1).$$

Then the state variables are x_t, ρ_t, and π_t^r. From equation (5), the distribution of future signals conditional on the information set is

$$s_{t+1}|\mathcal{F}_t \sim N\left(M + Ax_t, BB'\right).$$

Also, combining this prior with the state evolution equation (4), I obtain a truncated normal posterior distribution for $\rho|\mathcal{F}_{t+1}$ with mean ρ_{t+1} and variance $\sigma\sigma'\pi^r_{t+1}$, where the posterior parameters depend on the prior parameters and the new observation in the following way,

$$\begin{aligned}
\pi^r_{t+1} &= (x_t^2 + (\pi^r_t)^{-1})^{-1} \\
\rho_{t+1} &= \pi^r_{t+1}(\rho_t(\pi^r_t)^{-1} + x_{t+1}x_t).
\end{aligned}$$

This system, together with equation (4), completes the evolution equations for the state variables.

2.3.3 Case 3: Latent Variable

Under this information structure, the agent knows all the parameters of the model and observes signals. The persistent component of expected consumption growth x remains unobserved. As the state space system (4–5) is linear and Gaussian, the agent uses the Kalman filter to form her beliefs about the hidden process. Therefore, the filtered distribution of x_t given the history of signals is

$$x_t|\mathcal{F}_t \sim N(m_t, \pi^x_t),$$

where m_t and π^x_t are obtained from the signals using the Kalman filter; its explicit form can be found in the Appendix.

The state variables the agent considers are m_t, π^x_t. With this prior, the subjective distribution of signals conditional on the information set is

$$s_{t+1}|\,\mathcal{F}_t \sim N\left(M + Am_t, \pi^x_t AA' + BB'\right).$$

As the agent acquires more information, the parameters of the filtering distribution evolve, following the Kalman filter recursions:

$$
\begin{aligned}
\pi^x_{t+1} &= \rho^2 \pi^x_t + \sigma\sigma' - (\rho\pi^x_t A + B\sigma')'(AA'\pi^x_t + BB')^{-1}(\rho\pi^x_t A + B\sigma') \\
m_{t+1} &= \rho m_t + (\rho\pi^x_t A + B\sigma')'(AA'\pi^x_t + BB')^{-1}(s_{t+1} - M - Am_t).
\end{aligned}
$$

2.3.4 Case 4: Latent Variable and Learning ρ

In this coarser information set, $\mathcal{F}_t$ contains neither ρ nor $x^t = (x_1, \ldots, x_t)$. I propose a hierarchical prior structure. For ρ, I consider the same truncated normal prior as before. For x given ρ, I use a normal prior updated through the Kalman filter.

As I am looking for closed-form evolution equations for the state variables, I take a first-order expansion of the Kalman filter mean, with respect to ρ around ρ_t. I approximate the Kalman filter variance as the steady state variance at ρ_t. This approach leads to the following prior:

$$
\begin{aligned}
\rho \quad & |\mathcal{F}_t & \sim N(\rho_t, (\sigma\sigma')\pi^r_t) \text{ and } \rho \in (-1, 1) \\
x_t \quad & |\mathcal{F}_t, \rho & \sim N(a_t + b_t(\rho - \rho_t), \pi^x_{ss}(\rho_t)),
\end{aligned}
$$

with $\pi^x_{ss}(\rho_t)$ satisfying the equation:

$$
\pi^x_{ss}(\rho_t) = \rho_t^2 \pi^x_{ss}(\rho_t) + \sigma\sigma' - (\rho_t \pi^x_{ss}(\rho_t) A + B\sigma')'(AA'\pi^x_{ss}(\rho_t) + BB')^{-1}(\rho_t \pi^x_t A + B\sigma')
$$

The distribution of signals conditional on the information set at time t is

$$s_{t+1}|\,\mathcal{F}_t \sim N\left(M + Aa_t, (b_t^2\sigma\sigma'\pi_t^r + \pi_t^x)AA' + BB'\right).$$

And in this case, the states are $a_t, b_t, \pi_t^r, \rho_t$, and the evolution equations are

$$\begin{aligned}
a_{t+1} &= \rho_t a_t + K_t^0(y_{t+1} - M - Aa_t) \\
b_{t+1} &= a_t + b_t(\rho_t - K_t^0 A) + K_t^1(s_{t+1} - M - Aa_t) \\
\pi_{t+1}^r &= (\sigma\sigma' b_t^2 A'(\pi_{ss}^x(\rho_t)AA' + BB')^{-1}A + (\pi_t^r)^{-1})^{-1} \\
\rho_{t+1} &= \pi_{t+1}^r(\sigma\sigma' b_t A'(\pi_{ss}^x(\rho_t)AA' + BB')^{-1}\,(s_{t+1} - M - Aa_t + Ab_t\rho_t) + \rho_t(\pi_t^r)^{-1}),
\end{aligned}$$

where K_t^0, and K_t^1, are known quantities at time t and can be found in the Appendix.

3 Model Solution

I solve the model numerically, approximating the logarithm of the continuation-value consumption ratio

$$vc_t = \log\left(\frac{V_t}{C_t}\right).$$

Scaling by consumption and taking logarithms to equation (1), I obtain the following recursion that must be satisfied by the continuation-value consumption ratio:

$$vc_t = \begin{cases} \frac{1}{1-1/\psi} \log\left((1-\beta) + \beta \exp\left[(1-1/\psi)\mathbf{Q}_t\left(vc_{t+1} + \Delta c_{t+1}\right)\right]\right), & \psi \neq 1 \\ \beta \mathbf{Q}_t\left(vc_{t+1} + \Delta c_{t+1}\right), & \psi = 1, \end{cases} \tag{6}$$

where $\mathbf{Q}_t$ is the logarithm of the risk-adjusted expectation operator $\mathbf{R}$; that is,

$$\mathbf{Q}_t(z_{t+1}) = \log \mathbf{R}_t \exp(z_{t+1}) = \begin{cases} \frac{1}{1-\gamma} \log \mathbf{E}\left[\exp((1-\gamma)z_{t+1})|\mathcal{F}_t\right], & \gamma \neq 1 \\ \log \mathbf{E}\left[\exp(z_{t+1})|\mathcal{F}_t\right], & \gamma = 1, \end{cases} \tag{7}$$

As introduced in section 2.3, a set of state variables describes each information structure. I differentiate between two types of states that I will denote by y and π:

$$\begin{aligned} y &= (y^1, y^2, \ldots, y^k) \\ \pi &= (\pi^1, \pi^2, \ldots, \pi^l). \end{aligned}$$

The states y denote, refer either to the level (if observed) or to the mean of a conditional distribution. For example, if the agent observes the x process, x_t would be a y state. Also, m_t would be a y state when the representative agent uses the signals and the Kalman filter to estimate means of x at each point in time.

The state variables denoted by π correspond to the estimates of the conditional variance of the quantity of interest, for example, π_t^r in the case where the agent learns about ρ.

This separation between y and π variables allows me to nest the problems with different information structures. That is, as we go from finer to coarser information sets, we should include the solutions to the former in the latter. To do so, I require that the solution under the richer information set is a special case of the solution under the smaller information set.

The state variables summarize the information of the economy available to the agent at each point in time. Therefore, I follow Judd (1998), to approximate vc by an n^{th}-order complete polynomial of the state variables, with the following form:

$$vc_t \equiv f(y_t, \pi_t; a) = \sum_{\substack{i_1+i_2+\ldots+i_{k+l}\leq n \\ i_1,\ldots,i_{k+l}\geq 0}} a_{i_1,\ldots,i_{k+l}} T_{i_1}(y_t^1)\cdots T_{i_k}(y_t^k)(\pi_t^1)^{i_{k+1}}\cdots(\pi_t^l)^{i_{k+l}}, \tag{8}$$

where $T_l(\cdot)$ denotes to the l^{th}-order general Chebyshev polynomial; π states contribute to the approximated function as a power series, whereas y-states contribute through the Chebyshev polynomials.

The cases where some of the π variables are equal to zero correspond to a solution under a finer information set. Under this formulation, problems about learning unobserved parameters always contain the full-information solution as a special case, that is, when the solution function is evaluated at at point with all the π variables equal to 0. This "boundary" condition is of the following form:

$$vc_t^b \equiv f(y_t, 0; a) = \sum_{\substack{i_1+i_2+\ldots+i_k\leq n \\ i_1,\ldots,i_k\geq 0}} a_{i_1,\ldots,i_k,0,\ldots,0} T_{i_1}(y_t^1)\cdots T_{i_k}(y_t^k). \tag{9}$$

My approach involves calculating the solution when all the variables are observed and then using this solution as the "boundary" condition for more complicated problems. I proceed by adding unobservable quantities one at a time until I have accounted for all. By doing so, I increase parsimoniously the number of parameters as I explore coarser information sets.

This approach shares common ground with finite horizon dynamic programming problems with boundary conditions. I give π state variables the same role that time has in finite horizon problems. A desirable feature of my solution is that in models where agents learn perfectly, the estimates of the volatilities approach zero as information accumulates and the solution under the coarser information set must adapt to converge to the finer information-set solution.

I now describe the parametrization of the continuation-value consumption ratio for each of the information sets, and the restrictions that the nesting approach impose on the function coefficients.

3.1 Case 2: Full-Information

Under full-information, I parameterize the solution to depend on the level of two variables, ρ and x; ρ is not properly a state variable, as it is a parameter assumed to be known, but I treat it as such in order to use this solution as the boundary condition in subsequent problems:

$$f(x, \rho; a^{FI}) = \sum_{i_1=0}^{n} \sum_{i_2=0}^{n-i_1} a^{FI}_{i_1,i_2} T_{i_1}(x) T_{i_2}(\rho). \tag{10}$$

3.2 Case 2: Learning ρ

When ρ is unknown, I use two state variables, ρ_t, and π^r_t, to parameterize its distribution under the agent's information set $\mathcal{F}_t$. So the states in this case are x, ρ, and π^r, and the form of the approximation is

$$f(x, \rho, \pi^r; a^r) = \sum_{i_1=0}^{n} \sum_{i_2=0}^{n-i_1} \sum_{i_3=0}^{n-i_1-i_2} a^r_{i_1,i_2,i_3} T_{i_1}(x) T_{i_2}(\rho) (\pi^r)^{i_3}. \tag{11}$$

I impose the boundary restriction that

$$a^r_{i_1,i_2,0} \doteq a^{FI}_{i_1,i_2}, \text{ for all } i_1, i_2 \in \{0, 1, \ldots, n\} \text{ s.t. } i_1 + i_2 \leq n.$$

3.3 Case 3: Latent Variable

The state variables in this case are the conditional mean (m), and the conditional variance (π^r) of the latent process x. I also parameterize the solution in terms of ρ because I will use this solution as a special case for the coarser information set problem:

$$f(m, \rho, \pi^x; a^x) = \sum_{i_1=0}^{n} \sum_{i_2=0}^{n-i_1} \sum_{i_4=0}^{n-i_1-i_2} a^r_{i_1,i_2,i_4} T_{i_1}(x) T_{i_2}(\rho) (\pi^x)^{i_4}. \tag{12}$$

I impose the boundary restriction that

$$a^x_{i_1,i_2,0} \doteq a^{FI}_{i_1,i_2}, \text{ for all } i_1, i_2 \in \{0, 1, \ldots, n\} \text{ s.t. } i_1 + i_2 \leq n$$

3.4 Case 4: Latent Variable and Learning ρ

In this case I use a, b, π^x, ρ, and π^r as states, and the approximation takes the following form:

$$f(a, b, \pi^x, \rho, \pi^r) = \sum_{\substack{i_1+i_2+i_3+i_4+i_5 \leq n \\ i_1,i_2,i_3,i_4,i_5 \geq 0}} a^{xr}_{i_1,i_2,i_3,i_4,i_5} T_{i_1}(a) T_{i_2}(b) T_{i_3}(\rho) (\pi^x)^{i_4} (\pi^r)^{i_5} \tag{13}$$

Using the previous solutions, I have the following restrictions:

$$\begin{aligned} a^{xr}_{i_1,i_2,i_3,0,0} &= a^{FI} \\ a^{xr}_{i_1,i_2,i_3,0,i_5} &= a^{r} \\ a^{xr}_{i_1,i_2,i_3,i_4,0} &= a^{x}. \end{aligned}$$

4 Data and Model Estimation

I use aggregate consumption of nondurables and services and corporate earnings from the National Income and Product Accounts. The data have quarterly frequency and spans from the first quarter of 1947 to the second quarter of 2009. I use the data to construct the time series of the logarithm of consumption growth and the logarithm of the earnings to consumption ratio. The series are seasonally adjusted and in real terms, and are deflated by the GDP deflator, also from the National Accounts.

I choose the signal in equation (5) to be the logarithm of corporate earnings divided by aggregate consumption. The work of Lettau and Ludvigson

(2001) and Santos and Veronesi (2005) motivates this measure of the economic environment that the work of Hansen, Heaton, and Li (2008) proved to be relevant in identifying long-run risk. The time series of the earnings-consumption ratio presents a high degree of persistence that would be crucial in identifying the latent process x.

I estimate the parameters and moments of the latent variable in the state-space system (4–5) using MCMC methods. In particular, I use the Gibbs sampler (see, for example, Johannes and Polson 2009, Gamerman and Lopes 2006, Rossi, Allenby, and McCulloch 2006). As my system is linear with Gaussian errors I choose the natural normal inverse Wishart conjugate priors.

In order to be able to identify the latent process, I restrict the coefficients of the system in the following way:

$$x_{t+1} = \rho x_t + (\sigma_x, 0, 0)\,\varepsilon_{t+1} \tag{14}$$

$$\Delta c_{t+1} = M_c + x_t + (0, B_c, 0)\,\varepsilon_{t+1} \tag{15}$$

$$e_{t+1} - c_{t+1} = M_s + A_s x_t + (0, B_{c,s}, B_s)\,\varepsilon_{t+1}, \tag{16}$$

where the error process ε_{t+1} has a three-dimensional normal distribution with mean 0 and covariance matrix equal to the identity. The innovations to the latent process are assumed, therefore to be orthogonal to the rest of the signals, and I allow for the signals to be conditionally correlated as well. I use normal priors for A, M, and ρ, inverse gamma for σ_x, and inverse Wishart for B.

The Gibbs sampler in this system works in a straightforward manner.

Start the iteration with an initial set of parameters and a realization of the latent process. Draw from the distribution of each of the parameters and the latent process, conditional on the observed signals and the rest of the unknown quantities. I use those samples as conditioning variables in the next draw to generate a new sample. I iterate this process until convergence.

To sample from the conditional distribution of the latent variable, I use the Forward Filtering Backward Sampling (FFBS) algorithm. FFBS, is a multiple-sampling method proposed simultaneously by Frühwirth-Schnatter (1992) and Carter and Kohn (1994), that provides a sequential method to sample from the smoothed distribution of $x^T = (x_1, \ldots, x_T)$.

To understand how FFBS works, first note that from standard Kalman filter results, the conditional distribution of $x_t|s^t, \Theta$ (for convenience I denote the vector of parameters by Θ) is normal with mean m_t and variance v_t. The most important insight from this filtering technique is noticing that the probability density function $\pi(x_t|x_{t+1}, s^T, \Theta)$ for $t = 1, \ldots, T-1$ satisfies the following:

$$\begin{aligned}
\pi(x_t|x_{t+1}, s^T, \Theta) &= \pi(x_t|x_{t+1}, s^{t+1}, \Theta) \propto \pi(x_{t+1}|x_t, \Theta)\pi(x_t|s^{t+1}, \Theta) \\
&\propto \pi(x_{t+1}|x_t, \Theta)\pi(s_{t+1}|x_t, \Theta)\pi(x_t|s^t, \Theta),
\end{aligned}$$

where the dynamic structure of the problem justifies the equality. All relevant information that future signals carry about x_t is redundant once x_{t+1} is known. Note also that the timing of the model makes s_{t+1} dependent on x_t. The proportionality relations come from applying Bayes theorem twice, first to $(x_t|x_{t+1}, s^{t+1}, \Theta)$ and then to $(x_t|s^{t+1}, \Theta)$, to make use of the filtered distribution $(x_t|s^t,\Theta)$ obtained in the previous step.

Therefore, the distribution of x_t given x_{t+1}, the history of signals, and the parameters is also normal with mean m_t^s and variance v_t^s,

$$
\begin{aligned}
v_t^s &= (A'(BB')^{-1}A + v_t^{-1} + \rho^2\sigma_x^{-2})^{-1} \\
m_t^s &= (v_t^s)^{-1}\left((s_{t+1}-M)'A'(BB')^{-1}A(s_{t+1}-M) + m_t v_t^{-1} + \rho x_{t+1}\sigma_x^{-2}\right).
\end{aligned}
$$

The conditional distribution of the parameters, given the history of signals and the history of x, follows the standard conjugate prior calculations (see for example Gamerman and Lopes 2006). The parameter estimates, corresponding to diffuse priors, are in Table 1 and the posterior histograms and convergence plots are in figures 1 and 2. Figure 3, plots the filtered hidden process.

5 Empirical Results

In order to evaluate the effect that different information sets have on asset prices, I compare maximum Sharpe ratios (the slope of the mean-standard deviation frontier) across economies. I focus on this measure, because, as noted by Hansen and Sargent (2009), different risks are priced under different information sets, and therefore risk prices are not directly comparable across information structures.

As Hansen and Jagannathan (1991) show, the standard deviation of the one-period stochastic discount factor divided by its expected value represents an upper bound for the Sharpe ratio of the economy. Therefore, I use the form of the stochastic discount factor from equation (3) and my solution for the continuation-value to calculate the required standard deviation and expected value. With these quantities in hand, I calculate the one-period maximum Sharpe ratio. I also compute the time series of one-period risk-free rates as the reciprocal of the expected value of the one-period stochastic discount factor.

I calibrate the preference parameters to provide plausible levels for the average risk free rate and the maximum Sharpe ratio of the model with learning about persistence and conditional mean of consumption growth. I choose, risk-aversion parameter to 4.5, IES to 1.5, and the time-preference parameter to .995. These give me an average one-year risk-free rate of 5% and a maximum Sharpe ratio just below .5. In order to obtain comparable results I keep these parameters constant across information structures.

As the agent updates her beliefs about the unknown quantities, I need to specify prior distributions. The prior for ρ is a truncated normal centered at .9671, the mean of the posterior distribution of ρ obtained in section 4 and the prior standard deviation is .0165, the standard deviation of the same posterior. The prior for the hidden process at time zero is normal with mean 0 and the Kalman filter steady-state standard deviation from equation (21) that is 4.2810^{-4}.

Figure 4 plots time series of maximum Sharpe-ratios implied by the state variables filtered by the agent under each information set. I note that being uncertain about both, expected consumption growth, and its persistence has a significant effect on the risk-return trade-off. This uncertainty increases maximum Sharpe ratios to .45, from .07 in the economy without learning.

I also find that uncertainty only about persistence of consumption growth, leads to a maximum Sharpe ratio around .16. While in the case when only expected consumption growth is unobserved my measure of the risk-return trade-off is .08. Therefore, when the agent learns about two quantities that are closely related, the uncertainty compounds producing sensibly higher risk prices, compared to the cases of learning about each quantity separately.

Figure 5 plots the time-series of annualized one-period risk-free rates. The effect that learning has on Sharpe ratios translates to risk-free rates, the learning economies see an increase on the risk-free rate with respect to the full-information benchmark. I also notice the compounding effect, with a significant increase in the rate when both quantities, expected consumption growth and its persistence, are unobserved.

Finally, I would like to note that the significant impact that learning has on risk prices is due to two factors, recursive preferences and the compound effect of dealing with two related unobserved quantities. As noted in Ai (2007), recursive preferences contribution comes through the separation between risk aversion and intertemporal elasticity of substitution. This separation makes risk prices to increase in the presence of learning. The compounded effect of learning about two related hidden quantities is the reason why I document a larger increase in prices of risk, than Ghosh (2008) and Ai (2007). These authors study learning problems in economies similar to mine and focus on the effect of learning about a single quantity in isolation.

My work can be applied to solve learning problems with non-linear dynamics, and opens the door to quantifying the complete learning premium in more realistic economies where the agents have the same information as the econometrician.

6 Conclusion

The contribution of this paper is three-fold. First, I quantify a sizeble effect of learning on risk prices. I solve a learning problem where both, expected consumption growth and its persistence are unobserved, and I show that this uncertainty has a significant effect for risk-prices, pushing maximum Sharpe ratios to .45 from .07 in an economy without learning. I also show that the compounding effect of learning about this two quantities is sizably larger than considering the learning problems separately.

Second, I propose a method to parameterize the agent's continuation-value, when she solves a learning problem. I suggest using sufficient statistics to summarize the conditional distribution of hidden processes and unknown variables. This approach is inspired by recent work in particle learning from Carvalho et al. (2009) and avoids using a discretized parameter space, as is common in the literature.

Third, the paper develops a new numerical solution for models with learning about parameter uncertainty. I solve a set of nested dynamic programming problems, taking advantage of the fact that information sets are naturally nested. My approach has the spirit of dynamic programming with a boundary condition. I use the solution to a problem with a finer information set as the "boundary" condition for problems with coarser information sets. To do this, I give to the volatility of the unknown quantity the same role that time has in dynamic programming problems with boundary conditions. I impose the restriction that, as the volatility approaches zero, the solution of the problem with the coarser information set converges to the solution of the finer information set. This reduces the dimensionality of the problem and ensures consistent solutions across information sets.

My work can be applied to solve learning problems with non-linear state evolutions and opens the door to quantify the complete learning premium in economies where the agents and the econometrician have the same information.

References

Ai, Hengjie. 2007. "Information about Long-Run Risk: Asset Pricing Implications." Working paper, Duke University. Fuqua School of Business. Forthcoming in *The Journal of Finance*.

Bakshi, G., and G. Skoulakis. 2009. "Do Subjective Expectations Explain Asset Pricing Puzzles?" Working paper, University of Maryland.

Bansal, Ravi, and Amir Yaron. 2004. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *The Journal of Finance* 59 (4): 1481–1509.

Carter, C. K., and R. Kohn. 1994. "On Gibbs Sampling for State Space Models." *Biometrika* 81 (3): 541–53.

Carvalho, Carlos M., Michael Johannes, Hedibert F. Lopes, and Nicholas Polson. 2009. "Particle Learning and Smoothing." Working paper, University of Chicago and Columbia University.

Epstein, Larry G., and Stanley E. Zin. 1989. "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework." *Econometrica* 57 (4): 937–69.

Frühwirth-Schnatter, Sylvia. 1992. "Data Augmentation and Dynamic Linear Models." Dissertation, Institut für Statistik. Wirtschaftsuniversität Wien.

Gamerman, Dani, and Hedibert F. Lopes. 2006. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. London, UK: Taylor & Francis.

Ghosh, Anisha. 2008. "Asset Pricing with Regime Shifts in Consumption and Dividend Growth." Job market paper, London School of Economics.

Hansen, Lars P., John C. Heaton, Junghoon Lee, and Nikolai Roussanov. 2007. "Chapter 61 Intertemporal Substitution and Risk Aversion." In *Handbook of Econometrics*, edited by James J. Heckman and Edward E. Leamer, Volume 6, Part 1 of *Handbooks in Economics*, 3967 – 4056. Oxford, UK: Elsevier.

Hansen, Lars P., John C. Heaton, and Nan Li. 2008. "Consumption Strikes Back? Measuring Long-Run Risk." *Journal of Political Economy* 116 (2): 260–302.

Hansen, Lars P., and Thomas J. Sargent. 2009. "Fragile Beliefs and the Price of Model Uncertainty." Working paper, University of Chicago and New York University.

Hansen, Lars Peter, and Ravi Jagannathan. 1991. "Implications of Security Market Data for Models of Dynamic Economies." *Journal of Political Economy* 99 (2): 225–62.

Johannes, Michael, and Nicholas Polson. 2009. "Chapter 13 MCMC Methods for Continuous-Time Financial Econometrics." In *Handbook of Financial Econometrics*, edited by Yacine Aït-Sahalia and Lars P. Hansen, Handbooks in Finance, 1 – 72. Oxford, UK: Elsevier.

Judd, Kenneth L. 1998. *Numerical Methods in Economics.* Cambridge, MA: The MIT Press.

Kreps, David M., and Evan L. Porteus. 1978. "Temporal Resolution of Uncertainty and Dynamic Choice Theory." *Econometrica* 46 (1): 185–200.

Lettau, Martin, and Sydney Ludvigson. 2001. "Consumption, Aggregate Wealth, and Expected Stock Returns." *Journal of Finance* 56 (3): 815–49.

Pakoš, Michal. 2008. "Asset Prices under Doubt about Fundamentals." Working paper, Tepper School of Business at Carnegie Mellon.

Rossi, Peter E., Greg M. Allenby, and Rob McCulloch. 2006. *Bayesian Statistics and Marketing.* Wiley Series in Probability and Statistics. DeKalb, IL: John Wiley & Sons Inc.

Santos, Tano, and Pietro Veronesi. 2005. "Labor Income and Predictable Stock Returns." *Review of Financial Studies* 19 (1): 1–44.

Veronesi, Pietro. 1999. "Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model." *Review of Financial Studies* 12 (5): 975–1007.

———. 2000. "How Does Information Quality Affect Stock Returns?" *Journal of Finance* 55 (2): 807–837.

Weil, Philippe. 1989. "The Equity Premium Puzzle and the Risk-free Rate Puzzle." *Journal of Monetary Economics* 24 (3): 401–21.

A Appendix

A.1 Learning

Bellow I present some calculations for the evolution of the sufficient statistics in the different information sets.

A.1.1 Latent variable. The Kalman Filter

In the case that the conditional expected value of consumption growth is not observed, I use the Kalman filter to obtain the evolution of the states. Consider the system (4–5). Suppose, the agent knows the model and the parameters and does not observe the latent process x. Suppose further, I start at period $t+1$ with knowledge of a filtered distribution from period t, $x_t|s^t \sim N(m_t, \pi_t^x)$, then using the system equations I have

$$\left.\begin{pmatrix} x_{t+1} \\ s_{t+1} \end{pmatrix}\right| s^t \sim N\left(m_t \begin{pmatrix} \rho \\ A \end{pmatrix}, \begin{pmatrix} \rho^2\pi_t^x + \sigma_x\sigma_x' & \rho A'\pi_t^x + \sigma_x B' \\ \rho A\pi_t^x + B\sigma_x & AA'\pi_t^x + BB' \end{pmatrix}\right) \tag{17}$$

the properties of the multivariate normal implies that $x_{t+1}|s^{t+1}$ is also normal, let's denote it's mean and variance by m_{t+1} and π_{t+1}^x

$$K_t = \left(\rho A\pi_t^x + B\sigma_x'\right)'\left(AA'\pi_t^x + BB'\right)^{-1} \tag{18}$$

$$\pi_{t+1}^x = \rho^2\pi_t^x + \sigma_x\sigma_x' + K_t\left(\rho A\pi_t^x + B\sigma_x'\right) \tag{19}$$

$$m_{t+1} = \rho m_t + K_t\left(s_{t+1} - M - Am_t\right) \tag{20}$$

Equation (19), describes the evolution of the variance of the hidden process conditioned on the signal history. Iterating this equation I can compute the steady state variance, that satisfies the following equation

$$\pi_{ss}^x = \rho^2 \pi_{ss}^x + \sigma_x \sigma_x' + \left(\rho A \pi_{ss}^x + B\sigma_x'\right)' \left(AA'\pi_{ss}^x + BB'\right)^{-1} \left(\rho A \pi_{ss}^x + B\sigma_x'\right) \tag{21}$$

Substituting the solution to equation (21) on equations (18) and (20) I obtain the steady state Kalman gain, and evolution of the mean of the filtered distribution in the steady state

$$K_{ss} = \left(\rho A \pi_{ss}^x + B\sigma_x'\right)' \left(AA'\pi_{ss}^x + BB'\right)^{-1} \tag{22}$$

$$m_{t+1}^{ss} = \rho m_t^{ss} + K_{ss}\left(s_{t+1} - M - A m_t^{ss}\right) \tag{23}$$

A.1.2 Learning ρ

The agent has a truncated normal prior distribution over ρ, at time t

$$\rho|\mathcal{F}_t \sim N\left(\rho_t, \sigma_x^2 \pi_t\right) \text{ and } \rho \in (-1,1) \tag{24}$$

Notice that I can rewrite equation (14) to get

$$x_{t+1} = \rho_t x_t + \sigma_x x_t \sqrt{\pi_t}\left(\frac{\rho - \rho_t}{\sigma_x \sqrt{\pi_t}}\right) + (\sigma_x, 0, 0)\varepsilon_{t+1}$$

So the uncertainty about the parameter ρ introduces heteroscedasticity in the "subjective" distribution of the state x. This implies the agent faces the following conditional distribution of evolution of x_{t+1}

$$x_{t+1}\,|\mathcal{F}_t \sim N\left(\rho_t x_t, \sigma_x \sqrt{x_t^2 \pi_t + 1}\right) \tag{25}$$

Using Bayes theorem

$$p(\rho|x^{t+1}) \propto p(x_{t+1}|\rho, x^t) p(\rho|x^t)$$

where the terms on the right hand side are Gaussian and truncated Gaussian, and therefore the posterior distribution $\rho|\mathcal{F}_{t+1}$ is also truncated Gaussian. Writing down the distributions in equation A.1.2 using the prior and 25, I complete the square to get

$$\begin{aligned}
\pi^r_{t+1} &= (x_t^2 + (\pi^r_t)^{-1})^{-1} \\
\rho_{t+1} &= \pi^r_{t+1}(\rho_t (\pi^r_t)^{-1} + x_{t+1} x_t).
\end{aligned}$$

A.1.3 Latent variable and learning ρ

In this problem I consider a hierarchical priors for the unknown quantities so

$$\rho|s^t \sim N(\rho_t, \pi^r_t) \text{ and } \rho \in (-1, 1) \tag{26}$$

$$x_t|s^t, \rho \sim N(m_t(\rho), \pi^x_t(\rho)) \tag{27}$$

where $m_t(\rho)$ and $\pi^x_t(\rho)$ follow the Kalman Filter evolutions. In order to obtain closed-form equations for the evolution of the states I approximate

$\pi_t^x(\rho)$ at the steady state variance at ρ_t. That is

$$\pi_t^x(\rho) \approx \pi_{ss}^x(\rho_t)$$

where $\pi_{ss}^x(\rho_t)$ satisfies the steady state equation for the variance of the Kalman filter (21)

$$\pi_{ss}^x(\rho_t) = \rho_t^2 \pi_{ss}^x(\rho_t) + \sigma_x \sigma_x' + \left(\rho_t A \pi_{ss}^x(\rho_t) + B\sigma_x'\right)' \left(AA'\pi_{ss}^x(\rho_t) + BB'\right)^{-1} \left(\rho A \pi_{ss}^x(\rho_t) + B\sigma_x'\right)$$

with this approximation of the variance, the Kalman gain is linear in ρ. From equation (18) the Kalman gain is now

$$\begin{aligned}
K_t(\rho) &= \left(\rho A \pi_{ss}^x(\rho_t) + B\sigma_x'\right)' \left(AA'\pi_{ss}^x(\rho_t) + BB'\right)^{-1} \\
&= \left(\rho_t A \pi_{ss}^x(\rho_t) + B\sigma_x'\right)' \left(AA'\pi_{ss}^x(\rho_t) + BB'\right)^{-1} + A' \left(AA'\pi_{ss}^x(\rho_t) + BB'\right)^{-1} \pi_{ss}^x(\rho_t)(\rho - \rho_t) \\
&= K_t^0 + K_t^1(\rho - \rho_t)
\end{aligned}$$

Then, from equation (20), and taking a first order approximation of $m_t(\rho)$ centered around ρ_t, ($m_t(\rho) \approx a_t + b_t(\rho - \rho_t)$) I have

$$\begin{aligned}
a_{t+1} + b_{t+1}(\rho - \rho_{t+1}) &= \rho(a_t + b_t(\rho - \rho_t)) + K_t(\rho)(s_{t+1} - M - Aa_t + b_t(\rho - \rho_t)) \\
&\approx a_t(\rho_t - K_t^0 A) + K_t^0(s_{t+1} - M) + \\
&\quad + [\rho_t - K_t^0 A + (1 - K_t^1 A)a_t + K_t^1(s_{t+1} - M)](\rho - \rho_t)
\end{aligned}$$

The previous equation give me evolutions for the states a_t and b_t.

Under the approximations described above I have the following evolution for the sufficient statistics that control the distribution of the observable signals given the information set

$$s_{t+1}|\rho, s^t \sim N\left(M + A\left(a_t + b_t(\rho - \rho_t)\right), \pi^x_{ss}(\rho_t)AA' + BB'\right)$$

so, combining this with the prior, using Bayes theorem, I have that the posterior distribution of ρ is also truncated normal with parameters

$$\begin{aligned}
\pi^r_{t+1} &= (\sigma\sigma' b_t^2 A'(\pi^x_{ss}(\rho_t)AA' + BB')^{-1}A + (\pi^r_t)^{-1})^{-1} \\
\rho_{t+1} &= \pi^r_{t+1}(\sigma\sigma' b_t A'(\pi^x_{ss}(\rho_t)AA' + BB')^{-1}\left(s_{t+1} - M - Aa_t + Ab_t\rho_t\right) + \rho_t(\pi^r_t)^{-1}),
\end{aligned}$$

A.2 Numerical approach

I approximate the continuation value consumption ratio by a function of the states $f(y, \pi; a)$, as in equation (8)

$$f(y, \pi; a) = \sum_{\substack{i_1+i_2+\ldots+i_{k+l}\leq n \\ i_1,\ldots,i_{k+l}\geq 0}} a_{i_1,\ldots,i_{k+l}} T_{i_1}(y^1)\cdots T_{i_k}(y^k)(\pi^1)^{i_{k+1}}\cdots(\pi^l)^{i_{k+l}},$$

where y and π are vectors that contain the two types of states (see section 3) and a is a vector of coefficients. The objective is to find a coefficients a, such that the equilibrium condition (6) and the boundary condition are satisfied. The boundary condition implies that

$$f(y, 0; a) = f_b(y; a_b),$$

and the boundary solution also satisfies the equilibrium condition (6) for a problem with a finer information set.

I therefore choose coefficients that minimize the discrepancy between both sides of condition (6) over a set of grid points in the states. Therefore, I will approximate two functions, one for the boundary problem and then I use the solution at the boundary to solve augmented problem.

I construct a grid for y, $y_1, \ldots, y_{M_y}$ where M_y is the number of grid points in the y variables. and another grid for π with M_π, $\pi_1, \ldots, \pi_{M_\pi}$. Then, the total number of grid points is $M = M_y M_\pi$.

First, I solve the boundary problem. That can be written in the following way

$$
\begin{aligned}
\min_{a_b,\lambda} \quad & \sum_{i=1}^{M_y} \lambda_i \\
\text{s.t.} \quad & \\
& -\lambda_i \leq B_b(y_i; a_b) \leq \lambda_i, \\
& \lambda_i \geq 0 \\
& \text{for } i = 1, \ldots, M_y
\end{aligned}
$$

where a_b are the coefficients of the boundary function and B_b is a measure of discrepancy between the two sides of (6), that is

$$
B_b(y; a_b) \doteq \begin{cases} (1 - 1/\psi) f_b(y; a_b) - \log\left((1-\beta) + \beta \exp\left[(1 - 1/\psi)\mathbf{Q}\left(f_b(y_+; a_b) + \Delta c_+\right)\right]\right), & \psi \neq 1 \\ f_b(y; a_b) - \beta \mathbf{Q}\left(f_b(y_+; a_b) + \Delta c_+\right), & \psi = 1, \end{cases} \tag{28}
$$

The operator $\mathbf{Q}$ is a distorted conditional expectation defined in 7, y_+ and Δc_+ represent the next "period" realization of the states and consumption growth. Then, conditioning on current information (y), I calculate $\mathbf{Q}$ using the conditional the conditional distributions derived on section 2.3. I compute $\mathbf{Q}$ numerically using quadrature methods. As the conditional distributions of the states and consumption growth are normal, I follow Judd (1998) and use Gauss-Hermite quadrature.

Second, I proceed similarly to solve the full problem, although now, I impose that the boundary condition is satisfied. As explained in section 3, I restrict the coefficients of the augmented problem, so that when $\pi = 0$ the solution satisfies the boundary condition. To write down the new minimization problem, parameterize the vector of coefficients as follows

$$a = (a_b, a_{nb})$$

where a_b are the coefficients at the boundary obtained in the previous step and a_{nb} are the parameters that only have influence if $\pi > 0$. With this parametrization I solve the following minimization problem

$$\begin{aligned}
\min_{a_{nb},\lambda} \quad & \sum_{i=1}^{M_y} \sum_{j=1}^{M_\pi} \lambda_{i,j} \\
\text{s.t.} \quad & \\
& -\lambda_{i,j} \leq B(y_i, \pi_j; (a_b, a_{nb})) \leq \lambda_{i,j}, \\
& \lambda_{i,j} \geq 0, \\
& \text{ for } i = 1, \ldots, M_y \text{ and } j = 1, \ldots, M_\pi
\end{aligned}$$

where B is also a scalar function that measures discrepancy between the left-hand-side and right-hand-side of the equilibrium condition (6). The only differences is that the function evaluated is that of the augmented problem and not that of the boundary, and that the expectations are taken under the coarser information set of the augmented problem.

A.2.1 Precision

In the t grid I use is of size $M = 5,000$, and the rule is to use 10 nodes for y variables and 5 for π variables. I use Chebychev and power polynomials of order 3 and I approximate each integral using 25 quadrature nodes. To test the accuracy of my approximation method, I compute a unit-less discrepancy measure outside the grid points. The highest errors were of the order of 10^{-3} in the case where I learn about ρ and x. For the full information economy I obtain error measures of the order of 10^{-7}.

B Tables and Figures

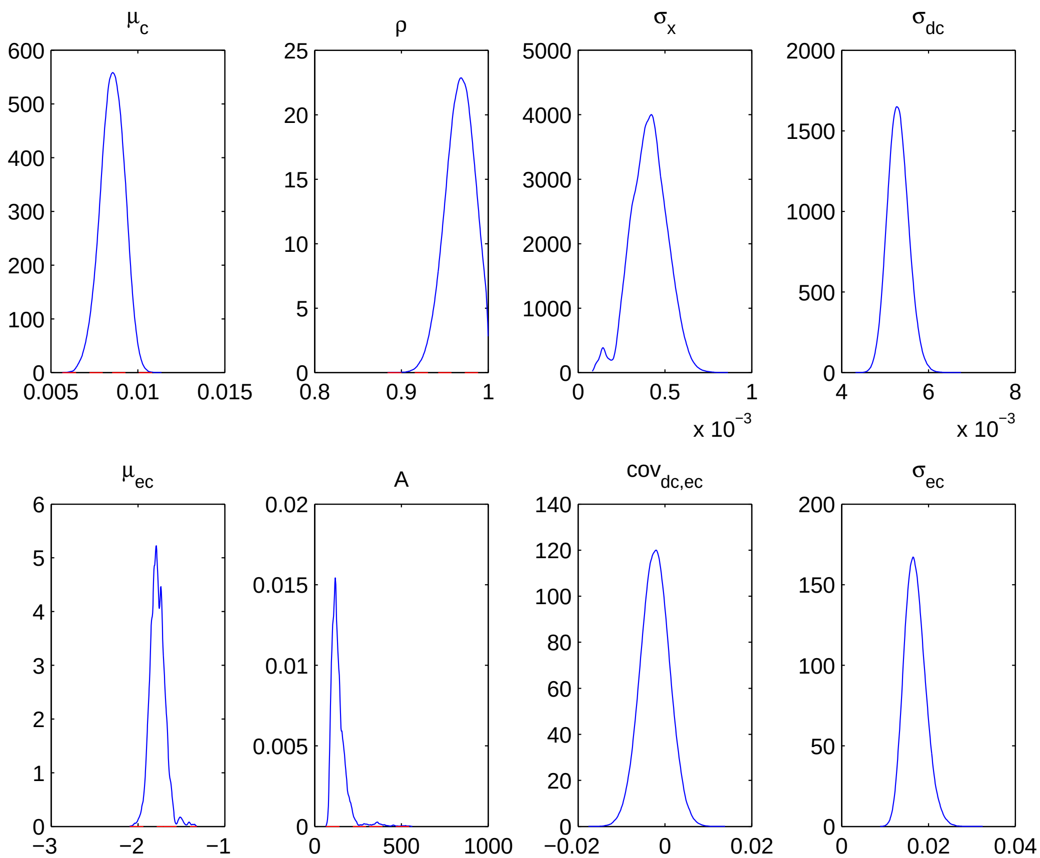

Figure 1: Parameter histograms for the model in equations (14) through (16). These estimates are based on 400,000 Gibbs sampler iterations with a burn-in period of 65,000.

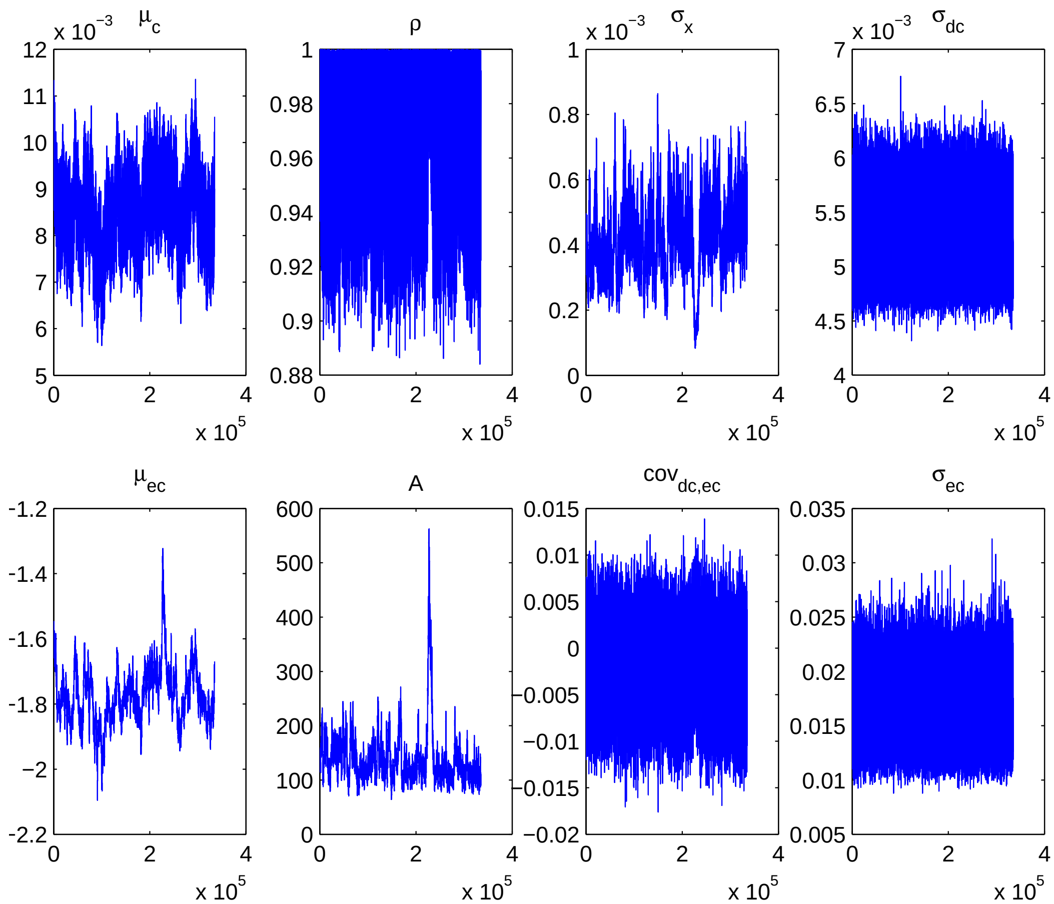

Figure 2: MCMC convergence plot for the parameters of the model in equations (14) through (16). These estimates are based on 400,000 Gibbs sampler iterations with a burn-in period of 65,000.

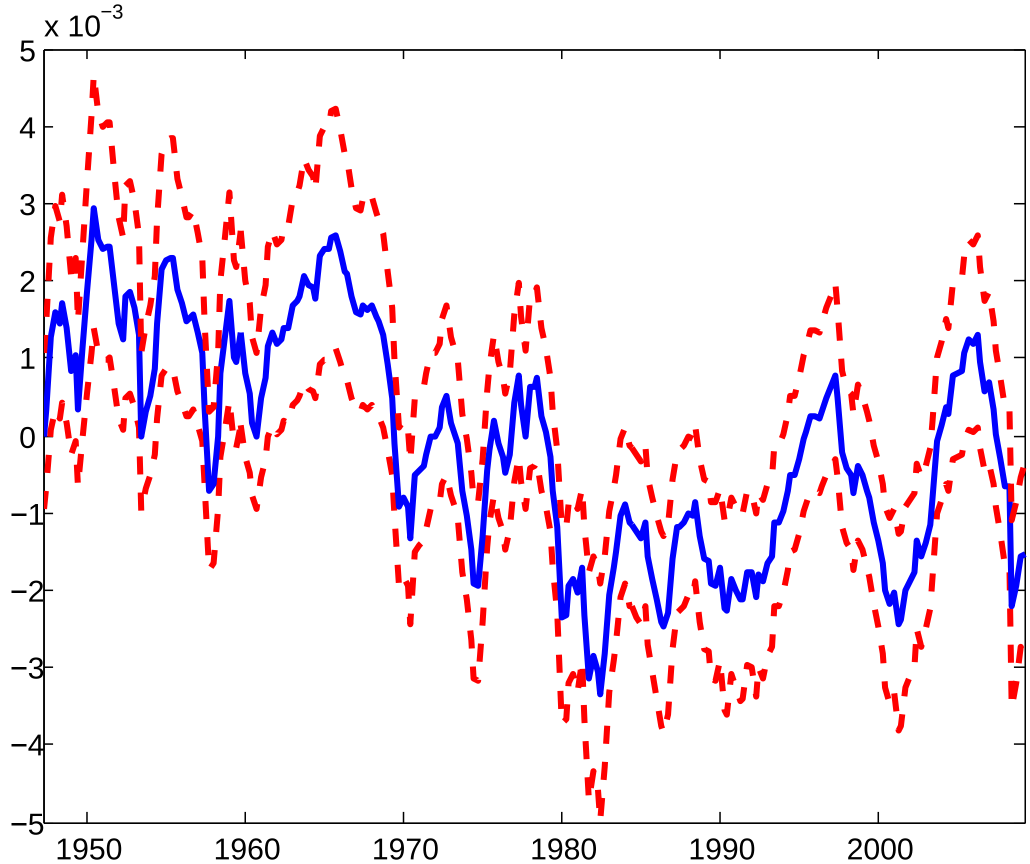

Figure 3: Median, 95, 5 percentiles estimates for the latent process of the model in equations (14) through (16). These estimates are based on 400,000 Gibbs sampler iterations with a burn-in period of 65,000.

	μ_c	ρ	σ_x	σ_{dc}	μ_{ec}	A	$corr_{dc,ec}$	σ_{ec}
mean	0.0085	0.9671	0.00041	0.0053	-1.7742	137.93	-0.1424	0.0164
st. deviation	0.0007	0.0165	0.00010	0.0002	0.0873	50.25	0.2086	0.0024
5th percentile	0.0074	0.9384	0.00025	0.0049	-1.9034	90.67	-0.4913	0.0128
median	0.0085	0.9679	0.00041	0.0053	-1.7799	126.27	-0.1386	0.0163
95th percentile	0.0096	0.9930	0.00058	0.0057	-1.6350	210.01	0.1947	0.0207

Table 1: Parameter estimates for the model in equations (14) through (16). These estimates are based on 400,000 Gibbs sampler iterations with a burn-in period of 65,000. The prior is plotted in red and the posterior in blue

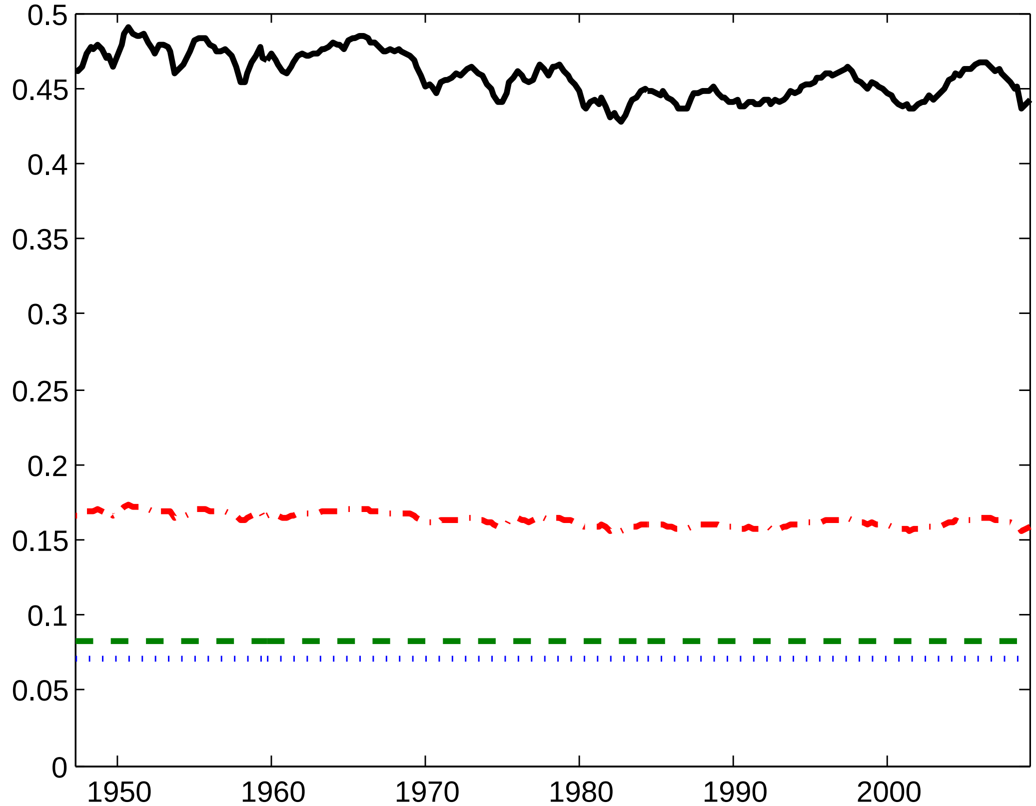

Figure 4: Time series of annualized maximum Sharpe ratios implied by the model for each information set. From the bottom to the top we have, full-information (dotted line), latent conditional mean (dashed line), learning ρ (dashed dotted), and combination of learning ρ and latent conditional mean (solid line).

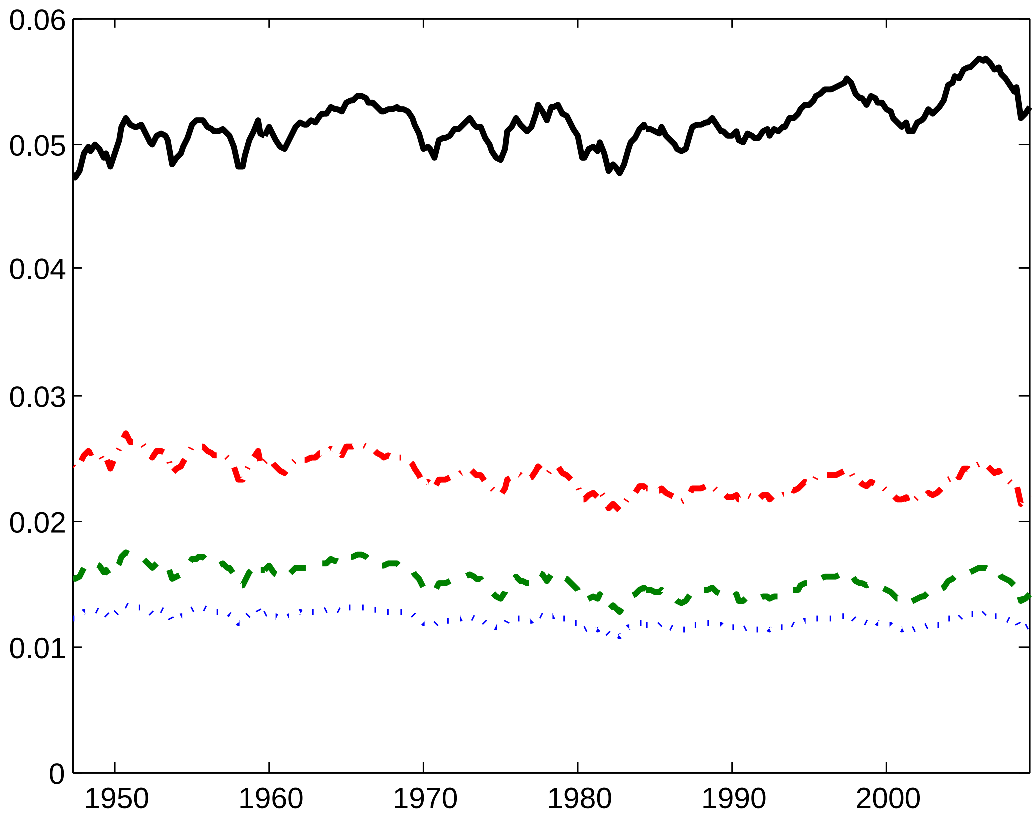

Figure 5: Time series of annualized risk-free rates implied by the model for each information set. From lower to higher we have, full-information (dotted line), latent conditional mean (dashed line), learning ρ (dashed dotted line) and combination of learning ρ, and latent conditional mean (solid line).