Efficient Estimation of Learning Models

Laurent Calvet and Veronika Czellar*

February 2011

Abstract

This paper develops a toolkit of inference and forecasting methods for a large class of nonlinear incomplete-information models. These methods apply most directly to environments with three levels of information: a Markov chain driving fundamentals, a representative agent receiving partially revealing signals about the Markov chain, and an econometrician observing the data generated by the agent. We provide a convenient and efficient method, based on indirect inference (Gouriéroux, Monfort and Renault 1993; Smith 1993), for the parameters specifying fundamentals, the signals, and the data. We develop a particle filter to recursively estimate the joint distribution of fundamentals and the agent's belief about fundamentals at each point in time. We also propose a particle filter-based test of a dynamic moment condition involving the hidden state. The good empirical performance of these methods is demonstrated on the multifrequency asset pricing model of Calvet and Fisher (2007) applied to over 80 years of daily aggregate equity excess returns.

Keywords: Learning, indirect inference, efficient method of moments, forecasting, particle filter, value at risk.

JEL Classifications: C10, C12, C15, C22, C14.

*Calvet: Department of Finance, HEC Paris, 1 rue de la Libération, 78351 Jouy en Josas, France; and NBER, calvet@hec.fr. Czellar: Department of Economics and Decision Sciences, HEC Paris, 1 rue de la Libération, 78351 Jouy en Josas, France, czellarv@hec.fr. We received helpful comments from Pavel Chigansky, Thierry Foucault, Christian Gouriéroux, Hashem Pesaran, Andrew Siegel, Ramon van Handel, and seminar participants at CORE, the University of Evry, the Second HEC Finance and Statistics Conference, the 16th International Conference on Computing in Economics and Finance in London, the 2010 Toulouse School of Economics Financial Econometrics Conference, the 2010 International Symposium on Business and Industrial Statistics in Portoroz, and the 2010 Econometric Society World Congress in Shanghai. We gratefully acknowledge the computer support of EUROFIDAI and the financial support of the American Statistical Association, the Europlace Institute of Finance, and the National Institute of Statistical Sciences.

1 Introduction

In asset and good markets, decision-makers do not directly observe the state of fundamentals, but must infer it from available economic and financial data. A growing literature shows that their sequential learning process has profound implications for security valuation and macroeconomic performance.[1] For instance, in financial economics, learning is used to explain phenomena as diverse as portfolio choice (Brennan 1998), the level of equity prices (Pastor and Veronesi 2009), time variations in volatility (Veronesi 1999), or the skewness of equity returns (Calvet and Fisher 2007).

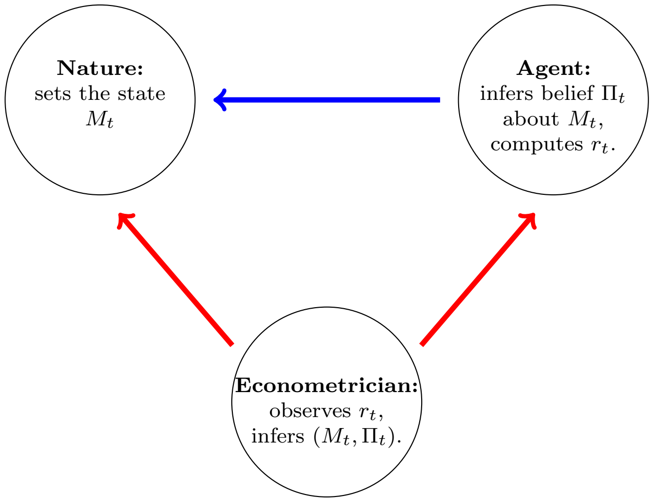
Incomplete-information models are typically calibrated, often at low frequencies. In one branch of the literature, agents are assumed to exhibit constant absolute risk aversion (CARA) and receive Gaussian signals, which permits the use of the Kalman filter (e.g. Wang 1994). These models, which provide valuable theoretical insights on the implications of learning, cannot be easily brought to the data because CARA utility is not a plausible description of investor behavior.[2] Another branch of the literature obtains tractability by combining Markov-switching fundamentals with more realistic preferences, such as constant relative risk aversion (CRRA), Epstein-Zin-Weil utility, or disappointment aversion (e.g. Bonomo and Garcia, 1991, 1994, 1996; Lettau, Ludvigson and Wachter, 2008). The econometrics of these incomplete-information economies, however, remains largely unexplored.

In this paper, we develop a toolkit of estimation, inference and forecasting methods for a large class of nonlinear learning models. These techniques readily apply to representative-agent economies with three levels of information, as illustrated in Figure 1. Fundamentals are driven by a latent Markov-switching *state of nature* M_t. The representative agent receives a partially revealing signal about M_t and uses Bayes' rule to recursively compute the conditional probability distribution Π_t over states of nature. The econometrician observes a data point r_t, which is a function of the current and possibly lagged values of the *state of the economy* $s_t = (M_t, \Pi_t)$. When the agent is fully informed about the underlying M_t, the state space is finite and the likelihood function is available analytically. Learning models generally do not have a closed-form likelihood but are easy to simulate, which serves as the basis for inference.

[1] See for instance Brennan (1998), Brennan and Xia (2001), David (1997), Guidolin and Timmermann (2003), Lettau, Ludvigson and Wachter (2008), Pastor and Veronesi (2009), Timmermann (1993, 1996), Van Nieuwerburgh and Veldkamp (2006), and Veronesi (1999, 2000).

[2] See for instance the discussion in Campbell and Viceira (2002, ch. 6).

Figure 1: INFORMATION STRUCTURE



We develop three main tools. First, we provide an estimation procedure, based on indirect inference (II), that closely parallels the investigation of learning models. II is a simulation-based method introduced by Smith (1993) and Gouriéroux, Monfort and Renault (1993), which has gained increasing importance in recent years (e.g. Calzolari, Fiorentini and Sentana 2004; Czellar, Karolyi and Ronchetti 2007; Czellar and Ronchetti 2010; Dridi, Guay and Renault 2007; Heggland and Frigessi 2004; Sentana, Calzolari and Fiorentini 2008). We define an auxiliary estimator by extending the maximum likelihood estimator (MLE) of the full-information (FI) model with a set of auxiliary sample statistics. We choose auxiliary statistics that quantify features of the data, such as skewness or kurtosis, that the incomplete information structure is designed to capture. For any choice of the structural parameter, we can simulate a sample path from the learning model and compute the corresponding auxiliary estimate. The II estimator is chosen so that the simulated auxiliary estimator matches as closely as possible the auxiliary estimator computed from the data.

Second, we develop a particle filter to recursively estimate the joint distribution of the latent state M_t and the representative agent's conditional distribution (or "belief") Π_t at every date t. The definition of the filter incorporates the joint dynamics of the state of nature and the signal received by the agent, as well as her learning process. Furthermore, the filter allows us to estimate the likelihood function of the estimated process, implement model selection methods, and provide forecasts.

Third, we propose a particle filter-based test of a moment condition involving the hidden state. Earlier research shows that in a large class of filters, the sample moments of the particles at a given point in time are asymptotically normal, and that the asymptotic variance is provided by a recursive formula that is generally challenging to estimate (Chopin, 2004). In this paper, we develop a simulation-based test of a moment condition at every point in time, which is shown to be correctly sized in Monte Carlo simulations.

These inference techniques are applied to a structural model of daily equity returns. Because the rich dynamics of daily returns requires a large state space, we employ the multifrequency learning economy of Calvet and Fisher ("CF" 2007). The underlying state of fundamentals is driven by an arbitrary number of components with heterogeneous degrees of persistence (CF 2001, 2004). Investor learning about volatility is an asymmetric process, which generates substantial negative skewness in equilibrium returns. We accordingly develop an II estimator based on the auxiliary full-information MLE expanded with a robust measure of return skewness. We verify by Monte Carlo simulation the accuracy of our II estimator, particle filter and dynamic test under the multifrequency model. We also provide evidence suggesting that it would be challenging to estimate volatility frequencies by moment-based methods.

We estimate the structural model on the daily excess returns of the CRSP U.S. value-weighted index between 1926 and 1999. The simulated likelihood increases rapidly with the number of components and significantly dominates the likelihood of full-information specifications. For the out-of-sample period (2000-2009), the incomplete-information model provides accurate value-at-risk forecasts, which significantly outperform the predictions obtained from historical simulations, GARCH(1,1), and the FI model.

The paper is organized as follows. Section 2 defines a class of recursive learning economies. In section 3, we introduce the indirect inference estimator, develop filtering methods, and construct a dynamic test. Section 4 applies these methods to the multifrequency investor learning model of CF (2007), and verifies their accuracy by Monte Carlo simulations. In section 5, we conduct inference on the daily returns of a U.S. aggregate equity index between 1926 and 2009. Section 6 concludes.

2 Recursive Learning Economies

We consider a class of dynamic stochastic economies parameterized by $\theta \in \Theta \subseteq \mathbb{R}^p$, $p \geq 1$, which are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Time is discrete and indexed by $t = 0, 1, ..., \infty$. In every period t, we define three levels of information, which respectively correspond to nature, a Bayesian agent, and the econometrician.

2.1 Nature

We assume that a *state of nature* M_t drives the economy. In applications, we will also refer to M_t as the state of fundamentals. The state of nature follows a first-order Markov chain on the set $\{m^1(\theta), ..., m^d(\theta)\}$, where $m^i(\theta) \neq m^j(\theta)$ for all $i \neq j$. For every $i, j \in \{1, .., d\}$, we denote by $a_{i,j}(\theta) = \mathbb{P}[M_t = m^j(\theta)|M_{t-1} = m^i(\theta); \theta]$ the transition probability from state i to state j. We assume that the Markov chain M_t is irreducible, aperiodic, positive recurrent, and therefore ergodic. For notational simplicity, we henceforth drop the argument θ from the states m^j and transition probabilities $a_{i,j}$.

2.2 Agent

At the beginning of every period t, the agent observes a signal vector $x_t \in \mathbb{R}^{n_X}$, which is partially revealing on the state of nature M_t. The probability density function of the signal conditional on the state of nature, $f_X(x_t|M_t; \theta)$, is known to the agent. Let $X_t = (x_1, ..., x_t)$ denote the information available to her at date t. For tractability reasons, we make the following hypotheses.

Assumption 1 (Signal). *The signal x_t may be informative about the current state of nature M_t, but contains no additional information on past or future states:*

(A1a) $\mathbb{P}(M_t = m^j|M_{t-1} = m^i, X_{t-1}; \theta) = a_{i,j}$ *for all* i, j ;

(A1b) $f_X(x_t|M_t, M_{t-1}, ..., M_0, X_{t-1}; \theta) = f_X(x_t|M_t; \theta)$.

The agent is Bayesian and uses X_t to compute the conditional probability of the states of nature.

Proposition 1 (Agent Belief). *The conditional probabilities* $\Pi_t^j = \mathbb{P}(M_t = m^j|X_t; \theta)$

satisfy the recursion:

$$\Pi_t^j = \frac{\omega^j(\Pi_{t-1}, x_t)}{\sum_{i=1}^d \omega^i(\Pi_{t-1}, x_t)} \text{ for all } j \in \{1, ..., d\} \text{ and } t \geq 1, \tag{1}$$

where $\Pi_{t-1} = (\Pi_{t-1}^1, ..., \Pi_{t-1}^d)$ *and* $\omega^j(\Pi_{t-1}, x_t) = f_X(x_t|M_t = m^j; \theta) \sum_{i=1}^d a_{i,j} \Pi_{t-1}^i$.

Proof. We infer from Bayes' rule that

$$\Pi_t^j \propto \underbrace{f_X(x_t|M_t = m^j, X_{t-1}; \theta)}_{=f_X(x_t|M_t=m^j;\theta) \text{ by (A1b)}} \mathbb{P}(M_t = m^j|X_{t-1}; \theta),$$

where

$$\mathbb{P}(M_t = m^j|X_{t-1}; \theta) = \sum_{i=1}^d \underbrace{\mathbb{P}(M_t = m^j|M_{t-1} = m^i, X_{t-1}; \theta)}_{=a_{ij} \text{ by (A1a)}} \mathbb{P}(M_{t-1} = m^i|X_{t-1}; \theta),$$

and Proposition 1 holds. $\diamond$

In applications, the agent's belief vector $\Pi_t = (\Pi_t^1, ..., \Pi_t^d)$ is a key driver of the economy. For instance, the agent may set asset prices or select production plans as a function of Π_t.

The *state of the learning economy* at a given date t consists of the state of nature and the agent's conditional probability distribution:

$$s_t = (M_t, \Pi_t) \in \{m^1, \cdots, m^d\} \times \Delta_+^{d-1},$$

where $\Delta_+^{d-1} = \{\Pi \in \mathbb{R}_+^d | \sum_{i=1}^d \Pi_i = 1\}$ denotes the $(d-1)$–dimensional unit simplex. The state space of the learning economy, $\{m^1, \cdots, m^d\} \times \Delta_+^{d-1}$, is considerably larger than the state space $\{m^1, \cdots, m^d\}$ of the full-information economy, which has profound implications for inference.

Proposition 2 (State of the Learning Economy). *The state* $s_t = (M_t, \Pi_t)$ *is first-order Markov.*

Proof. The joint distribution of $M_t = m^i$ and Π_t conditional on past states $s_1, \ldots, s_{t-1}$ is

$$\underbrace{f(\Pi_t|M_t, s_{t-1}, \ldots, s_1; \theta)}_{=f(\Pi_t|M_t, \Pi_{t-1};\theta) \text{ by (1)}} \cdot \underbrace{\mathbb{P}(M_t = m^i|s_{t-1}, \ldots, s_1; \theta)}_{=\mathbb{P}(M_t=m^i|M_{t-1};\theta) \text{ by (A1a)}} = f(\Pi_t|M_t, s_{t-1}; \theta) \cdot \mathbb{P}(M_t = m^i|s_{t-1}; \theta)$$

and we conclude that s_t is first-order Markov. $\diamond$

The state of the economy s_t thus preserves the first-order Markov structure of the state of nature M_t.

By assumption, the state of nature M_t is ergodic. In order to guarantee the ergodicity of the state of the learning economy $s_t = (M_t, \Pi_t)$, we impose:

Assumption 3 (Sufficient Condition for the Ergodicity of the Learning Economy). *The transition probabilities between states of nature are strictly positive:* $a_{i,j} > 0$ *for all* i, j*, and the signal's conditional probability density functions* $f_X(x|M_t = m^j; \theta)$ *are strictly positive for all* $x \in \mathbb{R}^{n_X}$ *and* $j \in \{1, ..., d\}$.

We can then show:

Proposition 3 (Ergodicity of the Learning Economy). *Under Assumption 3, the state of the learning economy,* $s_t = (M_t, \Pi_t)$*, is ergodic.*

Proof. We know from Kaijser (1975) that the belief process Π_t has a unique invariant distribution. Proposition 2.1 in van Handel (2009), also implies that (M_t, Π_t) has a unique invariant measure Λ. Chigansky (2006) derives a similar result in continuous time, and also shows that the Law of Large Numbers holds. A minor adaptation of his proof implies that for any function Φ, the sample average $T^{-1} \sum_{t=1}^{T} \Phi(M_t, \Pi_t)$ converges almost surely to the expectation of Φ under the invariant measure Λ. $\diamond$

The learning economy is therefore asymptotically independent of the initial state $s_0 = (M_0, \Pi_0)$.

2.3 Econometrician

Each period, the econometrician observes a data point $r_t \in \mathbb{R}^{n_R}$, which is assumed to be a deterministic function of the agent's signal and conditional probabilities over states of nature:

$$r_t = \mathcal{R}(x_t, \Pi_t, \Pi_{t-1}; \theta). \tag{2}$$

We include Π_{t-1} in this definition to accommodate the possibility that r_t is a growth rate or return. The econometrician's information vector $R_t = (r_1, ..., r_t)$ is contained in the agent's information set, which is a common assumption in financial economics. The parameter vector $\theta \in \mathbb{R}^p$ specifies the states of nature $m^1, ..., m^d$, their transition probabilities $(a_{i,j})_{1 \le i,j \le d}$, the signal's conditional density $f_X(\cdot|M_t, \theta)$, and the data function $\mathcal{R}(x_t, \Pi_t, \Pi_{t-1}; \theta)$.

The econometrician's inference problem consists of estimating the structural parameter θ and the distribution of $s_t = (M_t, \Pi_t)$ conditional on R_t at every date t. Since the econometrician has no informational advantage over the investor, the Law of Iterated Expectations implies that $\mathbb{E}\left[\mathbb{P}(M_t = m^j|X_t)|R_t\right] = \mathbb{P}(M_t = m^j|R_t)$ for all j,[3] or equivalently:

$$\mathbb{E}(\Pi_t^j|R_t) = \mathbb{P}(M_t = m^j|R_t). \tag{3}$$

This result is a consistency condition between the beliefs of the agent and the beliefs of the econometrician implied by the structural model.

For each learning model, we can define an auxiliary full information (FI) model in which the agent observes the state of nature M_t as well as the signal x_t. Her conditional probabilities are then

$$\Pi_t^j = \mathbb{P}(M_t = m^j|X_t, M_t; \theta),$$

and the belief vector Π_t is the vertex of the simplex corresponding to state M_t. The data point r_t is defined by the maintained equation (2). In many cases, the FI model has less parameters than the II model because of the simplification in the definition of Π_t. For now, we assume that the auxiliary FI models are parameterized by $\phi \in \mathbb{R}^q$, where $1 \le q \le p$. The probability density function of the data point r_t given $M_t = m^j$ and $M_{t-1} = m^i$ can be computed from the data point specification (2) and the conditional p.d.f. of the signal, $f_X(x_t|M_t = m^j, \theta)$. We make the following simplifying assumption.

Assumption 4 (Auxiliary FI Economies). *The probability density functions $f_{i,j}(r_t; \phi)$ are available analytically for all $i, j \in \{1, ..., d\}$.*

[3]Let $1_{\{M_t = m^j\}}$ denote the indicator function equal to unity if $M_t = m^j$ and 0 otherwise. The Law of Iterated Expectations implies that $\mathbb{E}\left[\mathbb{E}(1_{\{M_t=m^j\}}|X_t)|R_t\right] = \mathbb{E}(1_{\{M_t=m^j\}}|R_t)$.

The econometrician's filtering problem reduces to estimating the distribution of the state of nature M_t conditional on R_t, and the likelihood function $\mathcal{L}(\phi|R_T)$ is available analytically, as is shown in the appendix.

When the agent has incomplete information, the likelihood is generally unavailable in closed form because of the larger state space. The learning model can, however, be conveniently simulated. Given a state $s_{t-1} = (M_{t-1}, \Pi_{t-1})$, we can: (i) sample M_t from M_{t-1} using the transition probabilities $a_{i,j}$; (ii) sample the signal x_t from $f_X(\cdot|M_t;\theta)$; (iii) apply Bayes'rule (1) to impute the agent's belief Π_t; and (iv) compute $r_t = \mathcal{R}(x_t, \Pi_t, \Pi_{t-1};\theta)$. Estimation can therefore proceed by simulation-based methods. The simulated method of moments (SMM) is a possible estimation strategy, but moment selection can be challenging and efficiency quite poor, as in the multifrequency volatility economy considered in section 4. We now propose a related approach that builds more closely on the structure of incomplete-information models.

3 Indirect Inference Estimation and Filtering for Learning Models

3.1 Estimation

We assume that the data $R_T = (r_1, ..., r_T)$ is generated by the learning model with parameter θ^*. The II estimation of θ^* proceeds in two steps.

First, we define an empirical auxiliary estimator by expanding the full-information MLE with a set of statistics. Let

$$\hat{\phi}_T = \arg\max_{\phi} \mathcal{L}(\phi|R_T) \in \mathbb{R}^q$$

denote the empirical MLE of the full-information model. If $q < p$, we also consider a set of $p - q$ statistics $\hat{\eta}_T$ that quantify features of the dataset R_T that the learning model is designed to capture. The *empirical auxiliary estimator* is defined by

$$\hat{\mu}_T = \begin{bmatrix} \hat{\phi}_T \\ \hat{\eta}_T \end{bmatrix} \in \mathbb{R}^p. \tag{4}$$

By construction, $\hat{\mu}_T$ contains as many parameters as the structural parameter θ^*.

Second, we select a structural parameter estimate $\hat{\theta}_T$ that matches as closely as possible the empirical auxiliary estimate $\hat{\mu}_T$. More specifically, for any admissible parameter θ, we can simulate a sample path $R_{ST}(\theta)$ of length ST, $S \geq 1$, and compute the corresponding auxiliary estimator:

$$\hat{\mu}_{ST}(\theta) = \begin{bmatrix} \hat{\phi}_{ST}(\theta) \\ \hat{\eta}_{ST}(\theta) \end{bmatrix}, \tag{5}$$

where $\hat{\phi}_{ST}(\theta) = \arg\max_{\phi} \mathcal{L}[\phi | R_{ST}(\theta)]$. The indirect inference (II) matches as closely as possible the empirical auxiliary estimator $\hat{\mu}_T$ computed from the data:

$$\hat{\theta}_T = \arg\min_{\theta} \left[\hat{\mu}_{ST}(\theta) - \hat{\mu}_T\right]' \Omega \left[\hat{\mu}_{ST}(\theta) - \hat{\mu}_T\right], \tag{6}$$

where Ω is a positive definite weighting matrix. Since the auxiliary estimator $\hat{\mu}_T$ contains as many parameters as the structural parameter θ^*, the II estimator is exactly identified. In applications, we will therefore frequently obtain that $\hat{\mu}_{ST}(\hat{\theta}_T) = \hat{\mu}_T$.

Our methodology builds on the fact that the full-information economy can be efficiently estimated by ML and is therefore a natural candidate auxiliary model. Moreover, the theoretical investigation of a learning model often begins with the characterization of the FI case, so the estimation method we are proposing follows the natural progression commonly used in the literature.

We assume that the regularity conditions in Gouriéroux et al. (1993) and Gouriéroux and Monfort (1996) hold. Under the structural model θ^*, the auxiliary estimator $\hat{\mu}_T$ converges in probability to a deterministic function $\mu(\theta^*)$, called the *binding function*, and $\sqrt{T}\,[\hat{\mu}_T - \mu(\theta^*)] \xrightarrow{d} \mathcal{N}(0, W^*)$. Furthermore, when S is fixed and T goes to infinity, the II estimator is consistent and asymptotically normal:

$$\sqrt{T}(\hat{\theta}_T - \theta^*) \xrightarrow{d} \mathcal{N}(0, V),$$

where

$$V = \left(1 + \frac{1}{S}\right) \left[\frac{\partial \mu(\theta^*)}{\partial \theta'}\right]^{-1} W^* \left[\frac{\partial \mu(\theta^*)'}{\partial \theta}\right]^{-1}. \tag{7}$$

The II estimator has good asymptotic properties if: (a) the auxiliary estimator has a low variance-covariance matrix W^*, and (b) the Jacobian matrix of the binding function has a small inverse. Condition (a) motivates the choice of the full-information MLE and of auxiliary statistics with relatively low variances. Condition (b) suggests that the auxiliary statistics should also be sensitive to the parameters that are specific to the learning model.

The numerical implementation can be accelerated by the efficient method of moments when the auxiliary statistic $\hat{\eta}_T$ maximizes a criterion function of the form $\mathcal{H}(\eta, R_T)$. For instance, $\hat{\eta}_T$ may be a vector of sample moments or sample quantiles of the dataset R_T. The empirical auxiliary estimator $\hat{\mu}_T = (\hat{\phi}'_T, \hat{\eta}'_T)'$ then maximizes the composite criterion:

$$\mathcal{Q}_T(\mu, R_T) \equiv \frac{1}{T}\mathcal{L}(\phi, R_T) + \mathcal{H}(\eta, R_T) \quad \text{where } \mu = (\phi', \eta')'.$$

Similarly, the simulated auxiliary estimator $\hat{\mu}_{ST}(\theta)$ maximizes $\mathcal{Q}_{ST}[\mu, R_{ST}(\theta)]$ for every θ. When the structural model matches exactly the empirical auxiliary estimator: $\hat{\mu}_{ST}(\hat{\theta}_T) = \hat{\mu}_T$, the II estimator $\hat{\theta}_T$ minimizes the EMM-type objective function:

$$\left\{\frac{\partial \mathcal{Q}_{ST}}{\partial \mu}[\hat{\mu}_T, R_{ST}(\theta)]\right\}' W_T \left\{\frac{\partial \mathcal{Q}_{ST}}{\partial \mu}[\hat{\mu}_T, R_{ST}(\theta)]\right\}, \tag{8}$$

where W_T is a positive-definite weighting matrix. This property can be used to compute $\hat{\theta}_T$. For each iteration of θ, the evaluation of the EMM objective function (8) requires only the evaluation of the score. By contrast, the evaluation of the II objective function (6) requires the optimization of the FI likelihood in order to obtain $\hat{\mu}_{ST}(\theta)$. The computational advantage of EMM is substantial in applications where the calculation of the full-information MLE is expensive. We refer the reader to the appendix for a more detailed description of the numerical estimation.

3.2 Filtering

Once the structural parameter of the learning model has been estimated, the econometrician can impute the conditional distribution of the state of the economy $s_t = (M_t, \Pi_t)$ given R_t. Since s_t is first-order Markov, we approximate its conditional distribution by a particle filter based on the following analysis. The belief vector Π_t is by Bayes'rule (1) a deterministic function of the signal x_t and the lagged belief vector Π_{t-1}, and the data point r_t is a deterministic function of the agent's signal x_t and beliefs Π_t and Π_{t-1}. We infer that:

$$f(s_t, x_t, \Pi_{t-1}|R_t) = \frac{\delta[r_t - \mathcal{R}(x_t, \Pi_t, \Pi_{t-1})]}{f(r_t|R_{t-1})} f(s_t, x_t, \Pi_{t-1}|R_{t-1}), \tag{9}$$

where δ denotes the Dirac distribution. Let k denote a positive kernel defined on the real line, and for any $r \in \mathbb{R}^{n_R}$, let $K_h(r; V) = h^{-n_R}(\det V)^{-1/2} k\left(h^{-2} r' V^{-1} r\right)$ denote the corresponding probabiliy density kernel with bandwidth h. The kernel K_h converges to the Dirac distribution as h goes to zero, which we use to approximate (9).

We define recursively a particle filter based on sampling and importance resampling (SIR). Assume that $\{s_{t-1}^{(n)}\}_{1\leq n\leq N} = \{(M_{t-1}^{(n)}, \Pi_{t-1}^{(n)})\}_{1\leq n\leq N}$ is a filter that targets $f(s_{t-1}|R_{t-1})$. The computation of $(s_t^{(1)}, ..., s_t^{(N)})$ proceeds as follows.

Step 1 (Sampling): We simulate the economy one step forward. That is, for every $n \in \{1, ..., N\}$, we:

- draw a new state of nature $\tilde{M}_t^{(n)}$ from $M_{t-1}^{(n)}$ using the transition probabilities of the Markov chain;
- sample the signal observed by the agent $\tilde{x}_t^{(n)}$ from $f(x|\tilde{M}_t^{(n)})$;
- apply Bayes' rule (1) to infer her conditional distribution $\tilde{\Pi}_t^{(n)}$;
- compute the pseudo data point $\tilde{r}_t^{(n)} = \mathcal{R}(\tilde{x}_t^{(n)}, \tilde{\Pi}_t^{(n)}, \tilde{\Pi}_{t-1}^{(n)})$;
- let $\tilde{s}_t^{(n)} = (\tilde{M}_t^{(n)}, \tilde{\Pi}_t^{(n)})$.

Step 2 (Importance weights): We observe the new data point r_t and compute the probabilities

$$p_t^{(n)} = \frac{K_h\left(r_t - \tilde{r}_t^{(n)}; V_t\right)}{\sum_{n'=1}^{N} K_h\left(r_t - \tilde{r}_t^{(n')}; V_t\right)},\ n = 1, ..., N,$$

where V_t denotes the sample covariance matrix of $\tilde{r}_t^{(n)}$.

Step 3 (Multinomial Resampling): For every $n \in \{1, ..., N\}$, we draw $s_t^{(n)}$ from $\tilde{s}_t^{(1)}, ..., \tilde{s}_t^{(N)}$ with probabilities $p_t^{(1)}, ..., p_t^{(N)}$.

The set of particles $\{(s_{t-1}^{(n)}, \tilde{x}_t^{(n)}, \tilde{s}_t^{(n)})\}_{n=1,...,N}$ constructed in step 1 provide a discrete approximation to the conditional distribution of (s_{t-1}, x_t, s_t) given the data R_{t-1}. In step 2, we construct a measure of the distance between the pseudo and the actual data points, and in Step 3 we select particles for which this distance is small.

An alternative to Step 3 is:

Step 3′ (Residual Resampling): We select $\sum_{n=1}^{N} \lfloor N p_t^{(n)} \rfloor$ particles deterministically by setting $\lfloor N p_t^{(1)} \rfloor$ particles equal to $\tilde{s}_t^{(1)}$, ...,$\lfloor N p_t^{(N)} \rfloor$ particles equal to $\tilde{s}_t^{(N)}$, where $\lfloor \cdot \rfloor$ denotes the floor of a real number. The remaining $N_{r,t} = N - \sum_{n=1}^{N} \lfloor N p_t^{(n)} \rfloor$ particles are selected by drawing independent draws from the multinomial distribution that produces $\tilde{s}_t^{(n)}$ with probability $(N p_t^{(n)} - \lfloor N p_t^{(n)} \rfloor)/N_{r,t}$.

Earlier research (e.g. Chopin, 2004) shows that the residual resampling method has a smaller asymptotic variance and is numerically more efficient than multinomial resampling, so we use Step 3′ in applications.

The particle filter incorporates the joint dynamics of the state of nature and the signal, as well as Bayes' rule. We show in the appendix that for any bounded continuous function Φ, the expectation of $\Phi(s_t)$ conditional on the data R_t satisfies:

$$\lim_{h \to 0} \lim_{N \to \infty} \sum_{n=1}^{N} p_t^{(n)} \Phi(\tilde{s}_t^{(n)}) = \mathbb{E}\left[\Phi(s_t) | R_t\right], \tag{10}$$

where the limit with respect to N refers to almost sure convergence. The filter is therefore consistent.[4] We will verify in section 4 that it performs well in finite samples.

Simulated likelihood and forecasts. We estimate the conditional density of r_t given past observations by

$$\hat{f}(r_t | R_{t-1}) = \frac{1}{N} \sum_{n=1}^{N} K_h(r_t - \tilde{r}_t^{(n)}; V_t).$$

The Epanechnikov kernel $k(s) = (3/4) \max(1 - s; 0)$ is known to produce an asymptotically efficient density estimator and to also have excellent computational properties (e.g., Silverman 1986). In practice, we use it with the fixed optimal bandwidth $h = \left[8 c_{n_R}^{-1} (n_R + 4)(2\sqrt{\pi})^{n_R} / N\right]^{1/(n_R+4)}$, where c_{n_R} denotes the volume the unit sphere in $\mathbb{R}^{n_R}$. In order to avoid duplicating calculations, we employ the same bandwidth and kernel for the estimation of the importance probabilities $p_t^{(n)}$.

We estimate the log-likelihood function by $\sum_{t=1}^{T} \ln \hat{f}(r_t | R_{t-1})$. The calculation of the simulated likelihood is often too intensive to conduct simulated ML estimation, and we will rely for this reason on the indirect inference method described in section 3.1.[5] The

[4] See Crisan and Doucet (2002) for an excellent survey on the convergence of particle filters.

[5] For instance in the example considered in sections 4 and 5, we use a particle filter of size $N = 20$ million

computation of the likelihood of the learning model is very useful, however, for model selection; standard methods such as the likelihood ratio tests for nested models or the Vuong tests for nonnested models can be applied. The particle filter can also produce forecasts of the conditional density of returns and value at risk.

3.3 Dynamic Testing with a Particle Filter

In economics and finance, we are often interested in testing moment conditions of the form:

$$\mathbf{H_0} : \mathbb{E}[\Phi(s_t)|R_t] = 0,$$

at given period t, where Φ is a continuous function of the latent state taking values in a Euclidean space. Chopin (2004) shows that for a large class of filters, the weighted mean $\sum_{n=1}^{N} p_t^{(n)} \Phi(\tilde{s}_t^{(n)})$ is asymptotically normal:

$$\sqrt{N} \left\{ \sum_{n=1}^{N} p_t^{(n)} \Phi(\tilde{s}_t^{(n)}) - \mathbb{E}[\Phi(s_t)|R_t] \right\} \xrightarrow{d} N(0, V_t),$$

as $N \longrightarrow \infty$.[6] Because the particles $(\tilde{s}_t^{(1)}, ..., \tilde{s}_t^{(N)})$ all depend on past realizations of the filter, they are not independent random variables (conditional on R_t) and the sample variance of $\tilde{s}_t^{(n)}$ is an inconsistent estimator of V_t. Chopin (2004) derives a recursive expression for V_t, and explains that it is challenging to estimate it in practice. Perhaps for this reason, there has been, to the best of our knowledge, no previous attempt to develop dynamics tests based on particle filters.

One natural solution is to run A independent filters $\left\{ (\tilde{s}_{t'}^{(1,a)}, ..., \tilde{s}_{t'}^{(N,a)})_{0 \leq t' \leq t} \right\}$, $a \in \{1, ..., A\}$, starting at date 0. For each filter, we compute the sample mean:

$$\bar{\Phi}_t^{(a)} = \sum_{n=1}^{N} p_t^{(n,a)} \Phi(\tilde{s}_t^{(n,a)}).$$

and a dataset of about 20,000 observations. One evaluation of the likelihood function requires the evaluation 400 billion conditional densities. Since a typical optimization requires about 500 function evaluations, the simulated ML estimation of the II model would require the evaluation of 200 trillion conditional densities, which would be impractical.

[6] The sample sum $N^{-1} \sum_{n=1}^{N} \Phi(s_t^{(n)})$ is also asymptotically normal but has a larger variance than $\sum_{n=1}^{N} p_t^{(n)} \Phi(\tilde{s}_t^{(n)})$, so we focus our attention on $\sum_{n=1}^{N} p_t^{(n)} \Phi(\tilde{s}_t^{(n)})$

We know from Chopin (2004) that $\sqrt{N}\bar{\Phi}_t^{(a)} \xrightarrow{d} \mathcal{N}(0, V_t)$ as $N \to \infty$. Since we are working with nonlinear filters, however, $\mathbb{E}(\bar{\Phi}_t^{(a)})$ will generally differ from zero in final samples even if the null hypothesis holds. For a fixed value of N and as $A \to \infty$, the sample mean $\bar{\bar{\Phi}}_t = (\bar{\Phi}_t^{(1)} + ... + \bar{\Phi}_t^{(A)})/A$ converges to $\mathbb{E}(\bar{\Phi}_t^{(1)})$, which generally differs from zero.

The sample variance-covariance matrix of $\bar{\Phi}_t^{(1)}, ..., \bar{\Phi}_t^{(A)}$,

$$\hat{V}_t = \frac{1}{A-1} \sum_{a=1}^{A} (\bar{\Phi}_t^{(a)} - \bar{\bar{\Phi}}_t)(\bar{\Phi}_t^{(a)} - \bar{\bar{\Phi}}_t)'$$

converges to V_t as N and A get large. We infer that:

$$\sqrt{N}\hat{V}_t^{-1/2}\bar{\Phi}_t^{(1)} \xrightarrow{d} \mathcal{N}(0, I)$$

as $N, A \to \infty$. We will verify in section 4 the accuracy of the size of test in an asset pricing application.[7]

4 Inference in an Asset Pricing Model with Investor Learning

We now apply our methodology to a consumption-based asset pricing model. We adopt a Lucas tree economy with regime switching in fundamentals, which we use to specify the dynamics of daily equity returns.[8]

4.1 Dynamics of the State of Nature

The rich dynamics of daily returns requires a large state space. For this reason, we assume that the state of nature M_t follows a binomial Markov Switching Multifractal (MSM), as in CF (2001, 2008). That is, the state is a vector containing $\overline{k}$ components:

$$M_t = (M_{1,t}, \ldots, M_{\overline{k},t})' \in \mathbb{R}_+^{\overline{k}}.$$

[7]For a given filter size N, we know that in general, $\mathbb{E}(\bar{\Phi}_t^{(1)}) \neq 0$ and the distribution of $\sqrt{AN}\hat{V}_t^{-1/2}\bar{\bar{\Phi}}_t$ becomes degenerate as $A \to \infty$.

[8]Applications of this class of models under full information include Abel (1994, 1999), Bonomo and Garcia (1991, 1994, 1996), Cecchetti, Lam and Mark (1990). Investor learning is investigated within this class by Brandt, Zeng, and Zhang (2004), Calvet and Fisher (2007), David and Veronesi (2006), Lettau, Ludvigson and Wachter (2008), and Moore and Schaller (1996).

Each component is either in a high value $m_0 \in [1;2]$ or a low value $2 - m_0 \in [0;1]$. The state space therefore contains $d = 2^{\overline{k}}$ elements.

The components are mutually independent across k. Given a value $M_{k,t}$ for the k^{th} component at date t, the next-period multiplier $M_{k,t+1}$ is either:

$$\begin{cases} \text{drawn from a fixed distribution } M \text{ with probability } \gamma_k, \\ \text{equal to its current value } M_{k,t} \text{ with probability } 1-\gamma_k. \end{cases}$$

The distribution M is a Bernoulli taking the values m_0 and $2 - m_0$ with equal probability. The transition probabilities γ_k are parameterized by

$$\gamma_k = 1 - (1 - \gamma_{\overline{k}})^{b^{k-\overline{k}}}, \ k = 1, \ldots, \overline{k},$$

where $b > 1$. They are approximately geometric in k, since $\gamma_k \approx \gamma_{\bar{k}} b^{k-\overline{k}}$. Thus, $\gamma_{\overline{k}}$ controls the persistence of the highest-frequency component and b determines the spacing between frequencies.

4.2 Bayesian Agent in the Lucas Tree Economy

The agent receives an exogenous consumption stream $\{C_t\}$ and prices the stock, which is a claim on an exogenous dividend stream $\{D_t\}$. We let $c_t = \ln(C_t)$, $d_t = \ln(D_t)$, $\Delta c_t = c_t - c_{t-1}$, and $\Delta d_t = d_t - d_{t-1}$.

As in Calvet and Fisher (2007), the growth rates of consumption and dividends are specified by:

$$\begin{cases} \Delta c_t = g_c + \sigma_c \varepsilon_{c,t} \, , \\ \Delta d_t = g_d - \dfrac{\sigma_d^2(M_t)}{2} + \sigma_d(M_t)\varepsilon_{d,t} \, . \end{cases} \tag{11}$$

The innovations $\epsilon_{c,t}$ and $\epsilon_{d,t}$ are jointly normal and have zero means, unit variances, and correlation $\rho_{c,d}$. The volatility of dividend growth is specified by the product of the components of the state vector:

$$\sigma_d(M_t) = \overline{\sigma}_d \left(\prod_{k=1}^{\overline{k}} M_{k,t} \right)^{1/2},$$

where $\overline{\sigma}_d \in \mathbb{R}_+$. Multipliers deliver discrete switches, consistent with evidence of outliers and apparent nonstationarity in financial series (e.g., Schwert 1989; Pagan and Schwert 1990).

The signal x_t observed by the representative investor every period includes consumption growth Δc_t, dividend growth Δd_t, and noisy observations of the volatility components:

$$\delta_t = M_t + \sigma_\delta z_t.$$

We assume that $\sigma_\delta \in \mathbb{R}_+$ and that $z_t \in \mathbb{R}^{\bar{k}}$ is an i.i.d. vector of independent standard normals. The noise parameter σ_δ controls information quality.

Learning about the volatility state M_t is an asymmetric process. For expositional simplicity, assume that the noise parameter σ_δ is large, so that investors learn about M_t primarily through the dividend growth Δd_t. We consider two scenarios. First, when volatility switches from a low to a high state, the agent learn abruptly about the switch, because large realizations of Δd_t are implausible in a low-volatility regime. Learning about increases in volatility is therefore abrupt. Second, consider a switch from a high to a low volatility state. Because realizations of Δd_t near the mean are likely outcomes under any M_t, the agent learns only progressively that volatility has gone down. Learning about a decrease in volatility tends to be gradual. We will see shortly that this asymmetry in learning has important implications for the dynamics of stock returns.

The agent agent has isoelastic expected utility:

$$U_0 = \mathbb{E}_0 \left(\sum_{t=0}^{\infty} \delta^t \frac{C_t^{1-\alpha}}{1-\alpha} \right),$$

where δ is the discount rate and α is the coefficient of relative risk aversion. In equilibrium, the log interest rate is constant and satisfies $r_f = -\ln(\delta) + \alpha g_c - \alpha^2 \sigma_c^2/2$. The stock's price-dividend ratio, $\sum_{n=1}^{\infty} \delta^n \mathbb{E}\left[(C_{t+n}/C_t)^{-\alpha} D_{t+n}/D_t \middle| X_t \right]$, is negatively related to volatility,[9] and is linear in the belief vector:

$$Q(\Pi_t) = \sum_{j=1}^{d} Q^j \Pi_t^j.$$

[9] We infer from (11) that $\delta^n \mathbb{E}\left[(C_{t+n}/C_t)^{-\alpha} D_{t+n}/D_t \middle| X_t \right] = \mathbb{E}\left[\prod_{h=1}^{n} e^{g_d - r_f - \alpha \rho_{c,d} \sigma_c \sigma_d (M_{t+h})} | X_t \right]$. Since volatility is persistent, a high level of volatility at date t implies high forecasts of future volatility, and therefore a low period–t price-dividend ratio.

The price-dividend ratio is equal to Q^j if the agent is certain to be in state j. The linear coefficients are given by $\left(Q^1, \ldots, Q^d\right)' = (I-B)^{-1}\iota - \iota$, where $\iota = (1, ..., 1)'$, and $B = (b_{ij})_{1 \leq i,j \leq d}$ is the matrix with components $b_{ij} = a_{i,j} \exp\left[g_d - r_f - \alpha\, \rho_{c,d}\, \sigma_c\, \sigma_d(m^j)\right]$.

4.3 Econometric Specification of Stock Returns

The log excess return process is given by:

$$r_t = \ln\left[\frac{1+Q(\Pi_t)}{Q(\Pi_{t-1})}\right] + g_d - r_f - \frac{\sigma_d^2(M_t)}{2} + \sigma_d(M_t)\varepsilon_{d,t}\,. \tag{12}$$

Since learning about volatility is asymmetric, the stock price falls abruptly following bad news about volatility, but will increase only gradually following good news. The noise parameter σ_δ therefore controls the skewness of stock returns.

As is traditional in the asset pricing literature, we calibrate some of the parameters on aggregate data. Specifically, we set the consumption drift to $g_c = 0.75$ basis point (bp) (or 1.18% per year), excess dividend growth equal to $g_d - r_f = 0.5$ bp per day (about 1.2% per year), consumption volatility to $\sigma_c = 0.189\%$ (or 2.93% per year), and dividend volatility $\overline{\sigma}_d = 0.70\%$ per day (about 11% per year). The correlation coefficient is assumed to be $\rho_{c,d} = 0.6$. We calibrate the mean price-dividend ratio to a plausible long-run value

$$\mathbb{E}[Q(\Pi_t)] = \overline{Q}, \tag{13}$$

where $\overline{Q}$ is set equal to 25 in yearly units, or 6,000 in daily units.

We estimate the parameters:

$$\theta = (m_0, \gamma_{\overline{k}}, b, \sigma_\delta)' \in [1,2] \times (0,1] \times [1,\infty) \times \mathbb{R}_+,$$

where m_0 controls the variability of dividend volatility, $\gamma_{\overline{k}}$ is the transition probability of the most transitory volatility component, b controls the spacing of the transition probabilities, and σ_δ controls the precision of the signal received by the representative agent. Under full information, the vector of parameters reduces to $\phi = (m_0, \gamma_{\overline{k}}, b)'$.

Remark 1: An alternative approach would be to estimate all the parameters of the learning economy on aggregate excess return data, instead of calibrating some of the parameters on consumption data. In the 2005 NBER version of their paper, CF applied this method to the FI model and obtained broadly similar results to the ones reported in the published version. This alternative approach has the disadvantage of not taking into account the economic constraints imposed by the model, and we do not pursue it here.

Remark 2: The methodology highlighted above can accommodate a wide range of variants and extensions. For instance, it is straightforward to incorporate switches in the drift of dividend growth and in the drift and volatility of consumption growth and (CF 2007). Investors may exhibit Epstein and Zin preferences (e.g., CF 2007; Lettau, Ludvigson and Wachter 2008) or disappointment aversion (e.g., Bonomo and Garcia 1993), implement non-Bayesian learning rules (e.g., Brandt, Zeng and Zhang 2002; Cecchetti, Lam and Mark 2000), or have access to a production technology (e.g., Van Nieuwerburgh and Veldkamp 2006). Our inference methodology directly applies to these extensions.

4.4 Indirect Inference Estimator

The learning economy is specified by $p = 4$ parameters, $\theta = (m_0, \gamma_{\bar{k}}, b, \sigma_\delta)'$, while the FI economy is specified by $q = 3$ parameters, $\phi = (m_0, \gamma_{\bar{k}}, b)'$. For this reason, the definition of the auxiliary estimator requires an additional statistic $\hat{\eta}_T \in \mathbb{R}$. Since the noise parameter σ_δ controls the skewness of excess returns, the third moment seems like a natural choice. We are concerned however, that third moment may be too noisy to produce an efficient estimator of θ^*. For this reason, we consider a robust alternative based on the observation that by restriction (13), the mean return is nearly independent of the structural parameter:

$$\mathbb{E}(r_t) \approx \ln(1 + 1/\overline{Q}) + g_d - r_f - \overline{\sigma}_d^2/2. \tag{14}$$

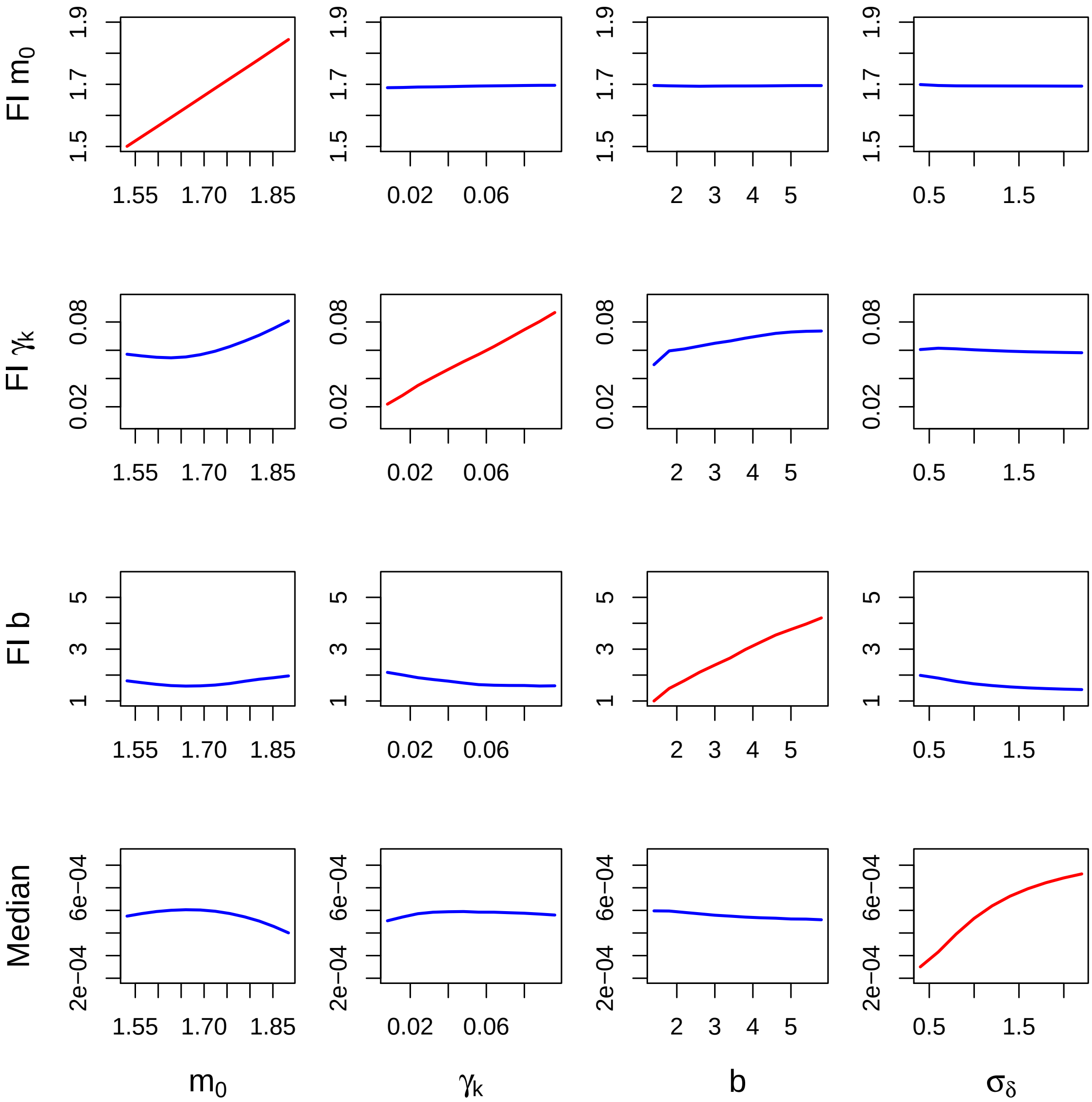
Since the mean is fixed, the median can be used as a robust measure of skewness.

The auxiliary estimator $\hat{\mu}_T = (\hat{\phi}_T, \hat{\eta}_T)'$ is defined by expanding the ML estimator of the full-information economy, $\hat{\phi}_T$, with either the third moment (Estimator 1), or the return median $\hat{\eta}_T = \text{median}(\{r_t\})$ (Estimator 2). The EMM methodology highlighted in section 3.1 applies to both II estimators.[10]

In Figure 2, we illustrate the relation between the auxiliary estimator $\hat{\mu}_T = (\hat{\phi}_T', \hat{\eta}_T)'$ and the structural parameter θ on a long simulated sample. The graphs correspond to the second II estimator and can be viewed as cuts of the binding function $\mu(\theta)$ discussed in section 3.1. The full-information MLE $\hat{\phi}_T$ provides almost unbiased estimators of the learning model parameters m_0 and $\gamma_{\bar{k}}$, and a negatively biased estimator of the spacing pa-

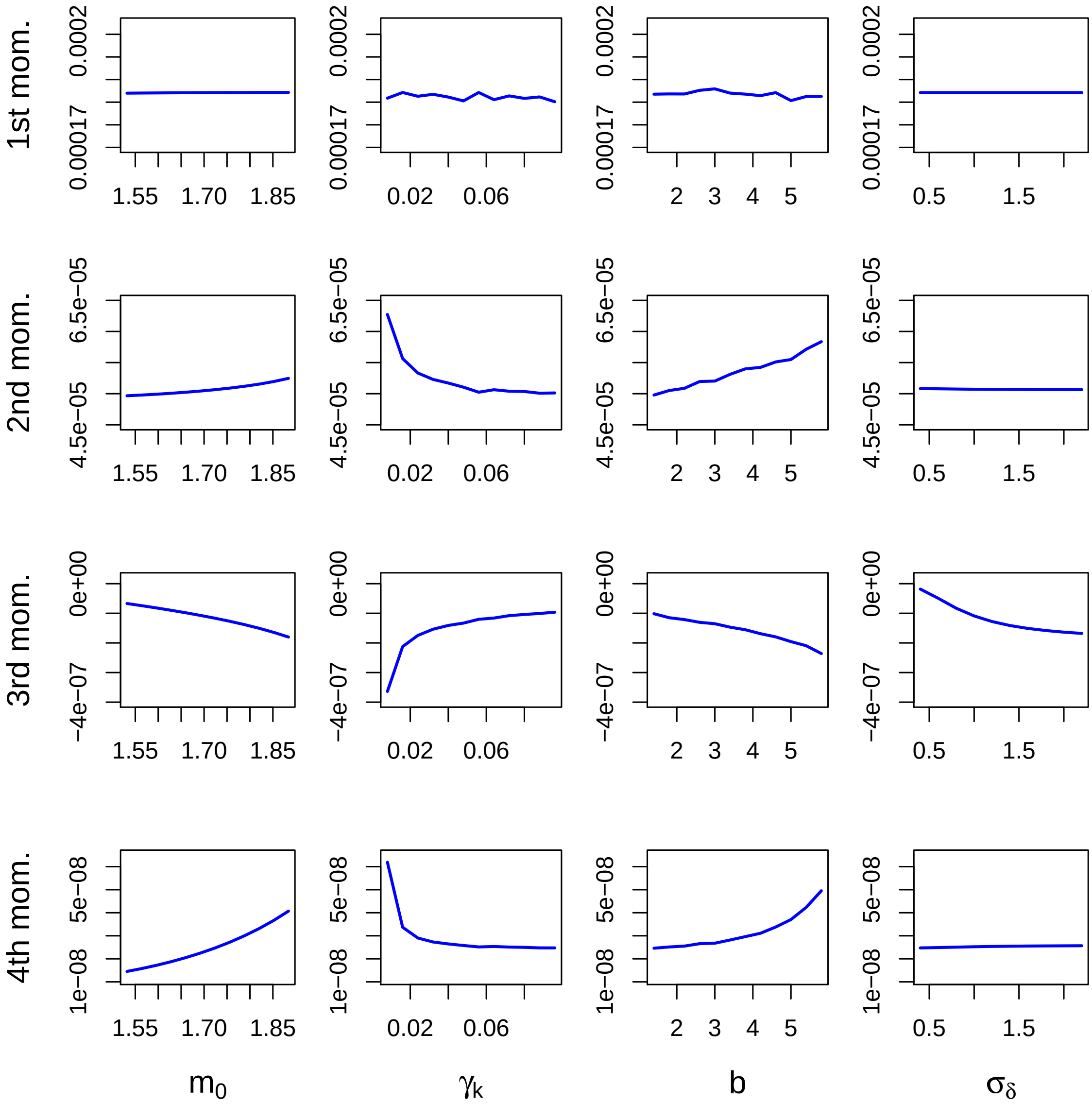
[10] The median maximizes the criterion $-\sum_{t=1}^{T} |r_t - \eta| / T$.

Figure 2: BINDING FUNCTION



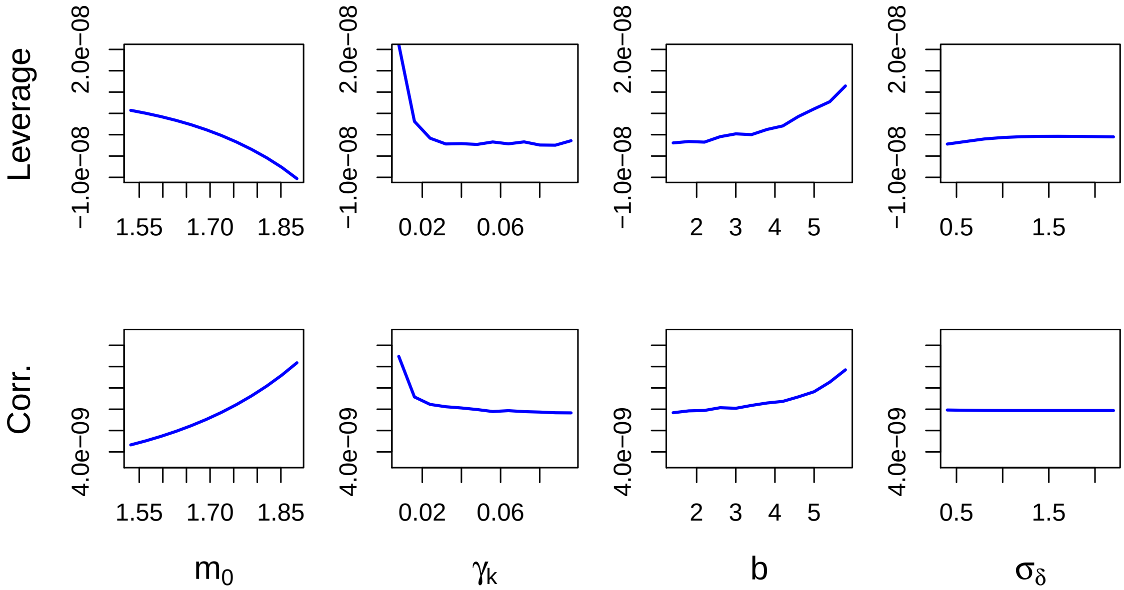
Notes. We report ML estimates of m_0, $\gamma_{\underline{k}}$, b and the sample median for a simulated sample of size 10^7 generated under the learning model with $\underline{k} = 3$. The reference parameter values are $m_0 = 1.7$, $\gamma_{\underline{k}} = 0.06$, $b = 2$ and $\sigma_\delta = 1$. In each column, one parameter is allowed to vary while the other three parameters are set to their reference values.

Figure 3: ALTERNATIVE AUXILIARY STATISTICS



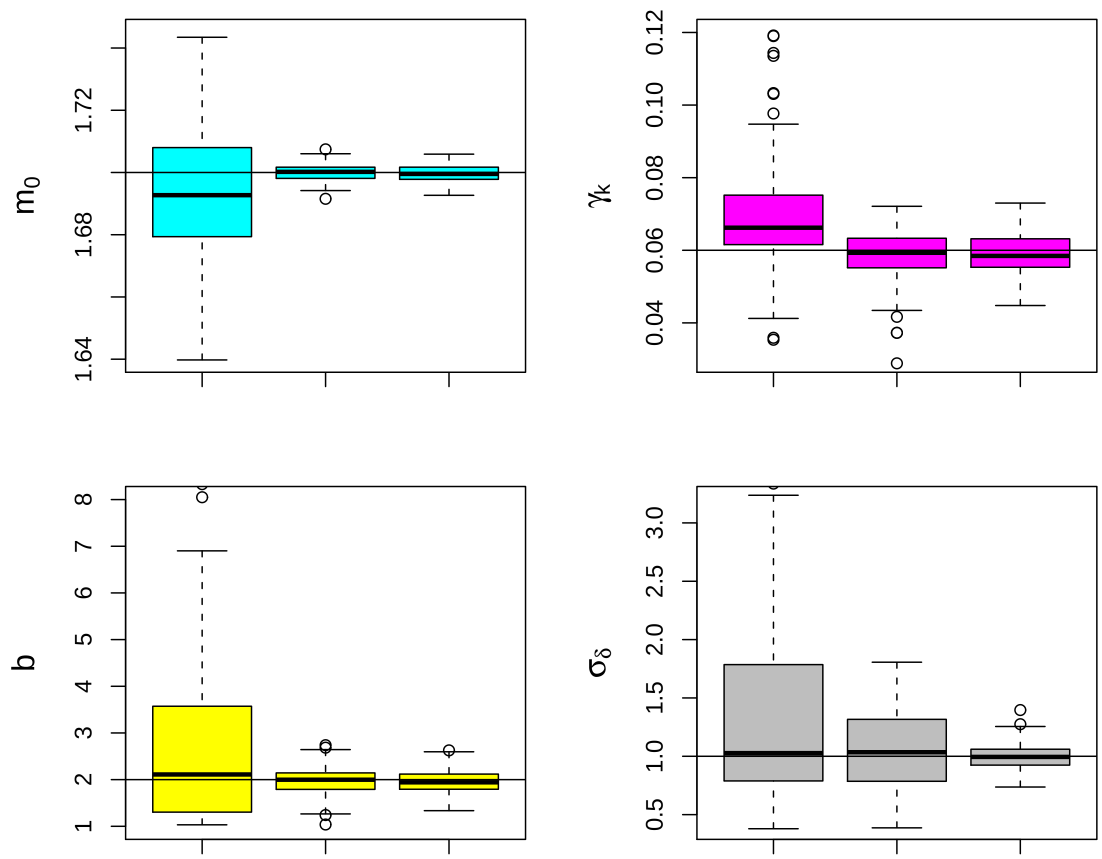
Notes. We report sample moments for data of size 10^7 simulated under the learning model with $\overline{k} = 3$ and parameter values $m_0 = 1.7$, $\gamma_{\overline{k}} = 0.06$, $b = 2$ and $\sigma_\delta = 1$. In a given sensitivity plot, one parameter varies and the others are fixed at their true parameter values.

Figure 4: ALTERNATIVE AUXILIARY STATISTICS



Notes. We report the leverage coefficient $\sum_{t=2}^{ST} r_{t-1} r_t^2/(ST)$, and the measure of volatility autocorrelation $\sum_{t=2}^{ST} r_{t-1}^2 r_t^2/(ST)$ with $ST = 10^7$ simulated under the learning model with $\overline{k} = 3$ and parameter values $m_0 = 1.7$, $\gamma_{\overline{k}} = 0.06$, $b = 2$ and $\sigma_\delta = 1$. In a given sensitivity plot, one parameter varies and the others are fixed at their true parameter values.

Figure 5: MONTE CARLO SIMULATIONS OF THE II ESTIMATOR



Notes. We report boxplots of the structural parameter estimates under three methods: a version of SMM (left boxplot of each panel), a version of II based on the third moment (middle boxplot), and a version of II based on the median (right boxplot). SMM employs the second, third and fourth moments of log excess returns, $\mathbb{E}(r_t^n)$, $n \in \{2,3,4\}$, and the leverage measure $\mathbb{E}(r_{t-1}r_t^2)$ as auxiliary statistics. For the two II methods, the auxiliary statistics are the MLE of the full-information model extended by the third moment (middle boxplots) or the median (right boxplots) of log excess returns. For each technique, the boxplots are based on estimates obtained from 100 simulated datasets of size 20,000 sampled from the learning model with $\overline{k} = 3$ components and parameters $m_0 = 1.7$, $\gamma_{\overline{k}} = 0.06$, $b = 2$ and $\sigma_\delta = 1$.

rameter b. As the noise parameter σ_δ increases, the median return increases monotonically, consistent with the fact that returns become more negatively skewed.

In Figures 3 and 4, we illustrate the impact of the structural parameters $m_0, \gamma_{\bar{k}}, b$ and σ_δ on the sample averages of r_t^n, $n \in \{1, \ldots, 4\}$, the leverage coefficient $r_{t-1}r_t^2$, and the volatility autocorrelation measure $r_{t-1}^2 r_t^2$. The noise parameter σ_δ has an impact on the third moment, as one expects, but not on the other moments. We conclude from Figures 2 and 4 that the median and the third moment are the only reasonable candidates for the auxiliary statistic $\hat{\eta}_T$.

Consistent with (14), the first moment of returns does not vary with the structural parameters. All reported moments tend to be insensitive to b and nonsmooth with respect to $\gamma_{\bar{k}}$ and b due to Markov-switching. Figures 3 and 4 thus illustrate the difficulty of using a moment-based approach to estimate the frequency parameters of the model.[11]

By (7), an II estimator is asymptotically efficient if: (a) its auxiliary estimator has a low variance, and (b): the Jacobian matrix of the binding function has a small inverse. Condition (b) seems to be satisfied by both II estimators. Since the median is less noisy than the third moment, condition (a) suggests that the median-based II estimator may be a more efficient choice. We now confirm this intuition by Monte Carlo simulations.

4.5 Monte Carlo Simulations

In Figure 5, we report boxplots of the SMM and II estimates of the structural parameter $\theta = (m_0, \gamma_{\bar{k}}, b, \sigma_\delta)'$ obtained from 100 simulated sample paths of length $T = 20,000$. The sample paths are generated from the learning model with $\bar{k} = 3$ components and parameters $m_0 = 1.7$, $\gamma_{\bar{k}} = 0.06$, $b = 2$ and $\sigma_\delta = 1$. We set the SMM and II simulation sizes to $S = 500$, so that each simulated path contains $ST = 10^7$ simulated data points. The II auxiliary estimator consists of the full-information MLE and the third moment or the median. The boxplots show that II estimation not only corrects for the negative bias of the auxiliary estimator of b apparent in Figure 2, but also provides an accurate estimate of the parameter σ_δ. The Monte Carlo simulations therefore confirm the accuracy of the median-based II estimator of θ, which dominates the II estimator based on the third moment.

In Figure 6, we report the accuracy of the particle filter on a simulated sample of size

[11] Calvet and Fisher (2002) and Lux (2008) consider predetermined values of the persistence parameters $\gamma_{\bar{k}}$ and b, and conduct the GMM estimation of the other parameters of a binomial MSM process, a much simpler process than the structural model considered in the present paper. To the best of our knowledge, no GMM method has so far been developed to accurately estimate $\gamma_{\bar{k}}$.

Table 1: SUMMARY STATISTICS ON LOG EXCESS RETURNS

Sample Period	Mean (%)	Standard dev. (%)	Skewness coeff.	Kurtosis coeff.
Jan 2, 1926 - Dec 31, 2009	0.021	1.064	-0.434	21.13
Jan 2, 1926 - Dec 31, 1999	0.025	1.012	-0.495	24.35

Notes. We report summary statistics on the daily log excess returns of the CRSP U.S. Value-Weighted Equity Index. The estimated are calculated for the full sample (1926-2009) and the in-sample period (1926-1999).

20,000, from a learning model with $\overline{k} = 3$ and parameters $m_0 = 1.7$, $\gamma_{\overline{k}} = 0.06$, $b = 2$ and $\sigma_\delta = 1$. In the left panel we report the estimated likelihood function for various number of particles N. In the right panel we illustrate the efficiency of the filter by reporting the following R^2 measure:

$$\tilde{R}^2_Q = 1 - \frac{\sum_{t=1}^{T}(\hat{Q}_t - Q_t)^2}{\sum_{t=1}^{T}(\hat{Q}_t - \bar{Q}_t)^2},$$

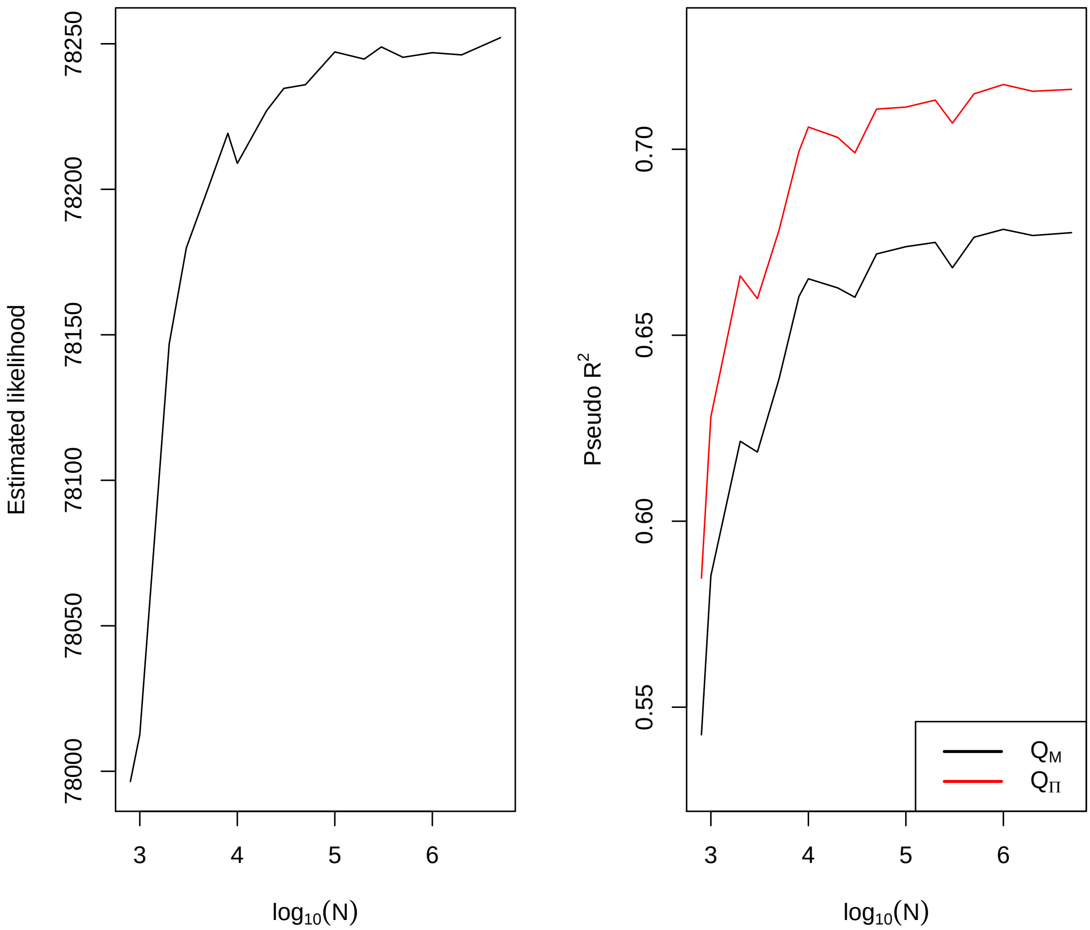
where $Q_t = Q(\Pi_t)$ denotes the price-dividend ratio computed by the agent at date t, $\hat{Q}_t = \sum_{n=1}^{N} Q(\Pi_t^{(n)})/N$ the sample mean of the N particles for Q_t at date t, and $\bar{Q}_t = \sum_{t=1}^{T} Q_t/T$. We also consider the price-dividend ratio that nature could compute, $Q(M_t)$, and report the R^2 of the estimates $\sum_{n=1}^{N} Q(M_t^{(n)})/N$. This figure shows that both the estimated likelihood and the efficiency increases with the number of particles used and the patterns settle down around $N = 10^7$. The right panel shows that $Q(\Pi_t)$ is better estimated than $Q(M_t)$ which is consistent with the information structure in Figure (1).

We test the accuracy of our dynamic test by considering the consistency condition (3) between the beliefs of the agent and the beliefs and the econometrician. The equality between conditional probabilities can be mapped into a condition about the price-dividend ratio. For a given state $s_t = (M_t, \Pi_t)$, we can compute the price-dividend ratio that nature could compute, $Q(M_t) \equiv Q[\Pi(M_t)]$, as well as the price-dividend ratio $Q(\Pi_t)$ that the investor computes. Since the price-dividend ratio is linear in beliefs, we infer that

$$\mathbb{E}[Q(M_t)|R_t] = \mathbb{E}[Q(\Pi_t)|R_t].$$

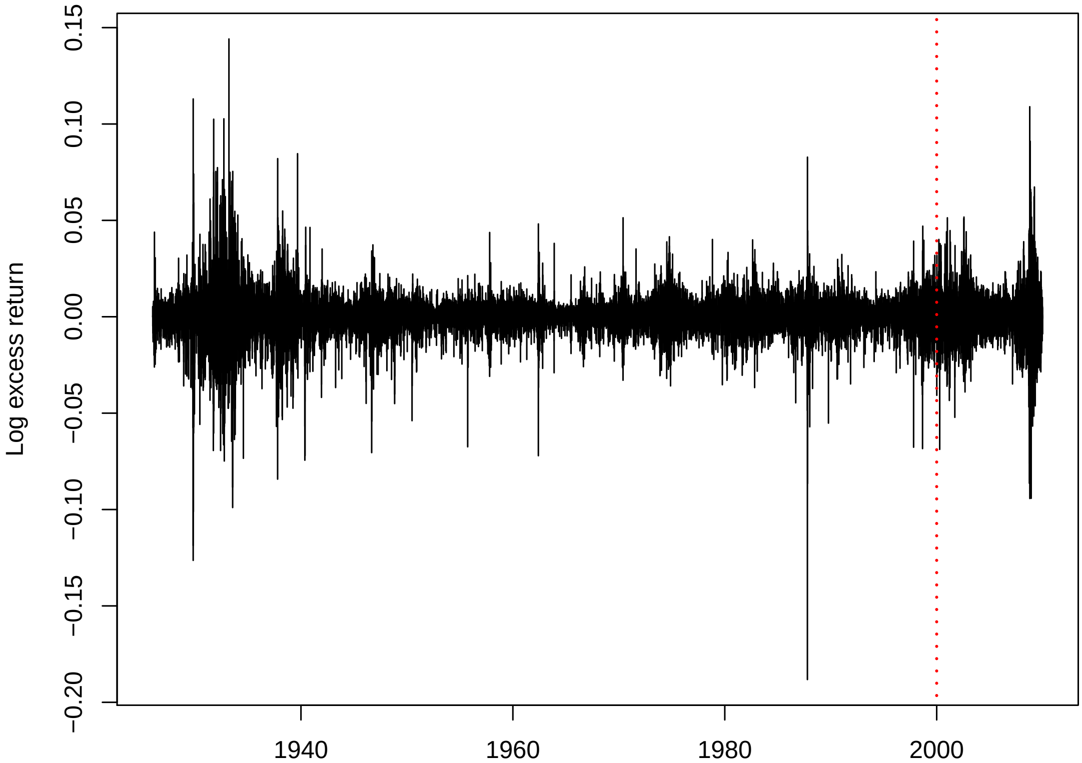
We test this assumption on a simulated sample of size 20,000 by following the methodology discussed in section 3.3, with the function $\Phi(M_t, \Pi_t) = Q(\Pi_t) - Q(M_t)$. We check that the 5% test is correctly sized.

Figure 6: ACCURACY OF THE PARTICLE FILTER



Notes. This figure illurtates the simulated likelihood (left panel) and the pseudo R^2 (right panel) as a function of the number of particles N. The original sample of size 20,000 was simulated from a learning model with $\overline{k} = 3$ and parameters $m_0 = 1.7$, $\gamma_{\overline{k}} = 0.06$, $b = 2$ and $\sigma_\delta = 1$.

Figure 7: CRSP VALUE-WEIGHTED EQUITY INDEX (1926-2009)



Notes: The figure illustrates the time series of log excess returns on the CRSP U.S. value-weighted equity index from 2 January 1926 to 31 December 2009. The dashed line separates the in-sample period (1926-1999) from the out-of-sample period (2000-2009).

Table 2: EMPIRICAL ESTIMATES

A. Auxiliary Estimates

$\overline{k}$	Parameter Estimates			Likelihood (in logs)	Moments of Daily Returns			
	m_0	$\gamma_{\overline{k}}$	b	$\mathcal{L}$	Mean (%)	S.d. (%)	Skew. coeff.	Kurt. coeff.
1	1.736 (0.0058)	0.068 (0.0063)	-	64,455.9	0.019	0.701	-0.036	4.61
2	1.708 (0.0052)	0.062 (0.0063)	5.698 (0.6485)	66,743.6	0.019	0.711	-0.072	6.69
3	1.688 (0.0070)	0.084 (0.0068)	6.109 (0.3228)	67,787.6	0.019	0.750	-0.092	17.01
4	1.591 (0.0074)	0.068 (0.0079)	3.358 (0.1918)	68,172.0	0.019	0.750	-0.091	16.39
Data					0.025	1.012	-0.495	24.35

B. II Estimates of the Learning Model and Simple Filter

$\overline{k}$	Parameter Estimates				Estimated Likelihood (in logs)	Moments of Daily Returns			
	m_0	$\gamma_{\overline{k}}$	b	σ_δ		Mean (%)	S.d. (%)	Skew. coeff.	Kurt. coeff.
1	1.732 (0.0091)	0.063 (0.0033)	-	93.807 (61616.8162)	66,782.6 (10.8825)	0.019	0.701	-0.154	4.64
2	1.714 (0.0061)	0.054 (0.0036)	21.104 (10.5573)	4.001 (1.1036)	67,531.4 (7.4977)	0.019	0.741	-0.683	9.27
3	1.690 (0.0055)	0.071 (0.0055)	16.471 (9.9115)	2.401 (1.5599)	67,742.5 (7.3317)	0.020	0.876	-2.508	116.00
4	1.587 (0.0059)	0.047 (0.0049)	5.089 (0.5387)	1.411 (0.1714)	68,200.5	0.019	0.828	-1.250	42.35
Data						0.025	1.012	-0.495	24.35

Notes. In Panel A, we report empirical estimates based on the daily excess returns of the CRSP U.S. value-weighted equity index between 2 January 1926 and 31 December 1999. Standard errors are reported in parentheses. The likelihood estimates are based on a particle filter containing $N = 2 \cdot 10^7$ elements. For each value of $\bar{k}$, the moments are computed on a long simulated sample containing $ST = 10^7$ elements. In Panel B, we report empirical estimates of the full information model based on the daily excess returns on the CRSP value-weighted index between 2 January 1926 and 31 December 1999. Standard errors are reported in parentheses.

5 Empirical Results

5.1 Estimation

We apply our estimation methodology to the daily log excess returns on the U.S. CRSP value equity index from 2 January 1926 to 31 December 2009. The dataset contains 22,276 observations, which are illustrated in Figure 7. We partition the dataset into an in-sample period, which runs between 2 Jan 1926 and 31 Dec 1999, and an out-of-sample period, which covers the remaining 10 years. Summary statistics are reported in Table 1. Daily returns have a mean of about 2 basis points, a standard deviation of 1%, a skewness coefficient of the order of -0.5, and a kurtosis of about 20.

In panel A of Table 2, we report the in-sample MLE of the full-information model for $\overline{k}$ ranging between 1 and 4. Standard errors are reported in parentheses. The parameter m_0, which controls the variability of volatility, monotonically declines from 1.74 when $\overline{k} = 1$ to 1.59 when $\overline{k} = 4$. As more volatility components are added, each of them needs to be less variable to match the data. The transition probability of the most transitory component, $\gamma_{\overline{k}}$, hovers around 0.07. The spacing parameter b, which is undefined when there is a single component, is close to 6 when $\bar{k} = 2$ and then tends to decrease with $\bar{k}$. The likelihood function gains almost 4,000 points as $\overline{k}$ increases from 1 to 4, which is strongly significant at all standard significance levels.

In the last four columns, we report the moments of daily excess returns under the estimated FI models. The mean return is approximately constant, as implied by (14). The standard deviation and kurtosis of returns goes up monotonically with $\bar{k}$. Additional components tend to be increasingly persistent and therefore have increasingly large effects on the price-dividend ratio. The skewness coefficient becomes slightly more negative as $\bar{k}$ increases, but remains far from its empirical estimate.

In panel B of Table 2, we report the II estimates of θ. We let $ST = 10^7$ and report standard errors in parentheses. The noise parameter σ_δ is significantly positive, which is a first indication that the learning model provides a better fit than its FI version. The point estimate σ_δ declines with $\overline{k}$. As more volatility components are added, the effect of learning becomes increasingly powerful, so the estimated value of σ_δ declines in order to match the fixed negative skewness of excess returns.

The II estimate of m_0 coincides with the auxiliary estimate when $\overline{k} = 1$, and takes slightly lower values for larger $\overline{k}$. Consistent with the sensitivity plots reported in Figure 2, the difference between the ML and II estimates of m_0 is very modest. The II estimates of $\gamma_{\bar{k}}$ are slightly smaller than the corresponding MLEs. The II estimates of b are substantially higher, and the difference is significant. Overall, the volatility components of the inferred learning model are slightly more volatile and persistent than the volatility components of the inferred FI model. Intuitively, this suggests that the learning model can capture the tails and the skewness of returns better than the full information model; at the same time, investor learning attenuates the day-to-day variations of changes in volatility components, so that typical returns are not too extreme.

We also report the estimated likelihood of each specification, which is computed by using a particle filter of $N = 20$ million particles every period. We observe that the likelihood function of the II model increases steadily with $\bar{k}$. A Vuong test shows that the specification with $\bar{k} = 4$ components dominates the specifications with $\bar{k} \leq 3$. We can also compare the likelihoods of the II and FI models with the same $\bar{k}$. Since the FI model is nested in the II specification, a likelihood ratio can be used. We verify that the II model strongly dominates the FI model. The exception is $\bar{k} = 3$, but this is due to simulation error and we are currently reestimating this value.

The moments of daily returns under the estimated II models are reported in the last set of columns of Table 2. The first moment is approximately constant. For each $\bar{k}$, the standard deviation of returns is substantially higher and closer to the empirical estimate under the II model than under full-information. The skewness coefficient is substantially negative.

5.2 Value at Risk

We now turn to the out-of-sample implications of the incomplete information model. The value at risk VaR^p_{t+1} constructed on day t is such that the return on day $t+1$ will be lower than $-Var^p_{t+1}$ with probability p. The failure rate is specified as the fraction of observations where the actual return exceeds the value at risk. In a well specified VaR model, the failure rate is on average equal to p.

We use as a benchmark historical simulations, which are widely used in practice (e.g. Christoffersen 2009). The historical VaR estimates are based on a window of 60 days, which corresponds to a calendar period of about three months. At a given date t, the 5% and the 10% VaR estimates of the one-step ahead return are the third and the sixth smallest elements in the set $\{r_{t-59}, \ldots, r_t\}$.

Table 3: FAILURE RATES OF VALUE-AT-RISK FORECASTS

Models	One Day			Five Days			Ten Days		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
Historical VaR	–	**0.069** (0.0051)	**0.119** (0.0065)	–	0.066 (0.0111)	0.129 (0.0150)	–	0.064 (0.0155)	0.112 (0.0199)
GARCH	**0.081** (0.0054)	**0.154** (0.0072)	**0.197** (0.0079)	**0.048** (0.0095)	**0.123** (0.0147)	**0.165** (0.0166)	0.032 (0.0111)	0.064 (0.0155)	0.135 (0.0216)
$FI, \overline{k} = 4$	0.016 (0.0025)	**0.070** (0.0051)	**0.132** (0.0067)	0.012 (0.0048)	0.068 (0.0112)	**0.143** (0.0156)	0.020 (0.0088)	0.048 (0.0134)	0.087 (0.0178)
$II, \overline{k} = 4$	0.012 (0.0022)	0.060 (0.0047)	0.114 (0.0063)	0.008 (0.0040)	0.054 (0.0101)	0.131 (0.0151)	0.016 (0.0079)	0.036 (0.0117)	0.095 (0.0185)

Notes. This table reports the failure rates of the 1-day, 5-day and 10-day value at risk forecasts produced by various methods in the out-of-sample period (2000-2009). The historical VaR is based on a rolling window of 60 days. The FI, and II forecasts are computed using the in-sample parameter estimates reported in Table 2. The FI forecasts are based on $N = 40,000$ simulations. The II forecasts are based on a particle filter with $N = 10^7$ elements. Significance level is 1%.

In Table 3, we report the failure rates of the VaR forecasts at horizons of 1, 5 and 10 days produced by: historical simulations, GARCH, the full information model and the learning model with $\bar{k} = 4$. Standard deviations are computed as in Kupiec (1995) and are reported in parentheses. A failure rate is in bold characters if it differs from its theoretical value at the 5% significance level.

Historical simulations provide inaccurate VaR forecasts at the 1-day horizon. The failure rates are significantly higher than their theoretical values, which suggests that historical simulations provide overly optimistic estimates of value at risk. GARCH VaR estimates are rejected at the 5% significance level in seven out of nine cases. The failure rates all exceed their theoretical values, which suggests that GARCH predictions tend to overestimate risk. The FI model's VaR predictions are rejected at the 5% significance level in four out of nine cases.

We next turn to the VaR predictions from the learning model. The VaR estimate is rejected only once out of nice cases at the 5% level. Our empirical findings suggest that the learning model captures well the dynamics of stock returns, and outperforms in and out of sample some of the best reduced-form specifications, as exemplified by our value at risk estimates. We note that this is an excellent result for a consumption-based asset pricing model.

6 Conclusion

In this paper, we have developed a powerful estimation method based on indirect inference for a wide class of learning models. The approach can be applied when the structural model has a closed-form likelihood under full information. An auxiliary estimator can then be defined by expanding the full-information MLE with a set of statistics that investor learning is designed to capture. For given parameter estimates, we have developed a particle filter tracking the joint distribution of the state nature and the investor belief. The filter is useful to compute the simulated likelihood of the learning model, conduct specification tests, and provide forecasts. We have also developed an accurate simulation-based test of a moment condition at a given point in time.

We have applied this method to a model of investor learning about multifrequency volatility. The moment-based estimation of such a model is challenging because the volatility components are latent and their frequencies do not affect the unconditional distribution of returns. By contrast, II estimation permits to use Bayesian filtering via the closed-form likelihood function of the full-information model. We have verified the accuracy of our estimator by Monte Carlo simulations, and have applied it to a long U.S. aggregate equity series. We have estimated the information quality received by investors and demonstrated the good performance of the inferred model both in- and out-of-sample.

The paper opens multiple directions for future research. In an ongoing investigation, we are applying the particle filter to design asset allocation and trading strategies that take into account both fundamentals and the distribution of investor beliefs about fundamentals. The particle filter can be used to test a variety of hypotheses about the state of the learning economy, and price more complex instruments, such as derivatives contracts, which crucially depend on the distribution of this state. Further extensions could include estimation and inference in equilibrium models with asymmetric information (e.g. Bossaerts, Biais, and Spatt, 2010), and the development of value-at-risk models that incorporate the cross-sectional dispersion of investor beliefs.

Appendix

A.1. Likelihood of the Full-Information Model (Section 2.3)

The econometrician recursively applies Bayes'rule:

$$\mathbb{P}(M_t = m^j | R_t; \phi) = \frac{f(r_t | M_t = m^j, R_{t-1}; \phi)\mathbb{P}(M_t = m^j | R_{t-1}; \phi)}{f(r_t | R_{t-1}; \phi)}$$

Since $f(r_t|M_t = m^j, R_{t-1}; \phi) = \sum_{i=1}^{d} f_{i,j}(r_t; \phi)\mathbb{P}(M_{t-1} = m^i | M_t = m^j, R_{t-1}; \phi)$, we infer that $f(r_t|M_t = m^j, R_{t-1}; \phi)\mathbb{P}(M_t = m^j|R_{t-1}; \phi) = \sum_{i=1}^{d} f_{i,j}(r_t; \phi)\mathbb{P}(M_{t-1} = m^i, M_t = m^j|R_{t-1}; \phi)$, and therefore

$$\mathbb{P}(M_t = m^j | R_t; \phi) = \frac{\sum_{i=1}^{d} a_{i,j} f_{i,j}(r_t; \phi)\mathbb{P}(M_{t-1} = m^i | R_{t-1}; \phi)}{f(r_t | R_{t-1}; \phi)}.$$

The econometrician's conditional probabilities are therefore computed recursively.

Since the conditional probabilities $\mathbb{P}(M_t = m^j|R_t; \phi)$ add up to unity, the conditional density of r_t satisfies

$$f(r_t | R_{t-1}; \phi) = \sum_{i=1}^{d} \sum_{j=1}^{d} a_{i,j} f_{i,j}(r_t; \phi)\mathbb{P}(M_{t-1} = m^i | R_{t-1}; \phi).$$

The likelihood function $\ln \mathcal{L}(\phi|R_T) = \sum_{t=1}^{T} \ln f(r_t|R_{t-1}; \phi)$ thus has an analytical expression.

A.2. Efficient Method of Moments (Section 3.1)

The numerical implementation is facilitated by the Efficient Method of Moments when the auxiliary estimator $\hat{\mu}_T$ defined by (4) maximizes a criterion function $\mathcal{Q}_T$:

$$\hat{\mu}_T = \arg\max_{\mu} \mathcal{Q}_T(\mu, R_T). \tag{15}$$

We also require two additional conditions. First, under the structural model with parameter θ^*, the criterion function $\mathcal{Q}_T(\mu, R_T)$ converges in probability to a deterministic function $\mathcal{Q}_\infty(\mu, \theta)$ as $T \to \infty$. Second, we also assume that the score function can be decomposed as:

$$\frac{\partial \mathcal{Q}_T}{\partial \mu}(\mu, R_T) \equiv \frac{1}{T} \sum_{t=1}^{T} \psi(r_t | R_{t-1}; \mu).$$

For instance, the criterion function can be the sum of the conditional densities and a sum of moments. By (15), the auxiliary parameter satisfies the first-order condition:

$$\frac{\partial \mathcal{Q}_T}{\partial \mu}(\hat{\mu}_T, R_T) = \frac{1}{T}\sum_{t=1}^{T} \psi(r_t|R_{t-1}; \hat{\mu}_T) = 0. \tag{16}$$

Let $W^* = \lim_T \mathrm{Var}_{\theta^*}\left\{\sqrt{T}(\partial \mathcal{Q}_T/\partial\mu)[\mu(\theta^*), R_T]\right\}$ denote the limiting variance of the rescaled score under the structural model with parameter θ^*.

When the weighting matrix Ω converges to W^*, the II estimator $\hat{\theta}_T$ is asymptotically efficient and has asymptotic variance-covariance matrix:

$$V = \left(1 + \frac{1}{S}\right)\left\{\frac{\partial^2 \mathcal{Q}_\infty}{\partial\theta\partial\mu'}[\mu(\theta^*), \theta^*] W^* \frac{\partial^2 \mathcal{Q}_\infty}{\partial\mu\partial\theta'}[\mu(\theta^*), \theta^*]\right\}^{-1}.$$

Note that this result uses the fact that the auxiliary and structural parameters $\hat{\mu}_T$ and $\hat{\theta}_T$ have the same dimension p.

In practice, we begin the procedure by computing $\hat{\mu}_T$ and setting the weighting matrix W_T equal to the Newey and West variance-covariance matrix:

$$W_T = \left[\hat{\Gamma}_0 + \sum_{v=1}^{\tau}\left(1 - \frac{v}{\tau+1}\right)\left(\hat{\Gamma}_v + \hat{\Gamma}_v'\right)\right]^{-1}, \tag{17}$$

where $\hat{\Gamma}_v = T^{-1}\sum_{t=v+1}^{T} \psi(r_t|R_{t-1}; \hat{\mu}_T)\psi(r_t|R_{t-1}; \hat{\mu}_T)'$. All the results reported in the paper are based on $\tau = 10$ lags. We then solve the optimization problem (**??**). Once $\hat{\theta}_T$ has been obtained, we approximate $(\partial^2 \mathcal{Q}_\infty/\partial\theta\partial\mu')[\mu(\theta^*), \theta^*]$ by $(\partial^2 \mathcal{Q}_{ST}/\partial\theta\partial\mu')[\hat{\mu}_T, R_{ST}(\hat{\theta}_T)]$, and obtain a finite-sample estimate of the asymptotic variance-covariance matrix V.

A.3. Convergence of the Particle Filter (Section 3.2)

For any bounded continuous function Φ, we infer from (9) that the expectation of $\Phi(s_t)$ conditional on the data R_t satisfies:

$$\mathbb{E}\left[\Phi(s_t)|R_t\right] = \int \Phi(s_t)\frac{\delta[r_t - \mathcal{R}(x_t, \Pi_t, \Pi_{t-1})]}{f(r_t|R_{t-1})} f(s_t, x_t, \Pi_{t-1}|R_{t-1}) d(s_t, x_t, \Pi_{t-1}).$$

The kernel K_h converges to the Dirac distribution as the bandwidth h converges to zero, which implies

$$\mathbb{E}\left[\Phi(s_t)|R_t\right] = \lim_{h\to 0} \int \Phi(s_t) \frac{K_h[r_t - \mathcal{R}(x_t, \Pi_t, \Pi_{t-1}); V_t]}{f(r_t|R_{t-1})} f(s_t, x_t, \Pi_{t-1}|R_{t-1}) d(s_t, x_t, \Pi_{t-1}).$$

Since $\{(s_{t-1}^{(n)}, \tilde{x}_t^{(n)}, \tilde{s}_t^{(n)})\}_{n=1,...,N}$ targets $f(s_{t-1}, x_t, s_t|R_{t-1})$, we infer from standard filtering theory (e.g. Crisan and Doucet, 2001) that $N^{-1}\sum_{n=1}^{N}\Phi(\tilde{s}_t^{(n)})$ converges almost surely to a deterministic limit $L(h)$. Since K_h converges to the Dirac distribution as $h \to 0$, we infer that $L(h)$ converges to $\mathbb{E}\left[\Phi(s_t)|R_t\right]$, and conclude that (10) holds.

A.4. Out-of-Sample Analysis (Section 3.2)

We can provide provide forecasts of the data points p steps ahead. We estimate the structural parameter vector θ in the in-sample period $t = 1, ..., T$. We then recursively forecast the conditional distribution of returns as follows.

Step 1. Consider the particle filter $s_t^{(1)}$, ... $s_t^{(n)}$ at date t.

- We pick the first particle $s_t^{(1)}$.
- We simulate forward the process one-step ahead from $s_t^{(1)}$. We reach the new state $\tilde{s}_{t+1}^{(1)}$ and compute the simulated data point $\tilde{r}_{t+1}^{(1)}$.
- We simulate forward the process one-step ahead from $\tilde{s}_{t+1}^{(1)}$. We reach the new state $\tilde{s}_{t+2}^{(1)}$ and compute the simulated data point $\tilde{r}_{t+2}^{(1)}$.
- We repeat until we obtain $\tilde{s}_{t+p}^{(1)}$ and compute the simulated data point $\tilde{r}_{t+p}^{(1)}$.

We repeat this procedure with the other particles $s_t^{(2)}, ..., s_t^{(N)}$. We now have N vectors $(\tilde{r}_{t+1}^{(n)}, ..., \tilde{r}_{t+p}^{(n)})$, $n \in \{1, ..., N\}$, which track $f(r_{t+1}, ..., r_{t+N})$ conditional on the history R_t.

Step 2. We observe the data points $r_{t+1}, ..., r_{t+p}$. That is, we recursively compute the particle filter at dates $t+1, ..., t+p$, as was explained in the previous section. We are ready to start again.

A.5. Bayesian Updating (Section 4.2)

The signal received by an investor at date t is $x_t = (\Delta c_{t+1}, \Delta d_{t+1}, \delta'_{t+1})'$. We know that conditional on $M_t = m^j$,

$$x_t \sim \mathcal{N}\left[\begin{pmatrix} g_c \\ g_d - \frac{\sigma_d^2(m^j)}{2}\end{pmatrix}, \begin{pmatrix} \sigma_c^2 & \sigma_c\sigma_d(m^j)\rho_{c,d} \\ \sigma_c\sigma_d(m^j)\rho_{c,d} & \sigma_d^2(m^j)\end{pmatrix}\right] \mathcal{N}\left(m^j, \sigma_\delta^2 I_{\overline{k}}\right),$$

The probabilities Π_t^j are recursively determined from the initial value $\Pi_1 = \iota/d$ and Bayes's rule (1).

A.6. Generator of Pseudo Data (Section 4.5)

We can estimate the structural parameter θ using pseudo-data $\{r_t^*(\theta)\}_{t=2,\ldots,ST}$, $S \geq 1$, simulated from the structural model F_θ. The pseudo-data can be generated from F_θ for a given θ in the following way.

Simulate the following variates:

- j_0 from a discrete uniform $\mathcal{U}\{1,\ldots,d\}$;
- $\{u_{k,t}^*\}_{t=1,\ldots,ST-1}^{k=1,\ldots,\overline{k}}$ with $u_{k,t}^* \sim \mathcal{U}[0,1]$;
- $\{\varepsilon_t^*\}_{t=1}^{ST-1}$ with
$$\varepsilon_t^* = \begin{pmatrix}\varepsilon_{c,t}^* \\ \varepsilon_{d,t}^*\end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}1 & \rho_{c,d} \\ \rho_{c,d} & 1\end{pmatrix}\right];$$
- $\{z_t^*\}_{t=1,\ldots,ST-1}$ where $z_t^* = (z_{t,1}^*,\ldots,z_{t,\overline{k}}^*)'$ with $z_{t,k}^* \sim \mathcal{N}(0,1)$.
- $\{N_{k,t}^*\}_{t=1,\ldots,ST-1}^{k=1,\ldots,\overline{k}}$ with
$$N_{k,t}^* \sim Bernoulli(0.5).$$

The pseudo-data is then obtained by calculating $M_1^* = m^{j_0}$ and

$$M_{k,t}^* = \begin{cases} \left(2 - m_0\right)^{N_{k,t}^*} m_0^{1-N_{k,t}^*} & \text{if } u_{k,t}^* < \gamma_k, \\ M_{k,t-1}^* & \text{otherwise.}\end{cases}$$

the other variables are defined by:

$$\begin{cases} \Delta c_t^*(\theta) = g_c + \sigma_c \varepsilon_{c,t}^* \,, \\ \Delta d_t^*(\theta) = g_d - \dfrac{\sigma_d^2(M_t^*(\theta))}{2} + \sigma_d(M_t^*(\theta))\varepsilon_{d,t}^* \,, \\ \delta_t = M_t^*(\theta) + \sigma_\delta z_t^* \\ r_t^*(\theta) = \ln \dfrac{1 + Q[\pi_t^*(\theta)]}{Q[\pi_{t-1}^*(\theta)]} + g_d - r_f - \dfrac{\sigma_d^2(M_t^*(\theta))}{2} + \sigma_d(M_t^*(\theta))\varepsilon_{d,t}^* \,. \end{cases} \tag{18}$$

for all $t \geq 2$.

References

Abel, A. B. (1994). Exact Solutions for Expected Rates of Return under Markov Regime Switching: Implications for the Equity Premium Puzzle. *Journal of Money, Credit, and Banking* **26(3)**, 345–361.

Abel, A. B. (1999). Risk Premia and Term Premia in General Equilibrium. *Journal of Monetary Economics* **43(1)**, 3–33.

Biais, B., Bossaerts, P., and Spatt, C. (2010). Equilibrium Asset Pricing and Portfolio Choice Under Asymmetric Information. Forthcoming, *Review of Financial Studies.*

Bollerslev, T., Litvinova, L., and Tauchen, G. (2006). Leverage and Volatility Feedback Effects in High-Frequency Data. *Journal of Financial Econometrics* **4**, 353–384.

Bonomo, M., and Garcia, R. (1994). Can a Well-Fitted Equilibrium Asset-Pricing Model Produce Mean Reversion. *Journal of Applied Econometrics* **9(1)**, 19–29.

Brennan, M. (1998). The Role of Learning in Dynamic Portfolio Decisions. *European Finance Review* **1**, 295–396.

Brennan, M. and Xia Y. (2001). Stock Price Volatility and the Equity Premium. *Journal of Monetary Economics* **47**, 249–283.

Calvet, L. E., and Fisher, A. J. (2001). Forecasting Multifractal Volatility. *Journal of Econometrics* **105**, 27–58.

Calvet, L. E., and Fisher, A. J. (2004). How to Forecast Long-Run Volatility: Regime-Switching and the Estimation of Multifractal Processes. *Journal of Financial Econometrics* **2**, 49–83.

Calvet, L. E., and Fisher, A. J. (2007). Multifrequency News and Stock Returns. *Journal of Financial Economics* 86: 178-212.

Calvet, L. E., and Fisher, A. J. (2008). *Multifractal Volatility: Theory, Forecasting and Pricing.* Elsevier - Academic Press.

Calvet, L. E., Fisher, A. J. and Thompson S. (2006). Volatility Comovement: A Multifrequency Approach. *Journal of Econometrics* **131**, 179–215.

Calzolari G., Fiorentini, G. and Sentana, E. (2004). Constrained Indirect Estimation. *Review of Economic Studies* **71**, 945–973.

Campbell, J. Y., and Viceira, L. (2002). *Strategic Asset Allocation.* Oxford University Press.

Cecchetti, S. G., Lam, P. and Mark, N. C. (1990). Mean Reversion in Equilibrium Asset Prices. *American Economic Review* **80(3)**, 398–418.

Cecchetti, S. G., Lam, P. and Mark, N. C. (2000). Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good to Be True? *American Economic Review* **90(4)**, 787–805.

Chigansky, P., (2006). An Ergodic Theorem for Filtering with Applications to Stability. *Systems and Control Letters* **55(11)**, 908–917.

Chopin, N. (2004). Central Limit Theorem for Sequential Monte Carlo Methods and its Applications to Bayesian Inference. *Annals of Statistics* **32(6)**, 2385–2411.

Christoffersen, P. (2009). Backtesting. In: *Encyclopedia of Quantitative Finance*, R. Cont (ed)., John Wiley and Sons.

Crisan, D., and Doucet, A. (2002). A Survey of Convergence Results on Particle Filtering Methods for Practitioners. *IEEE Transactions on Signal Processing* **50(3)** 736–746.

Czellar, V., Karolyi, G. A. and Ronchetti, E. (2007). Indirect Robust Estimation of the Short-Term Interest Process. *Journal of Empirical Finance* **14**, 546–563.

Czellar, V. and Ronchetti E. (2010). Accurate and Robust Tests for Indirect Inference. *Biometrika* **97(3)**, 621–630.

David, A. (1997). Fluctuating confidence in stock markets: implications for returns and volatility. *Journal of Financial and Quantitative Analysis* **32**, 427–462.

Dridi, R., Guay, A. and Renault, E. (2007). Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Model. *Journal of Econometrics* **136**, 397–430.

Epstein, L. G. and Zin, S. E. (1989). Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. *Econometrica* **57**, 937–969.

Gouriéroux, C., Monfort, A. and Renault, E. (1993). Indirect Inference. *Journal of Applied Econometrics* **8**, S85–S118.

Guidolin, M. and Timmermann, A. (2003). Option prices under Bayesian learning: implied volatility dynamics and predictive densities. *Journal of Economic Dynamics and Control* **27**, 717–769.

Heggland, K. and Frigessi, A. (2004). Estimating Functions in Indirect Inference. *Journal of the Royal Statistical Society Series B*, **66**, 447–462.

Huber, P. J. (1981). *Robust Statistics.* New York: John Wiley and Sons.

Idier, J. (2010). (Re)Correlation: A Markov Switching Multifractal Model with Time Varying Correlations. Working paper, Bank of France.

Kaijser, T. (1975). A Limit Theorem for Partially Observed Markov Chains. *Annals of Probability* **3(4)**, 677–696.

Kandel and Stambaugh (1990). Expectations and Volatility of Consumption and Asset Returns. *Review of Financial Studies* **3(2)**, 207–232.

Kupiec, P. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *Journal of Derivatives* **3**, 73–84.

Lettau, M., Ludvigson, S. and Wachter J. (2008). The Declining Equity Premium: What Role Does Macroeconomic Risk Play? *Review of Financial Studies* **21**, 1653–1687.

Ljungqvist, L., and Sargent, T. J. (2000). *Recursive Macroeconomic Theory.* MIT Press.

Lux, T. (2008). The Markov-Switching Multifractal Model of Asset Returns: GMM Estimation and Linear Forecasting of Volatility. *Journal of Business and Economic Statistics* **26**, 194–210.

Pagan, A. R. and Schwert, G. W. (1990). Alternative Models for Conditional Stock Volatility. *Journal of Econometrics* **45**, 267–290.

Pastor, L., and Veronesi, P. (2009). Technological Revolutions and Stock Prices. *American Economic Review* **99**, 1451–1483.

Schwert, G. W. (1989). Why Does Stock Market Volatility Change over Time? *Journal of Finance* **44**, 1115–1153.

Schwert, G. W. (1990). Stock Returns and Real Activity: A Century of Evidence *Journal of Finance* **45**, 1237–1257.

Sentana, E., Calzolari, G. and Fiorentini, G. (2008). Indirect Estimation of Large conditionally Heteroskedastic Factor Models, with an Application to the Dow 30 Stocks, *Journal of Econometrics* **146**, 10–25.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall.

Smith, A. A. (1993). Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics* **8**, S63–S84.

Stokey, N. and Lucas, R. (1989). *Recursive Methods in Economic Dynamics.* Harvard University Press.

Timmermann, A. G. (1993). How learning in financial markets generates excess volatility and predictability in stock prices. *Quarterly Journal of Economics* **108**, 1135–1145.

Timmermann, A. G. (1996). Excess Volatility and Predictability of Stock Prices in Autoregressive Dividend Models with Learning. *Review of Economic Studies* **63**, 523–557.

Van Handel, R. (2009). Uniform Time Average Consistency of Monte Carlo Particle Filters. *Stochastic Processes and Their Applications* **119**, 3835–3861.

Van Nieuwerburgh, S. and Veldkamp, L. (2006). Learning Asymmetries in Real Business Cycles. *Journal of Monetary Economics* **53**, 753–772.

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

Veronesi, P. (2000). How does information quality affect stock returns? *Journal of Finance* **55**, 807–837.

Wang, J. (1994). A model of competitive stock trading volume. *Journal of Political Economy* **102(1)**, 127–168.

Weil, P. (1989). The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics* **24**, 401-421.

Whitelaw (2000). Stock Market Risk and Return: An Equilibrium Approach. *Review of Financial Studies* **13(3)**, 521–547.