Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Empirical Finance 12 (2005) 374–417
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/econbase

# Equilibrium analysis of volatility clustering

Joel M. Vanden*

*Tuck School of Business, Dartmouth College, Hanover, NH 03755, USA*

Accepted 5 April 2004
Available online 19 December 2004

**Abstract**

Volatility clustering is a pervasive feature of equity markets. This article studies volatility clustering in an equilibrium setting by generalizing the CRRA and CARA representative agent models of finance. In equilibrium, the market portfolio follows a volatility regime-switching process in which the volatility level is determined by the agent's local risk aversion. Using monthly data, the empirical tests reveal that at least four volatility regimes are necessary to fit the data. While one of the models explains the GARCH effects in the data, an analysis of the Euler equation pricing errors suggests that both models are likely misspecified. Since the models can be used to closely approximate any state-independent utility function, it is doubtful that there exists any representative agent equilibrium (with state-independent utility) that is consistent with the data. An equivalent interpretation is that the market portfolio price process is not a diffusion process of the type studied by Bick [Bick, A., On viable diffusion price processes of the market portfolio, J. Finance 45 (1990) 673–689] and He and Leland [He, H., Leland, H., On equilibrium asset price processes, Rev. Financ. Stud. 6 (1993) 593–617].


*JEL Classification:* G0; G1; C5
*Keywords:* Implied risk aversion; Volatility clustering; Time-varying risk aversion; Equilibrium price process

## 1. Introduction

Over the past few decades the finance industry has taken an increased interest in understanding the volatility pattern of stock market returns. Indeed, a clear knowledge of return volatility is necessary for making sound investment decisions, for implementing risk

* Tel.: +1 603 646 1832; fax: +1 603 646 1308.
*E-mail address:* joel.m.vanden@dartmouth.edu.

0927-5398/$ - see front matter 
doi:10.1016/j.jempfin.2004.04.007

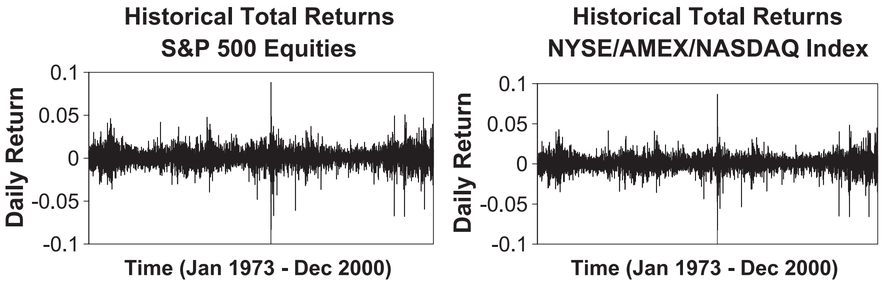


Fig. 1. The unannualized historical daily total return is shown for two popular market proxies, the value-weighted S&P 500 portfolio and the value-weighted NYSE/AMEX/NASDAQ portfolio.

management techniques and for pricing derivative assets. A particular feature of volatility that has received considerable attention is volatility clustering (see Fig. 1). As described by Campbell et al. (1997, pp. 481–482), small returns (of either sign) tend to be followed by more small returns and large returns (of either sign) tend to be followed by more large returns. According to these authors, the returns vary through time within a range that also appears to be changing through time. While the statistical properties of volatility clustering are well understood, the economic drivers of volatility clustering have received considerably less attention.[1] This article sheds some light on the economic primitives that underlie volatility clustering by examining two new representative agent models. Collectively, the two models are generalizations of the popular constant relative risk aversion (CRRA) and constant absolute risk aversion (CARA) "workhorse" models of finance.

The interest in studying volatility clustering in a representative agent setting stems from the observation that there should be an equilibrium connection between the agent's risk aversion level and the volatility level. For example, it is well known that the Black and Scholes (1973) model, with its assumption of constant volatility, does not perform well for most option markets. Evidently, a model that incorporates some form of time varying volatility into the underlying asset's price process is needed. However, as the equilibrium model of Bick (1987) demonstrates, if the volatility is time varying, then the economy's representative agent cannot possibly have constant relative risk aversion.[2] Apparently, if observed volatility patterns are to be reconciled within the representative agent framework, then the assumption of constant relative (or absolute) risk aversion must be abandoned in favor of more realistic assumptions. The two models presented in this article take a step in this direction.

Both models use a novel technique to exploit the economic relationship between the risk aversion level and the volatility level. Essentially, it is shown that a representative agent whose utility function exhibits several "risk aversion regimes" is tantamount to

[1] For the statistical properties of volatility, see Campbell et al. (1997) and the comprehensive survey of Palm (1996).

[2] The relationship between the Black-Scholes model and constant relative risk aversion is also discussed in various forms in Breeden and Litzenberger (1978), Brennan (1979), Rubinstein (1976) and Stapleton and Subrahmanyam (1984a,b).

having an equilibrium price process that exhibits several "volatility regimes". More specifically, in the first model, the agent's relative risk aversion (RRA) at the terminal date of the economy is assumed to be a step function of wealth (i.e., each step in the step function represents a different risk aversion regime). In the special case of no steps, the CRRA model is recovered. In the second model, the representative agent's absolute risk aversion (ARA) is assumed to be a step function of wealth. In the case of no steps, the CARA model is recovered. In both models, the representative agent's indirect utility function exhibits a time varying, step function pattern of risk aversion. In turn, this induces volatility clustering in the market portfolio return process, similar to what is observed in Fig. 1.

Due to the stepwise nature of the representative agent's risk aversion function, the problems studied in this article are free boundary problems. While free boundary problems are notoriously difficult to solve, the problems analyzed here have simple closed form solutions. In the case of the RRA step function model, the boundaries are all exponential functions of time and are all proportional to one another. In the case of the ARA step function model, the boundaries are all linear functions of time and are parallel to one another.

The boundaries play an important role in equilibrium. In the first model, the representative agent's RRA switches to a different level whenever the market portfolio price process crosses one of the boundaries. In order for the representative agent to continue to find it optimal to hold the market portfolio when the RRA changes, the volatility of the market portfolio in equilibrium also changes. The boundaries in the second model play a similar role except that now it is the agent's ARA that changes to a different level whenever the market portfolio price process crosses a boundary. Since both models allow for arbitrary step function patterns of risk aversion, myriad patterns of time varying volatility can be achieved by simply specifying an appropriate pattern of risk aversion for the representative agent. This fact makes both models quite useful for empirical analysis.

The two models are analyzed along several dimensions in order to determine their empirical relevance. The first set of empirical results is geared towards the regime estimation problem, i.e., determining the number of volatility regimes and the level of volatility within each regime that best fits the data. Using various model selection criteria, it is shown that the multi-regime models clearly outperform their CRRA and CARA counterparts. For the RRA step function model, five volatility regimes offer the best fit to the data. For the ARA step function model, four regimes offer the best fit. In addition, it is demonstrated that the fitted RRA regime-switching model is capable of explaining the GARCH effects in the data. While this seemingly lends some support to the theory, further analysis reveals that the fitted RRA regime-switching model is inconsistent with observed option market prices. This suggests that option prices do not fully reflect the regime-switching information that is present in the underlying asset market.

The second set of empirical results exploits the equilibrium relationship between the volatility level and the agent's risk aversion level by extracting the empirical risk aversion function from each model's volatility estimates. While estimating the empirical risk aversion function is not a new topic, there is a fundamental difference in the empirical methodology used here versus that of prior work. In particular, Aït-Sahalia and Lo (2000), Jackwerth (2000) and Rosenberg and Engle (2002) use S&P index options data to estimate the empirical risk aversion function while this article relies solely on stock market data. This is an important difference since it allows for both of the models to be estimated over a much longer sample period, including time periods for which there is no index options data.[3] Specifically, Aït-Sahalia and Lo (2000) use 1 year of options data (covering 1993), Jackwerth (2000) uses slightly less than 10 years of data (covering April 1986–December 1995) and Rosenberg and Engle (2002) use 5 years of data (covering 1991–1995). In contrast, the sample period used here covers the 28-year period from 1973 until 2000. Remarkably, even with the difference in methodology and the longer sample period, the risk aversion estimates presented here for 1993, 1986–1995 and 1991–1995 largely agree with those of Aït-Sahalia and Lo (2000), Jackwerth (2000) and Rosenberg and Engle (2002), respectively.

The final set of empirical results exploits the step function nature of the risk aversion functions in order to derive an economically powerful model specification test. In particular, it is well known that any continuous function (such as a continuous RRA function) can be approximated arbitrarily closely by using either piecewise linear approximating functions or piecewise constant approximating functions. The equilibria analyzed in this article are consistent with the latter approximating technique. In fact, since any step function pattern of risk aversion is allowed, the models presented here can be used to approximate any representative agent equilibrium that employs a state-independent utility function. Viewed in this manner, the two models encompass the entire class of "viable diffusion price processes" that are studied by Bick (1990) and He and Leland (1993). This fact greatly strengthens the empirical results since a rejection of the models can quite literally be interpreted as a rejection of the representative agent framework. In turn, this can be viewed as a rejection of the hypothesis that the market portfolio is described by a diffusion process in which the drift and volatility are at most functions of the market level and time.[4]

The model specification test is implemented by using size sorted equity portfolios and by applying Hansen's (1982) generalized method of moments (GMM) to the Euler equation pricing errors. Although several versions of the models cannot be rejected at traditional significance levels, the average pricing errors are inversely related to portfolio size. In fact, the average pricing error for the smallest stocks is roughly twice as large as the average pricing error of any other portfolio. Using Berk's (1995) critique of size-related anomalies, this fact suggests that both models are misspecified. Hence, even though the models are consistent with existing risk aversion estimates, the models' stochastic discount factors do not capture all of the priced risks in the economy.[5] Given the above discussion, this result clearly drives a stake in the heart of the representative agent framework (with state-independent utility).

[3] S&P 500 index options began trading on the CBOE in 1983 but did not have substantial trading volume until several years later.

[4] In addition, since the equilibria presented here are available in closed form, there is no need to solve He and Leland's (1993) nonlinear partial differential equation for the price coefficients. This fact may be of independent interest to researchers and finance practitioners.

[5] This result is in the spirit of Bollen et al. (2000) who take the level of investigation one step further by examining not only the ability of regime-switching models to capture the GARCH effects in the data, but also the ability of regime-switching models to provide accurate security valuation.

In terms of the existing literature, there are several competing explanations for time varying volatility. While some of these explanations do not explicitly address volatility clustering, they are nevertheless closely connected. The alternative explanations include the leverage effect of Black (1976) and Christie (1982), the volatility feedback hypothesis of Pindyck (1984) and French et al. (1987), the macroeconomic variables approach of Schwert (1989), French and Sichel (1993) and Hamilton and Lin (1996), and the contemporaneous return-volatility analysis of Duffee (1995). For additional analysis, see Poterba and Summers (1986), Campbell and Hentschel (1992), and the recent works of Bekaert and Wu (2000), Wu (2001), Gaunersdorfer and Hommes (2002) and Lux and Marchesi (2000). In contrast, the explanation put forth here relies solely on time varying risk aversion as the driver of time varying volatility. Hence, time varying volatility and volatility clustering are directly tied to the economic primitives.

The step function models also differ from existing volatility regime-switching models. In terms of the model structure, the transition probabilities in the step function models are endogenously determined and are state-dependent. This immediately sets the step function models apart from the model of Hamilton (1990, 1994), which employs constant transition probabilities (CTP). In terms of the economics, Ang and Bekaert (2002a) analyze the asset allocation problem in a partial equilibrium setting and show that a CRRA agent may alter his portfolio in response to a volatility regime shift. This is quite different from the equilibrium results presented here. In fact, as shown below, it is precisely because the representative agent does not have constant relative (or absolute) risk aversion that we get shifts in the volatility regime. Hence, this article offers an explanation for volatility regime shifting itself, while Ang and Bekaert (2002a) take regime shifting as given and analyze its economic implications.

The remainder of the article is organized as follows. Section 2 introduces the two models and summarizes the resulting equilibria. Section 3 briefly examines the implications for asset pricing, including the fact that the models produce time varying risky asset betas, an implied volatility skew and a term structure of implied volatility. Section 4 then presents the empirical results, followed by Section 5 that offers concluding remarks.

## 2. The economic models

The basic framework used in this article is similar to that of Bick (1990) and He and Leland (1993).[6] Each model is a continuous-time pure exchange economy with a representative agent. Time is indexed by $t \in [0,T]$ and at each date $t$ there are two assets

[6] The approach is similar but not identical. For example, He and Leland's (1993) regularity conditions are different than the ones used here. In particular, He and Leland assume that the drift and diffusion functions of the equilibrium price process are differentiable and that the agent's utility function is twice continuously differentiable. These assumptions do not apply to the models studied here. Due to the stepwise nature of the agent's risk aversion function, the agent's utility function does not have a continuous second derivative. In turn, this leads to an equilibrium price process for each model in which the diffusion function is not even once continuously differentiable.

available for trading, a riskless asset and a risky asset. The riskless asset is in zero net supply and pays a constant instantaneous rate of interest equal to $r$, whereas the risky asset has a positive supply of one unit and is interpreted to be the market portfolio. The representative agent is endowed at date 0 with one unit of the risky asset and none of the riskless asset. The agent's preferences are exogenously specified and the market portfolio price process is then endogenously determined such that the agent optimally holds the market portfolio and has a zero demand for the riskless asset. The market portfolio price process that satisfies this market clearing condition is called an equilibrium price process. Lastly, at date $T$, the economy ends and the representative agent consumes the final realized value of the market portfolio. There is no intermediate consumption.

Since the two models are new, it seems appropriate to discuss the assumed form of the representative agent's preferences. For clarity, this discussion proceeds along two dimensions: (i) the wealth-dependent nature of the agent's risk aversion level and (ii) the use of a step function to implement the wealth dependency. The need for (i) is well established in the literature. Indeed, the idea that an individual consumer's risk attitude changes with his wealth level goes back at least to Friedman and Savage (1948). Upon aggregating the individual consumers, it is apparent that the risk attitude of the economy's representative agent will also be wealth-dependent (Constantinides, 1982; Benninga and Mayshar, 2000). A wealth-dependent risk aversion level is also reflected in the experimental work of Levy (1994) and is consistent with the results of several recent empirical studies (Aït-Sahalia and Lo, 2000; Jackwerth, 2000; Mehra and Prescott, 1985; Rosenberg and Engle, 2002).

As for (ii), the existing literature provides no guidance as to whether or not the risk aversion level is a continuous function of wealth. While most researchers have chosen to work with a continuous function for reasons of tractability (He and Leland, 1993; Levy, 1994), there is some evidence to support the notion that a step function might be a better fit to the data. First, Lamoureux and Lastrapes (1990) have observed that the source of persistence in GARCH models might be "structural breaks" or "regime shifts" in the volatility process rather than long memory. A similar line of reasoning can be found in Diebold (1986), Hamilton and Susmel (1994), Perez-Quiros and Timmerman (2001) and Gray (1996). This observation is consistent with the theoretical models presented here since a risk aversion step function induces volatility regime-shifting in the equilibrium price process. Second, Bliss and Panigirtzoglou (2004) provide evidence that the implied risk aversion level is lower during periods of high volatility and is higher during periods of low volatility. They go on to suggest that theoretical models must evolve in order to capture this effect. The step function models presented here can be interpreted as one way of implementing their suggestion.

Third, Ang and Schwarz (1985) provide evidence of time varying risk aversion in which the presence of speculative traders leads to higher price volatility. If the speculative traders follow a technical trading rule based on the stock market level, then the aggregate risk aversion level will vary through time as the speculators periodically enter and exit the market. This type of behavior is consistent with a representative agent whose risk aversion level is described by a step function of wealth. Lastly, the step function approach is parsimonious, it is amenable to empirical analysis, and it reduces to the CARA and CRRA equilibrium models in the case of a single regime. It is also capable of closely approximating any continuous utility function, which implies that it can accommodate many of the features found in previous studies.

### 2.1. The RRA step function model

For the first model, the date $T$ utility function of the representative agent is described in terms of the agent's RRA function along with two continuity conditions. Specifically, it is assumed that the agent's RRA is a step function of wealth, the agent's utility function is continuous, and the agent's marginal utility is continuous. To illustrate, consider a collection of constants $\{\eta_n: n=1,2, \ldots ,N\}$ and let

$$0<\eta_1<\eta_2<\cdots<\eta_N \tag{1}$$

Each $\eta_n$ is a switching point, i.e., a point at which the RRA at date $T$ changes to a different level. For notational purposes, let $U(\cdot)$ denote the agent's direct utility function and let $W(T)$ denote the agent's wealth level at date $T$. Given the step function assumption, the agent's RRA takes the following form:

$$\frac{-W(T)U''(W(T))}{U'(W(T))} = \begin{cases} \gamma_1, & 0<W(T)\leq\eta_1 \\ \gamma_2, & \eta_1<W(T)\leq\eta_2 \\ \vdots & \vdots \\ \gamma_{N+1}, & \eta_N<W(T) \end{cases} \tag{2}$$

where $\{\gamma_n: n=1, \ldots ,N+1\}$ is a collection of positive constants. By integrating the step function in Eq. (2), the agent's utility function within the region $\eta_{n-1}<W(T)\leq\eta_n$ takes the form

$$U(W(T)) = k_n\frac{(W(T))^{1-\gamma_n}}{1-\gamma_n} + d_n \tag{3}$$

where $k_n$ and $d_n$ are the constants of integration.[7] To satisfy the continuity conditions, the constants $\{k_n: n=1, \ldots ,N+1\}$ and $\{d_n: n=1, \ldots ,N+1\}$ are chosen such that $U(\cdot)$ and $U'(\cdot)$ are continuous functions of $W(T)$. Since there are $N$ switching points and both $U(\cdot)$ and $U'(\cdot)$ are continuous at each switching point, there are a total of $2N$ equations. However, since there are $N+1$ regions with two constants of integration for each region, there are a total of $2N+2$ constants. It is therefore assumed that $k_1=1$ and $d_1=0$. This arbitrary choice of $k_1$ and $d_1$ is without loss of generality and has no effect on the results of the article. The remaining constants are now determined recursively via the expressions

$$k_n = k_{n-1}(\eta_{n-1})^{\gamma_n-\gamma_{n-1}} \tag{4}$$

$$d_n = d_{n-1} + \frac{k_{n-1}}{1-\gamma_{n-1}}(\eta_{n-1})^{1-\gamma_{n-1}} - \frac{k_n}{1-\gamma_n}(\eta_{n-1})^{1-\gamma_n} \tag{5}$$

for $n=2,3,\ldots,N+1$.

[7] If $\gamma_n=1$, then preferences are logarithmic for the region $\eta_{n-1}<W(T)\leq\eta_n$.

Since there are $N$ steps in the agent's RRA function, it is initially conjectured that the market portfolio price process has $N$+1 different volatility regimes. Proposition 1 below formally verifies that the conjectured price process coincides with the endogenously determined equilibrium price process. Letting $S(t)$ denote the price of the market portfolio at date $t$, the market portfolio price process is

$$\mathrm{d}S(t) = (r+\lambda)S(t)\mathrm{d}t + \sigma(t)S(t)\mathrm{d}z(t) \tag{6}$$

where $z(t)$ is a standard Brownian motion and $\lambda>0$ is the constant instantaneous risk premium.[8] Due to the RRA step function, the function $\sigma(t)$ is itself a step function. To describe $\sigma(t)$, first introduce the functions

$$b_n(t) = \eta_n \mathrm{e}^{-r(T-t)-\frac{1}{2}\lambda(T-t)} \tag{7}$$

for $n$=1,2, … ,$N$. Each $b_n(t)$ represents the switching boundary at date $t$ that is associated with the switching point $\eta_n$. The function $\sigma(t)$ can now be written as

$$\sigma(t) = \begin{cases} \sqrt{\frac{\lambda}{\gamma_1}}, & 0<S(t)\leq b_1(t) \\ \sqrt{\frac{\lambda}{\gamma_2}}, & b_1(t)<S(t)\leq b_2(t) \\ \vdots & \vdots \\ \sqrt{\frac{\lambda}{\gamma_{N+1}}}, & b_N(t)<S(t) \end{cases} \tag{8}$$

As illustrated in Eq. (8), the volatility of the market portfolio within a region depends on the level of RRA within that region. Hence, the market portfolio has the highest instantaneous volatility in the region in which the agent has the lowest RRA, has the next highest instantaneous volatility in the region in which the agent has the next lowest RRA, and so forth.

Turning next to the agent's optimization problem, it is verified that Eqs. (6)–(8) is an equilibrium price process. Letting $\theta(t)$ denote the representative agent's fraction of date $t$ wealth allocated to the market portfolio, the agent solves the following dynamic investment problem:

$$\max_{\theta(t)} E[U(W(T))] \tag{9}$$

subject to the constraints

$$\theta(0) = 1 \tag{10}$$

and

$$\mathrm{d}W(t) = (r+\theta(t)\lambda)W(t)\mathrm{d}t + \theta(t)\sigma(t)W(t)\mathrm{d}z(t) \tag{11}$$

[8] Nakao (1972) proves that stochastic differential equations of the type given in Eq. (6) have pathwise unique solutions. In addition, a constant risk premium $\lambda$ is for ease of exposition only. The model can easily be generalized to allow $\lambda=\lambda(t)$, a deterministic function of time.

In words, the agent dynamically allocates his wealth between the market portfolio and the riskless bond in order to maximize the expected utility from consuming the final wealth at date $T$. The first constraint (i.e., expression (10)) is simply the endowment of the agent. This constraint implies $W(0)=S(0)$. The second constraint is the dynamic budget constraint that characterizes the evolution of the agent's wealth through time. The following proposition summarizes the resulting equilibrium.

**Proposition 1.** *The market portfolio price process in Eqs. (6)–(8) is an equilibrium price process. The representative agent's optimal policy is $\theta(t)=1$ for all $t\in[0,T]$.*

**Proof.** In Appendix A. □

Since the volatility of the market portfolio is time varying as described by Eq. (8), the equilibrium Sharpe ratio is also time varying. For the region $b_{n-1}(t)<S(t)\leq b_n(t)$, the Sharpe ratio is given by $\sqrt{\lambda\gamma_n}$ so that the equity risk premium per unit of volatility is proportional to the square root of the RRA. In addition, for the region $b_{n-1}(t)<W(t)\leq b_n(t)$, the agent's indirect utility function is

$$J(W(t),t) = k_n \frac{(W(t))^{1-\gamma_n}}{1-\gamma_n}(g(t))^{1-\gamma_n} + d_n \tag{12}$$

where $g(t) = \mathrm{e}^{\left(r+\frac{\lambda}{2}\right)(T-t)}$. Expression (12) implies a time varying, step function pattern of RRA.

One of the interesting features of the equilibrium is that the risk premium $\lambda$ does not vary with $S(t)$. To understand this, note that the equilibrium price process under CRRA preferences is a geometric Brownian motion (Bick, 1987; He and Leland, 1993). In addition, if the CRRA agent's risk aversion parameter is $\gamma$, the coefficients of the geometric Brownian motion satisfy $\lambda=\gamma\sigma^2$. Solving for $\sigma$ gives $\sigma = \sqrt{\lambda/\gamma}$. Hence, ex ante, it seems reasonable to expect that time varying volatility can be achieved by allowing both the risk premium $\lambda$ and the RRA parameter $\gamma$ to vary with $S(t)$. However, as Proposition 1 demonstrates, if the RRA is described by a step function, $\lambda$ cannot vary with $S(t)$. The main reason for this is that the endogenous switching boundaries in Eq. (7) depend on $\lambda$ and these boundaries, by definition, can at most be a function of $t$.

Fig. 2 illustrates the model for the case of three different RRA levels. The upper left panel shows the agent's RRA function, while the lower left panel shows the two switching boundaries and a sample path of the equilibrium price process. Using the same sample path, the upper right panel shows the instantaneous volatility while the lower right panel shows the market's raw return over each time step. As the figure illustrates, the sample path starts in the regime with the highest volatility and then enters the regime with the lowest volatility. At $t$=8.05, there is brief excursion back into the regime with the highest volatility. Relative to the low volatility regime, this brief excursion shows up as an outlier in the sample path returns. The lower right panel clearly shows that the model is consistent with volatility clustering.

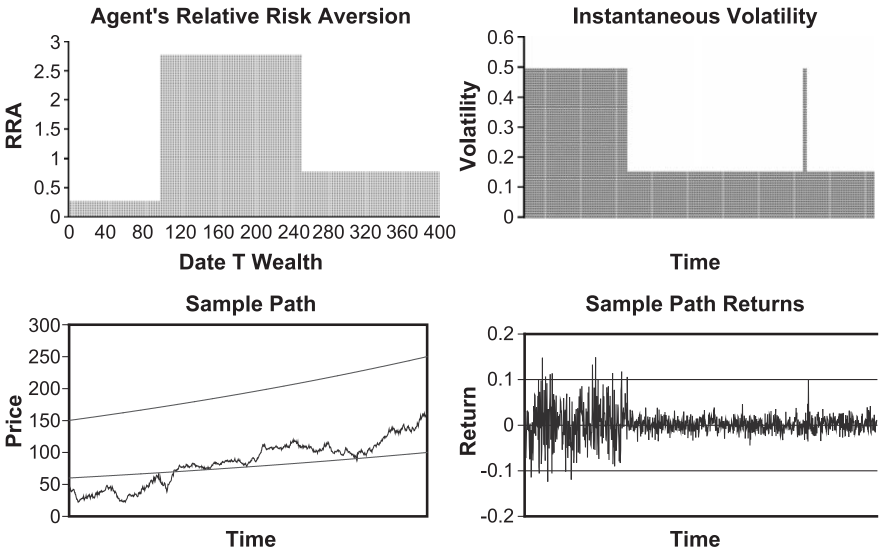


Fig. 2. The upper left panel shows the agent's relative risk aversion using $\eta_1$=100, $\eta_2$=250, $\gamma_1$=0.25, $\gamma_2$=2.75 and $\gamma_3$=0.75. Using these same parameter values, the lower left panel shows the switching boundaries and a sample path of the equilibrium price process. The two right panels show the instantaneous volatility and the return over each time step, respectively, that correspond to this sample path. The other inputs used are $\lambda$=0.06, $r$=0.05, $T$=10 and $S(0)$=45.

## 2.2. The ARA step function model

In the second model, the following three conditions describe the agent's date $T$ utility function: (i) the agent's ARA is a step function of wealth; (ii) the agent's utility function is continuous; and (iii) the agent's marginal utility is continuous. To illustrate, again consider a collection of constants $\{\eta_n: n=1,2, \ldots ,N\}$ and let

$$\eta_1<\eta_2<\cdots<\eta_N \tag{13}$$

Each $\eta_n$ is a switching point, i.e., a point at which the agent's ARA changes to a different level. Note that unlike the first model, the constants $\{\eta_n\}$ are not restricted to be positive numbers. Due to condition (i), the representative agent's ARA takes the following form:

$$\frac{-U''(W(T))}{U'(W(T))}=\begin{cases}\gamma_1, & W(T)\le\eta_1\\ \gamma_2, & \eta_1<W(T)\le\eta_2\\ \vdots & \vdots\\ \gamma_{N+1}, & \eta_N<W(T)\end{cases} \tag{14}$$

where $\{\gamma_n: n=1, \ldots ,N+1\}$ is a collection of positive constants. Upon integrating the ARA step function, the agent's utility function within the region $\eta_{n-1}<W(T)\le\eta_n$ is

$$U(W(T))=-\frac{k_n}{\gamma_n}\mathrm{e}^{-\gamma_n W(T)}+d_n \tag{15}$$

where $k_n$ and $d_n$ are the constants of integration. To satisfy (ii) and (iii), the constants $\{k_n: n=1, \dots, N+1\}$ and $\{d_n: n=1, \dots, N+1\}$ are chosen such that $U(\cdot)$ and $U'(\cdot)$ are continuous functions of $W(T)$. Choosing $k_1=1$ and $d_1=0$, the remaining constants are determined recursively via the expressions

$$k_n = k_{n-1}\exp[(\gamma_n - \gamma_{n-1})\eta_{n-1}] \tag{16}$$

$$d_n = d_{n-1} + \frac{k_n}{\gamma_n}\mathrm{e}^{-\gamma_n \eta_{n-1}} - \frac{k_{n-1}}{\gamma_{n-1}}\mathrm{e}^{-\gamma_{n-1}\eta_{n-1}} \tag{17}$$

for $n=2,3, \dots, N+1$.

Taking the riskless bond as numeraire (i.e., setting the interest rate equal to zero) and fixing the risk premium $\lambda>0$, the market portfolio price process is conjectured to be

$$\mathrm{d}S(t) = \lambda \mathrm{d}t + \sigma(t)\mathrm{d}z(t) \tag{18}$$

which is formally verified in Proposition 2 below. In Eq. (18), the function $\sigma(t)$ is given by

$$\sigma(t) = \begin{cases} \sqrt{\frac{\lambda}{\gamma_1}}, & S(t)\le b_1(t) \\ \sqrt{\frac{\lambda}{\gamma_2}}, & b_1(t)<S(t)\le b_2(t) \\ \vdots & \vdots \\ \sqrt{\frac{\lambda}{\gamma_{N+1}}}, & b_N(t)<S(t) \end{cases} \tag{19}$$

where the switching boundaries are

$$b_n(t) = \eta_n - \frac{1}{2}\lambda(T-t) \tag{20}$$

for $n=1,2, \dots, N$. Unlike the first model, the switching boundaries in this model are linear functions of time. Given the utility function and the price process above, the following proposition summarizes the equilibrium.

**Proposition 2.** *The market portfolio price process in Eqs. (18)–(20) is an equilibrium price process. The representative agent's optimal policy is $\theta(t)=1$ for all $t\in[0,T]$.*

**Proof.** In Appendix A. □

For the region $b_{n-1}(t)<W(t)\le b_n(t)$, the agent's indirect utility function is

$$J(W(t),t) = -\frac{k_n}{\gamma_n}\mathrm{e}^{-\gamma_n W(t)}(g(t))^{\gamma_n} + d_n \tag{21}$$

where $g(t) = \mathrm{e}^{-\frac{\lambda}{2}(T-t)}$. Expression (21) implies a time varying, step function pattern of ARA. In turn, due to the market clearing condition, this time varying pattern of ARA induces a time varying pattern of volatility clustering, as described by Eq. (19).

Fig. 3 illustrates the ARA step function model for the case of three different levels of ARA. For the particular sample path shown in the figure, the price process has excursions in all three of the volatility regimes. This is clearly seen in the upper right panel that shows the volatility level at each point in time. As the lower right panel shows, the magnitude of

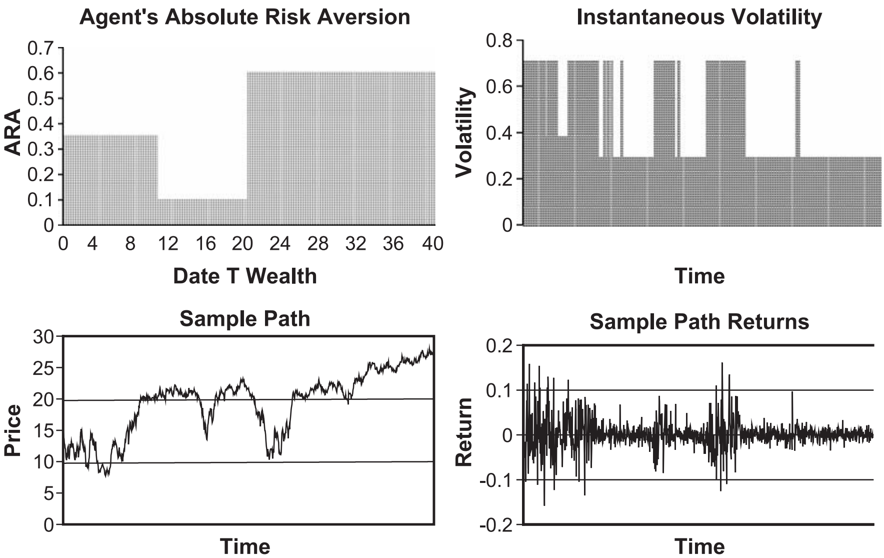


Fig. 3. The upper left panel shows the agent's absolute risk aversion using $\eta_1$=10, $\eta_2$=20, $\gamma_1$=0.35, $\gamma_2$=0.10 and $\gamma_3$=0.60. Using these same parameter values, the lower left panel shows the switching boundaries and a sample path of the equilibrium price process. The two right panels show the instantaneous volatility and the return over each time step, respectively, that correspond to this sample path. The other inputs used are $\lambda$=0.05, $T$=10 and $S(0)$=15.

the returns varies substantially over time and is tied to the contemporaneous volatility level. This illustrates the volatility clustering effect that is induced by the time varying risk aversion.

## 3. Implications for asset pricing

The implications of the step function models for asset pricing are threefold. First, it is demonstrated that volatility clustering of the market portfolio induces time varying betas for all risky assets. Second, since the market volatility depends on preferences, the prices of options are also explicitly preference-dependent. Lastly, the prices of options exhibit a term structure of implied volatility and an implied volatility skew. In the interest of brevity, only the RRA step function model is formally discussed with the understanding that the results also apply to the ARA step function model.

### *3.1. Time varying betas*

In the case of a representative agent with a RRA step function, recall from Eq. (6) that the market portfolio price process is

$$\mathrm{d}S(t) = (r + \lambda)S(t)\mathrm{d}t + \sigma(t)S(t)\mathrm{d}z(t) \tag{22}$$

Now suppose that there exists an individual risky asset with price process

$$\mathrm{d}y(t) = \alpha y(t)\mathrm{d}t + vy(t)\mathrm{d}w(t) \tag{23}$$

where $w(t)$ is a standard Brownian motion that has correlation $\rho$ with $z(t)$. Since $y(t)$ is the price of the asset, it must be the case that

$$y(t) = E_t\left[\frac{\xi(T)}{\xi(t)}y(T)\right] \tag{24}$$

where $\xi(t)$ is the pricing kernel for the economy (see expression (48) in Appendix A). Since Eq. (24) must hold at each date $t$, this places a restriction on the drift $\alpha$ of the risky asset. Indeed, it must be the case that

$$\alpha = r + \left[\frac{\rho v}{\sigma(t)}\right]\lambda \tag{25}$$

The expression in square brackets is the instantaneous beta of the asset. The fact that this term involves $\sigma(t)$ shows that time varying risk aversion induces time varying betas (and hence time varying excess returns). For the case of constant $v$ and $\rho$, the beta and the excess return are high when risk aversion is high, and vice versa. In addition, there are as many different betas for the risky asset as there are volatility regimes for the market portfolio.

Since $\lambda$ is a constant, the security market line has a constant slope in the regime-switching model. However, since risky assets have time varying betas, the location of where an individual asset plots along the security market line changes over time. Empirical tests that ignore the possibility of a regime change are likely misspecified, and the misspecification will be particularly acute if the beta measurement period and the return sample period involve different regimes.

### 3.2. Derivatives pricing and implied volatility

Using the RRA step function model, the pricing problem for a standard European call option on the market portfolio is examined. For illustration, consider an option with strike price $L$ and time to expiration $T-t$. Furthermore, suppose that there is a single switching point $\eta$ and two risk aversion levels, $\gamma_1$ and $\gamma_2$. Since $S(t)$ may be greater than or less than the switching boundary $b(t)$, the call option pricing formula satisfies a pair of partial differential equations subject to appropriate boundary conditions. If $S(t)>b(t)$, the value of the call option, $C(S(t),t)$, solves

$$\frac{1}{2}\frac{\lambda}{\gamma_2}S^2C_{SS} + rSC_S + C_t - rC = 0 \tag{26}$$

where subscripts on $C$ denote partial differentiation. On the other hand, if $S(t)\leq b(t)$, the value of the call option solves

$$\frac{1}{2}\frac{\lambda}{\gamma_1}S^2C_{SS} + rSC_S + C_t - rC = 0 \tag{27}$$

In addition, $C$ and $C_S$ are continuous at $b(t)$ and $C$ satisfies the condition

$$C(S(T), T) = \max[S(T) - L, 0] \tag{28}$$

While solving Eqs. (26)–(28) analytically is a formidable task, it is straightforward to approach this problem numerically.

Since Eqs. (26)–(27) involve $\gamma_1$ and $\gamma_2$, the call option value explicitly depends on the preferences of the representative agent. In addition, the option value depends on the equity risk premium $\lambda$ and on the switching point $\eta$, which show up in the boundary $b(t)$. One way to characterize the preference-dependence of option prices is to examine how the risk aversion parameters affect the implied volatility. For example, if the risk aversion level is lower below the boundary and higher above the boundary, the equilibrium price process will have higher volatility below $b(t)$ and lower volatility above $b(t)$. This leads to a downward sloping volatility skew as shown in the top left panel of Fig. 4. On the other hand, if the risk aversion levels are reversed, the volatility skew is upward sloping as shown in the top right panel of Fig. 4. As $T-t$ increases, the exponential switching boundary converges to zero and hence the equilibrium price process at date $t$ behaves more like a geometric Brownian motion. Because of this, longer-dated options generate a relatively "flat" volatility skew while shorter-dated options generate a "steeper" volatility skew. This is illustrated by comparing the solid and dotted lines in the two top panels of the figure, which correspond to 3 and 12 months until expiration, respectively.

If the risk aversion level is lower below the boundary and higher above the boundary, the exponential form of the switching boundary implies that the term structure of implied

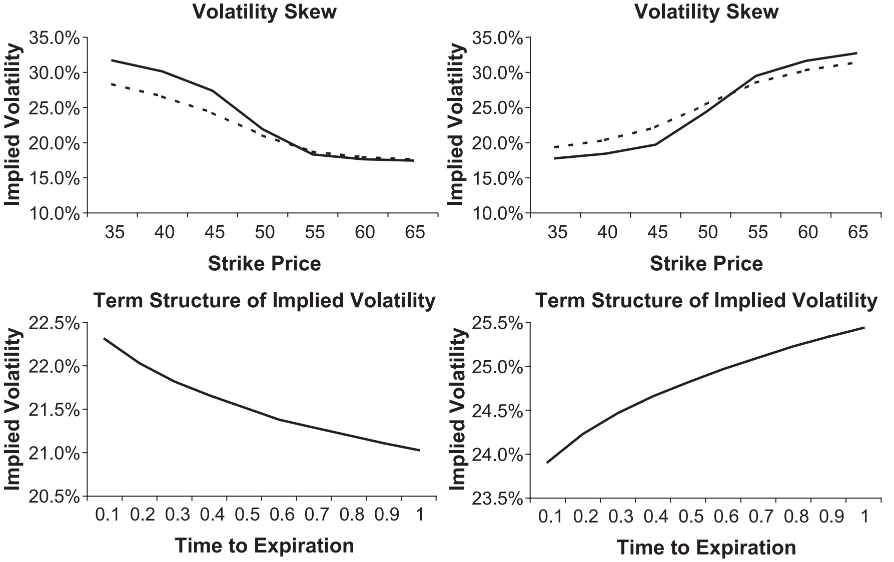


Fig. 4. The two top panels show the implied volatility skew for terms to expiration of 3 months (solid line) and 12 months (dotted line). The two bottom panels show the term structure of implied volatility for the at-the-money option. The two left panels use $\gamma_1$=0.5, $\gamma_2$=2.0, while the two right panels use $\gamma_1$=2.0, $\gamma_2$=0.5. The other parameter values are $\lambda$=0.06, $r$=0.035, $\eta$=50 and $S$=50.

volatility is downward sloping as a function of the time until expiration. If the risk aversion levels are reversed, the term structure is upward sloping. These features are illustrated in the two bottom panels of the figure. Lastly, if the risk aversion is lower below the boundary and higher above the boundary, the overall level of the term structure is lower. These issues will be revisited in Section 4.4, which assesses the step function model's ability to empirically match the options data.

## 4. Empirical results

The RRA and ARA step function models are empirically estimated in order to evaluate the relevance of the theory. The first set of results uses maximum likelihood to estimate the model parameters and their standard errors. Various model selection criteria are then used to identify the number of regimes that offers the best fit to the data. For comparison, a regime-switching model of the type discussed in Hamilton (1990, 1994) is also estimated. The main difference between the step function models and Hamilton's (1990, 1994) model is that the former has time varying transition probabilities whereas the latter has CTP.[9] The maximum likelihood estimation errors are then used to determine whether or not the models capture the GARCH effects in the data. As shown below, the RRA models tend to outperform their ARA counterparts and, within the RRA class of models, the step function model with time varying transition probabilities offers the best fit to the data.

### *4.1. Regime estimation*

The maximum likelihood estimation is carried out using monthly S&P 500 total returns over the sample period January 1973–December 2000. The S&P data for the analysis is obtained from the CRSP database and monthly T-bill total returns are used to determine the riskless rate $r$. The main reason for focusing on S&P 500 data in this section is to be able to provide a comparison to the other literature that analyzes empirical risk aversion functions (e.g., Aït-Sahalia and Lo, 2000; Jackwerth, 2000; Rosenberg and Engle, 2002 use the S&P 500 as the market proxy). Since the theoretical model does not account for intermediate consumption, it is assumed that all dividends and capital gains are reinvested, i.e., $S(t)$ is taken to be the total return process for the S&P 500. For reference, with dividends and capital gains reinvested, a US$1 investment in the S&P 500 at the close of trading on December 31, 1972 is worth US$31.43 on December 31, 2000.

Before proceeding with the estimation, there is one issue worth discussing. Since the continuous price processes in Eqs. (6)–(8) and (18)–(20) do not have known transition densities, a suitably discretized version of each price process is used to derive the

[9] The CTP models are included as a benchmark against which the equilibrium step function models can easily be compared. Several studies, including Gray (1996) and Diebold et al. (1994), have concluded that regime-switching models can be significantly improved by allowing the transition probabilities to be state dependent (and therefore time varying). Along these lines, Perez-Quiros and Timmermann (2001) refer to CTP models as "first generation" models and regime-switching models with time varying transition probabilities as "second generation" models.

likelihood function. A similar technique is used by Christie (1982) and Ogden (1987). While there are several econometric issues associated with using a discretized model (Lo, 1988), there are also issues associated with using other, more involved techniques. For example, since the volatility functions in the RRA and ARA step function models are not differentiable, the discretely sampled diffusion techniques of Lo (1988) and Aït-Sahalia (2002) do not directly apply to the problem at hand. In addition, at a more basic economic level, the continuous processes in Eqs. (6)–(8) and (18)–(20) are to some extent misspecified since financial markets are not continuously open for trading. Hence, while the indirect inference technique of Gourieroux et al. (1993) could be applied to the regime estimation problem, Genton and Ronchetti (2003) show that, in the presence of model misspecification, the indirect estimators tend to lose their ability to correct for the discretization (and therefore lose their justification).[10] With these issues in mind, maximum likelihood estimation is simply applied to the discretized model.

The setup of the RRA model is described first and then several points are discussed about the ARA setup. The likelihood function is derived using a procedure that is based on Hamilton (1990, 1994). For an application of Hamilton's two-regime model to stock market data, see Pagan and Schwert (1990). For our more general multi-regime problem, let $\theta$ denote the vector of parameters to be estimated, i.e., the $N$ switching points $\{\eta_n\}_{n=1,N}$ and the $N$+1 volatility levels $\{\sigma_n\}_{n=1,N+1}$. As a first step, apply the transformation

$$x(t) = S(t)\mathrm{e}^{r(T-t)+\frac{\lambda}{2}(T-t)} \tag{29}$$

to the equilibrium price process in Eqs. (6)–(8) and use Itô's lemma to get

$$\mathrm{d}\log x(t) = \frac{\lambda}{2}\mathrm{d}t - \frac{1}{2}\sigma^2(t)\mathrm{d}t + \sigma(t)\mathrm{d}z(t) \tag{30}$$

where $\sigma(t)$ equals $\sigma_1 \equiv \sqrt{\lambda\gamma_1^{-1}}$ if $x(t) \le \eta_1$, equals $\sigma_2 \equiv \sqrt{\lambda\gamma_2^{-1}}$ if $\eta_1 < x(t) \le \eta_2$ and so forth. Hence, under the transformation given in Eq. (29), the boundaries are no longer time-dependent.

Next, discretize the interval $[0,T]$ into time steps of equal size $\delta t$ so that the time index $t \in \{0, \delta t, 2\delta t, \ldots, T\}$. Based on Eq. (30), the discrete time approximation

$$\log x_{t+\delta t} = \log x_t + \frac{\lambda}{2}\delta t - \frac{1}{2}\sigma_t^2\delta t + \sigma_t Z_{t+\delta t} \tag{31}$$

is used for estimation purposes where $Z_{t+\delta t}$ is normally distributed with mean zero and variance $\delta t$. To implement Eq. (31), $\lambda$ is replaced with its sample mean. In particular, since the average excess total return on the S&P 500 during the sample period is 5.68%, it is assumed that $\lambda$=0.0568.

Given this setup, if it is known that $\eta_{n-1} < x_t \le \eta_n$, then log $x_{t+\delta t}$ is normally distributed with mean

$$\log x_t + \frac{\lambda}{2}\delta t - \frac{1}{2}\sigma_n^2\delta t \tag{32}$$

[10] The indirect inference technique contains the simulated method of moments (SMM) of Duffie and Singleton (1993) as a special case.

and variance $\sigma_n^2\delta t$. Furthermore, given a known parameter vector $\theta$, the probability of transitioning from the regime with volatility $\sigma_n$ to the regime with volatility $\sigma_k$ over the time interval $\delta t$ is

$$P(\sigma_{t+\delta t} = \sigma_k | \sigma_t = \sigma_n) = \Phi\left(\frac{\log(\eta_k) - (\log x_t + \frac{\lambda}{2}\delta t - \frac{1}{2}\sigma_n^2\delta t)}{\sigma_n\sqrt{\delta t}}\right) - \Phi\left(\frac{\log(\eta_{k-1}) - (\log x_t + \frac{\lambda}{2}\delta t - \frac{1}{2}\sigma_n^2\delta t)}{\sigma_n\sqrt{\delta t}}\right) \tag{33}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

Since the true $\theta$ is unknown, an inference is made using the data. For notational purposes, let $\mathcal{Y}_t$ denote the observations of $x_t$ up through date $t$ and let $\phi(z;\mu,v^2)$ denote the normal density function with mean $\mu$ and variance $v^2$. At date 0, it is assumed that each regime is equally likely, i.e., $P(\sigma_0=\sigma_n)=(1+N)^{-1}$. Upon observing $x_{\delta t}$, the belief about the initial regime is updated and the optimal inference is

$$P(\sigma_0 = \sigma_n | \mathcal{Y}_{\delta t}) = \frac{\phi\left(\log x_{\delta t}; \log x_0 + \frac{\lambda}{2}\delta t - \frac{1}{2}\sigma_n^2\delta t, \sigma_n^2\delta t\right)}{\sum_{k=1}^{N+1}\phi\left(\log x_{\delta t}; \log x_0 + \frac{\lambda}{2}\delta t - \frac{1}{2}\sigma_k^2\delta t, \sigma_k^2\delta t\right)} \tag{34}$$

Using these probabilities and the transition probabilities given by Eq. (33), the optimal inference about the regime at date $\delta t$ is

$$P(\sigma_{\delta t} = \sigma_n | \mathcal{Y}_{\delta t}) = \sum_{k=1}^{N+1} P(\sigma_{\delta t} = \sigma_n | \sigma_0 = \sigma_k) P(\sigma_0 = \sigma_k | \mathcal{Y}_{\delta t}) \tag{35}$$

Next suppose that $x_{2\delta t}$ is observed. The optimal inference for the regime at date $\delta t$ conditional on $\mathcal{Y}_{2\delta t}$, denoted $P(\sigma_{\delta t} = \sigma_n | \mathcal{Y}_{2\delta t})$, is equal to

$$\frac{\phi\left(\log x_{2\delta t}; \log x_{\delta t} + \frac{\lambda}{2}\delta t - \frac{1}{2}\sigma_n^2\delta t, \sigma_n^2\delta t\right) P(\sigma_{\delta t} = \sigma_n | \mathcal{Y}_{\delta t})}{\sum_{k=1}^{N+1}\phi\left(\log x_{2\delta t}; \log x_{\delta t} + \frac{\lambda}{2}\delta t - \frac{1}{2}\sigma_k^2\delta t, \sigma_k^2\delta t\right) P(\sigma_{\delta t} = \sigma_k | \mathcal{Y}_{\delta t})} \tag{36}$$

while the optimal inference for the regime at date $2\delta t$ conditional on $\mathcal{Y}_{2\delta t}$ is

$$P(\sigma_{2\delta t} = \sigma_n | \mathcal{Y}_{2\delta t}) = \sum_{k=1}^{N+1} P(\sigma_{2\delta t} = \sigma_n | \sigma_{\delta t} = \sigma_k) P(\sigma_{\delta t} = \sigma_k | \mathcal{Y}_{2\delta t}) \tag{37}$$

Proceeding recursively, expressions similar to Eqs. (36)–(37) are constructed for each date. Since the denominator of Eq. (36), denoted $f(x_{2\delta t} | \mathcal{Y}_{\delta t}, \theta)$, is the conditional density of log $x_{2\delta t}$ given $\mathcal{Y}_{\delta t}$, an estimate of the parameter vector $\theta$ is obtained by solving the optimization problem

$$\max_\theta L_{N+1}(\theta) \equiv \sum_{t \in \{\delta t, 2\delta t, \ldots, T\}} \log[f(x_t | \mathcal{Y}_{t-\delta t}, \theta)] \tag{38}$$

subject to the constraints $\sigma_n>0$ for $n=1,2,\ldots,N+1$ and $0<\eta_1<\eta_2<\cdots<\eta_N$.

For the ARA step function model, the transformation

$$x(t) = S(t) + \frac{\lambda}{2}(T - t) \tag{39}$$

is used to obtain the discretized process

$$x_{t+\delta t} = x_t + \frac{\lambda}{2}\delta t + \sigma_t Z_{t+\delta t} \tag{40}$$

where $\sigma_t$ equals $\sigma_1$ if $x(t) \leq \eta_1$, equals $\sigma_2$ if $\eta_1 < x(t) \leq \eta_2$, etc. Since the ARA model takes the bond as numeraire, a normalized price system with $r=0$ is used to construct the total return process $S(t)$. In addition, $\lambda$ is replaced with the average monthly change in $S(t)$ using the normalized price system. For the sample period used here, $\lambda=0.1397$. The remainder of the setup follows that of the RRA model, i.e., expressions similar to Eqs. (33)–(37) are derived to get the likelihood function for the ARA model. A problem similar to Eq. (38) is then solved for the parameter estimates.

For comparison, two exogenous regime-switching models with CTP are also estimated. The first CTP model is identical to the RRA model described above with the exception that the transition probabilities in Eq. (33) are replaced by a constant Markov matrix of probabilities as in Hamilton (1990, 1994). The second CTP model is identical to the ARA model with the exception that it too uses a constant Markov matrix of transition probabilities. For future reference, note that the number of parameters in the step function model is linear in the number of regimes. For a given value of $N$, one must estimate $N+1$ volatility levels and $N$ switching points. This produces a total number of parameters equal to $2(N+1)-1$. On the other hand, for the CTP model, one must estimate $N+1$ volatility levels and a transition probability matrix whose dimension is $(N+1)\times(N+1)$. Since the rows of the matrix must sum to 1, this implies that the total number of estimated parameters for the CTP model is equal to $(N+1)^2$, which is a quadratic function of the number of regimes. Since the model selection is at least partially based on the information criteria of Akaike (1974, 1981) and Schwarz (1978), this suggests that a smaller number of regimes will likely be chosen for the CTP models.

Table 1 summarizes the estimated models and presents several model selection criteria. The Akaike Information Criterion (AIC) suggests that five regimes offer the best fit for the RRA step function model, while four regimes offer the best fit for the ARA step function model. For both of the CTP models, the AIC suggests that three regimes offer the best fit. As shown in the last column of Table 1, with the exception of the ARA step function model, the Schwarz (1978) criterion selects a model with a fewer number of regimes relative to the AIC. This is not too surprising given the larger penalty of the Schwarz criterion relative to that of the AIC. Since the AIC and Schwarz criterion disagree and since the different parameter growth rates naturally lead to a smaller number of regimes for the CTP models, a formal likelihood ratio test is used to help select the number of regimes for each type of model. For the step function models, the $n$-regime (restricted) model is tested against the $(n+1)$-regime (unrestricted) model by imposing the two restrictions $\eta_n=\eta_{n+1}$ and $\sigma_n=\sigma_{n+1}$ for some $n$. The test statistic in this case is asymptotically distributed $\chi^2(2)$, i.e., chi-squared with two degrees of freedom. For the CTP models, the standard likelihood ratio test is technically invalid since not all of the parameters are

Table 1
Model summary and selection criteria

| Model | Number of parameters | Maximum log-likelihood | Akaike criterion (AIC) | Schwarz criterion |
|---|---|---|---|---|
| RRA 1-regime | 1 | 551.45 | 550.45 | 548.54 |
| RRA 2-regime | 3 | 562.73 (0.0000) | 559.73 | 554.00 |
| RRA 3-regime | 5 | 575.46 (0.0000) | 570.46 | 560.92 |
| RRA 4-regime | 7 | 585.81 (0.0000) | 578.81 | 565.45 |
| RRA 5-regime | 9 | 590.19 (0.0124) | 581.19 | 564.01 |
| RRA 6-regime | 11 | 590.47 (0.7605) | 579.47 | 558.48 |
| ARA 1-regime | 1 | 294.38 | 293.38 | 291.47 |
| ARA 2-regime | 3 | 434.14 (0.0000) | 431.14 | 425.41 |
| ARA 3-regime | 5 | 447.24 (0.0000) | 442.24 | 432.70 |
| ARA 4-regime | 7 | 457.27 (0.0000) | 450.27 | 436.91 |
| ARA 5-regime | 9 | 457.61 (0.7104) | 448.61 | 431.43 |
| RRA CTP 1-regime | 1 | 551.45 | 550.45 | 548.54 |
| RRA CTP 2-regime | 4 | 572.76 (0.0000) | 568.76 | 561.13 |
| RRA CTP 3-regime | 9 | 579.29 (0.0228) | 570.29 | 553.12 |
| RRA CTP 4-regime | 16 | 582.85 (0.4179) | 566.85 | 536.31 |
| ARA CTP 1-regime | 1 | 294.38 | 293.38 | 291.47 |
| ARA CTP 2-regime | 4 | 446.49 (0.0000) | 442.49 | 434.86 |
| ARA CTP 3-regime | 9 | 460.06 (0.0001) | 451.06 | 433.89 |
| ARA CTP 4-regime | 16 | 462.81 (0.6002) | 446.81 | 416.27 |

Summary selection criteria are presented for the regime-switching models and their CTP counterparts. The number in parentheses is the $p$-value for a likelihood ratio test of that model against the preceding model (the null). It is assumed that twice the difference in the maximum log-likelihoods is distributed $\chi^2$ with degrees of freedom equal to the difference in the number of parameters between the alternative and the null. The AIC is calculated as the maximum log-likelihood less $n$, where $n$ is the number of parameters. The Schwarz criterion is equal to the maximum log-likelihood less $(n/2)\ln(T)$, where $T$=336 is the sample size.

identified under the restricted model. Table 1 nevertheless reports the test results based on the assumption that the test statistic is asymptotically distributed as a $\chi^2$ random variable with degrees of freedom equal to the difference in the number of parameters between the unrestricted and restricted models.[11]

The middle column of Table 1 gives the maximum value of the log-likelihood function for each of the estimated models. The number in parentheses that is next to the maximum log-likelihood is the $p$-value for a likelihood ratio test of that model against the immediately preceding model. For the RRA step function model, the restricted two-regime, three-regime and four-regime models are easily rejected in favor of their unrestricted counterparts (all three $p$-values are very close to zero). The evidence is less strong for the test of the restricted five-regime model (a $p$-value of 0.0124) and the data clearly cannot reject the restricted six-regime model (a $p$-value of 0.7605). Hence, it appears that at least four regimes but no more than five regimes are needed to fit the data.

[11] Although technically invalid, the test is useful as a descriptive measure of the models' fit. Hamilton and Susmel (1994) and Gray (1996), among others, take a similar approach. Hansen (1992, 1996) has proposed a valid likelihood ratio test that overcomes this "nuisance parameter" issue. However, his procedure is numerically burdensome unless the nuisance parameter grid is extremely coarse or the model being tested is relatively simple.

A similar exercise reveals that four regimes are needed for the ARA step function model, while three regimes are needed for the ARA CTP model. For the RRA CTP model, it is apparent that at least two regimes are needed but no more than three. In the sequel these models are scrutinized in more detail in order to determine whether or not they fully capture the GARCH effects in the data.

Table 2 shows the parameter estimates and the standard errors for the two-regime through six-regime RRA step function models and the two-regime through five-regime ARA step function models. It is apparent that both the five-regime RRA model and the four-regime ARA model, which from Table 1 appear to offer the best fit to the data, contain a regime with extremely high volatility. Specifically, see regime 3 of the five-regime RRA model and regime 2 of the four-regime ARA model. Furthermore, note that the high volatility regime in each model is quite "narrow", i.e., the two switching boundaries that define this regime are very close to one another. This suggests that the highest volatility regime for each model has very low persistence, a fact that is consistent with other regime-switching studies.[12]

This can also be seen graphically, as shown in Fig. 5 for the RRA step function model. Fig. 5 illustrates the estimated volatility level, the smoothed probability of the estimated regime, and the probability of transitioning to a regime other than the estimated regime.[13] To determine the estimated regime, the value of $x_t$ on each date is compared to the estimated switching points, $\eta_1, \eta_2, \ldots, \eta_N$. For each model in Fig. 5, note the asymmetrical nature of the smoothed probability and the probability of transitioning to different regime. When the smoothed probability is close to 1, the probability of transitioning to a different regime is very close to zero. On the other hand, when the smoothed probability is much less than 1, the probability of transitioning to a different regime is much greater than zero. The likelihood of transitioning to a different regime generally increases if $x_t$ is very near one of the switching boundaries or if the volatility of the current regime is very high. Both of these conditions hold for the high volatility regimes in Fig. 5. This illustrates the very low persistence of the volatility "spikes".

Focusing on the five-regime model in Fig. 5, note that the volatility spikes all occur during 1992. According to the University of Michigan's Consumer Sentiment Index and the Conference Board's Consumer Confidence Index, 1992 was a period of relatively low consumer confidence. The U.S. economy had technically been in a recession from July 1990 until March 1991. Although many economic sectors had recovered by mid-1991, real personal (non-agricultural) income did not reach its trough until January 1992, a full 10 months after the official end of the recession. In a period of low consumer confidence, one can imagine that the most risk averse households will pull their cash out of the stock market, leaving only the most risk tolerant investors to hold the existing pool of equities. Evidently, the estimated RRA step function model is capturing this phenomenon, which

[12] In a two-regime model, Turner et al. (1989) find a persistent, low variance state and a less persistent, high variance state. The variance of the high variance state is more than four times that of the low variance state. Campbell and Li (2003) find a very similar result. Also see Gray (1996), Schaller and Van Norden (1997), Perez-Quiros and Timmermann (2001) and Ang and Bekaert (2002b). Along similar lines, Chauvet and Potter (2000) find that bear market volatility is 10 times higher than bull market volatility.

[13] The smoothed probability is the estimated probability of each regime at date $t$ given the entire sequence of observations from date 0 through date $T$. See Hamilton (1994, p. 694).

Table 2
Maximum likelihood results for the RRA and ARA models

| | RRA model estimated value | RRA model standard error | ARA model estimated value | ARA model standard error |
|---|---|---|---|---|
| *Two regimes* | | | | |
| $\eta_1$ | 16.3838 | 1.2183 | 3.1622 | 0.0689 |
| $\sigma_1$ | 0.1496 | 0.0056 | 0.1830 | 0.0033 |
| $\sigma_2$ | 0.1792 | 0.0161 | 0.7805 | 0.0760 |
| *Three regimes* | | | | |
| $\eta_1$ | 15.1075 | 0.2284 | 2.6452 | 0.0670 |
| $\eta_2$ | 20.7381 | 0.7999 | 3.3534 | 0.3475 |
| $\sigma_1$ | 0.1681 | 0.0075 | 0.1633 | 0.0059 |
| $\sigma_2$ | 0.0778 | 0.0101 | 0.3372 | 0.0429 |
| $\sigma_3$ | 0.1721 | 0.0157 | 0.7987 | 0.0821 |
| *Four regimes* | | | | |
| $\eta_1$ | 14.9974 | 0.8231 | 2.6434 | 0.0391 |
| $\eta_2$ | 15.0530 | 0.2243 | 2.6582 | 0.0960 |
| $\eta_3$ | 20.7357 | 0.9951 | 3.1600 | 0.0699 |
| $\sigma_1$ | 0.1574 | 0.0067 | 0.1636 | 0.0059 |
| $\sigma_2$ | 0.8502 | 1.6715 | 1.1865 | 3.1730 |
| $\sigma_3$ | 0.0789 | 0.0109 | 0.1779 | 0.0250 |
| $\sigma_4$ | 0.1728 | 0.0159 | 0.7802 | 0.0759 |
| *Five regimes* | | | | |
| $\eta_1$ | 8.1108 | 0.3730 | 2.6243 | 6.2562 |
| $\eta_2$ | 15.0996 | 0.6034 | 2.6442 | 0.0412 |
| $\eta_3$ | 15.1486 | 0.5891 | 2.6589 | 0.1038 |
| $\eta_4$ | 20.7348 | 0.7515 | 3.1598 | 0.0700 |
| $\sigma_1$ | 0.2499 | 0.0794 | 0.1635 | 0.0059 |
| $\sigma_2$ | 0.1471 | 0.0071 | 0.1743 | 3.5266 |
| $\sigma_3$ | 1.0888 | 1.4146 | 1.2166 | 3.3929 |
| $\sigma_4$ | 0.0787 | 0.0109 | 0.1776 | 0.0258 |
| $\sigma_5$ | 0.1732 | 0.0160 | 0.7804 | 0.0760 |
| *Six regimes* | | | | |
| $\eta_1$ | 8.1094 | 0.3729 | – | – |
| $\eta_2$ | 15.0996 | 0.9406 | – | – |
| $\eta_3$ | 15.1489 | 0.1998 | – | – |
| $\eta_4$ | 20.7088 | 1.4974 | – | – |
| $\eta_5$ | 23.6107 | 50.7664 | – | – |
| $\sigma_1$ | 0.2502 | 0.0799 | – | – |
| $\sigma_2$ | 0.1472 | 0.0071 | – | – |
| $\sigma_3$ | 1.0644 | 1.4539 | – | – |
| $\sigma_4$ | 0.0787 | 0.0114 | – | – |
| $\sigma_5$ | 0.1609 | 0.2060 | – | – |
| $\sigma_6$ | 0.1747 | 0.0166 | – | – |

The maximum likelihood parameter estimates and their standard errors are shown for the two-regime through six-regime RRA model and the two-regime through five-regime ARA model. The six-regime ARA model is not estimated.

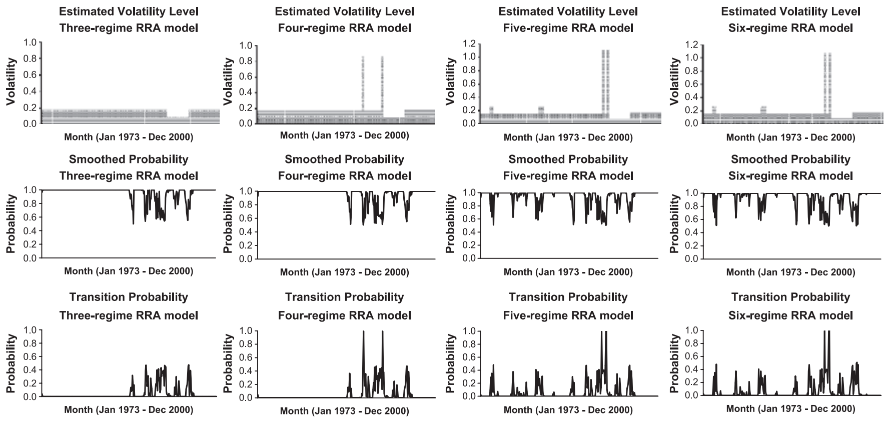


Fig. 5. The figure illustrates the volatility of the estimated regime, the smoothed probability of the estimated regime and the probability of transitioning to a different regime for the three-, four-, five- and six-regime RRA models.

shows up in the model as a low implied risk aversion level and a correspondingly high volatility for a portion of 1992.

The fact that low confidence and high volatility are related is consistent with much of the existing literature. Schwert (1989), Hamilton and Susmel (1994) and Hamilton and Lin (1996) all provide evidence that the volatility of stock returns is higher during recessions than during expansions. Along similar lines, Chauvet and Potter (2000) find that bear markets are less persistent and are associated with higher volatility than bull markets. Perez-Quiros and Timmermann (2001) also find that high volatility is associated with recessions. In fact, their Fig. 1 indicates that at least a portion of 1992 included a high volatility regime, which matches up nicely with the volatility spikes in Fig. 5. Lastly, Ang and Bekaert (2002b) provide similar evidence for interest rates. They show that the classification of regimes is closely related to the business cycle and that recessions are characterized by higher volatility.

### *4.2. Implied risk aversion estimates*

The results of Table 2 are benchmarked against the existing literature by examining the implied risk aversion estimates. These are recovered from the volatility estimates by inverting expressions (8) and (19). For the five-regime RRA model, the estimated risk aversion levels are $\gamma_1$=0.91, $\gamma_2$=2.62, $\gamma_3$=0.05, $\gamma_4$=9.16 and $\gamma_5$=1.89. For reference, this estimated RRA step function is compared to the empirical RRA function that is uncovered by Aït-Sahalia and Lo (2000). Using daily S&P 500 options data from 1993, Aït-Sahalia and Lo (2000) uncover a U-shaped function with a weighted average risk aversion level of 12.7. Since all of the observations of the S&P 500 during 1993 lie within regime 4 of the five-regime model, the step function model predicts a constant level of RRA during 1993 equal to 9.16. This compares favorably with Aït-Sahalia and Lo's relatively high average level of 12.7. On the other hand, the step function model is unable to detect their U-shape during 1993. This is probably due to the use of monthly observations here versus their use of daily observations.

The estimated five-regime RRA model is also compared to the empirical risk aversion dynamics of Rosenberg and Engle (2002). Analyzing the 60-month time period from 1991 through 1995, Rosenberg and Engle find that the empirical RRA level has a mean value of 7.36, a standard deviation of 2.58 and a range of 2.36–12.55. In addition, as their Fig. 7 illustrates, the RRA level varies substantially over time and appears to be trending upwards. Since all of the S&P 500 observations during 1991–1995 lie within regimes 2, 3 and 4 of the five-regime model, the range for the step function model is 0.05–9.16 and the weighted average RRA level (using the number of observations for each level) is 7.33. This latter number is very near Rosenberg and Engle's result. As for the dynamics of the RRA level, for the first 11 months of 1991, the S&P 500 is in regime 2 ($\gamma_2$=2.62) and, during the 36-month period including 1993–1995, the S&P 500 is in regime 4 ($\gamma_4$=9.16). The last month of 1991 and all of 1992 is a transitory period in which the S&P 500 makes 10 different regime changes, but ultimately the RRA level increases from 2.62 to 9.16. These results appear to be consistent with Rosenberg and Engle (2002). For example, in their Fig. 7, the average RRA level during 1991 is much lower than the average level during 1993–1995. In addition, while there are some missing observations in their figure, 1992 appears to have been a period of increasing risk aversion.

For the four-regime ARA step function model, the estimated risk aversion levels are $\gamma_1$=5.22, $\gamma_2$=0.10, $\gamma_3$=4.41 and $\gamma_4$=0.23. These results compare favorably to those of Jackwerth (2000). In particular, Jackwerth's results indicate that the empirical ARA function during the 1986–1995 time period is U-shaped. Since all of the S&P 500 observations during the 1986–1995 time period lie within the first three regimes of the estimated four-regime model, the estimated ARA step function is consistent with the U-shaped nature of Jackwerth's empirical ARA function.

### 4.3. GARCH tests

The errors from the fitted regime-switching models are analyzed in order to assess each model's ability to capture the GARCH effects in the data. For example, if the RRA step function model in Eq. (31) is correctly specified, the errors

$$v_{t+\delta t} \equiv \frac{\log x_{t+\delta t} - E_t[\log x_{t+\delta t}]}{\sigma_t}$$

should be i.i.d. normal with mean zero and variance $\delta t$. To construct $v_{t+\delta t}$, the maximum likelihood estimates for the five-regime RRA model are used and the value of $x_t$ on each date is compared to the estimated switching points $\eta_1,\eta_2,\ldots,\eta_N$ in order to identify the date $t$ regime.[14] Likewise, if the ARA step function model in Eq. (40) is correctly specified, the errors

$$v_{t+\delta t} \equiv \frac{x_{t+\delta t} - E_t[x_{t+\delta t}]}{\sigma_t}$$

[14] Alternatively, the date $t$ regime can be identified by using the regime with largest smoothed probability. The results presented here are invariant to which identification method is used. For the five-regime RRA step function model, the two methods disagree in only 5 months out of a total of 336 months. For the four-regime ARA model, the two methods disagree in only 2 months.

should also be i.i.d. normal with mean zero and variance $\delta t$. In this case, the maximum likelihood estimates for the four-regime ARA step function model are used to construct $v_{t+\delta t}$. Each step function model is compared to its one-regime counterpart, i.e., the RRA step function model is compared to a CRRA benchmark model and the ARA step function model is compared to a CARA benchmark. The CRRA and CARA models are also estimated using maximum likelihood. Lastly, given the results in Table 1, the errors from the three-regime RRA CTP model and the three-regime ARA CTP model are also analyzed. To construct the errors for the CTP models, the regime with the largest smoothed probability at each date is assumed to be the true regime.

Several tests are used to assess the models. First, the GARCH(1,1) model of Bollerslev (1986) is fit to each model's estimation errors. As discussed in Bollerslev et al. (1992), the GARCH(1,1) model fits the underlying data quite well and thus it seems natural to rely on this model for analyzing the estimation errors of the fitted regime-switching models. The GARCH(1,1) model is $v_t = c_0 + \varepsilon_t$, where $\varepsilon_t$ is normally distributed with zero mean and variance $h_t = c_1 + c_2 h_{t-1} + c_3 \varepsilon_{t-1}^2$. A statistically significant value for either $c_2$ or $c_3$ would appear to indicate that the regime-switching model under analysis does not fully capture the volatility dynamics of the data. Second, a formal likelihood ratio (LR) test is performed, which is a joint test for $c_2 = c_3 = 0$. Lastly, as shown by Lee (1991), Engle's (1982) test for the presence of ARCH($q$) against the null of white noise is equivalent to testing for the presence of GARCH($p$,$q$) against the null of white noise. Hence, Engle's test with 1 lag is performed as an alternative GARCH(1,1) test.

The test results are shown in Table 3. For reference, the top row of the table shows the results for the actual S&P 500 data. As the top row illustrates, the persistence in the S&P 500 volatility is quite high, as evidenced by the large and statistically significant estimate for $c_2$. Similar results for the S&P 500 can be found in Engle and Mustafa (1992) and Jacobsen and Dannenburg (2003). The other rows of Fig. 3 show the test results for the RRA models and the ARA models, respectively. Note that the CRRA model, the CARA model, the ARA step function model, and the ARA CTP model do not capture the GARCH effects in the data. Each of these four models produces a statistically significant estimate for $c_2$ whose magnitude is very close to that of the underlying data. Furthermore, the CRRA model, the CARA model and the ARA step function model have statistically significant estimates for $c_3$. Although the Engle test and the LR test do not reject some of these models at traditional significance levels, the persistence in the volatility of the fitted errors indicates that all four models should be rejected.

On the other hand, the RRA step function model and the RRA CTP model offer a much better fit to the data. The $c_2$ and $c_3$ coefficients for both models are statistically insignificant and neither model can be rejected based on the Engle test or the LR test. A close examination of the GARCH results, however, suggests that these two models are somewhat different. For the RRA CTP model, the point estimate for $c_2$ is 0.8644 and is nearly significant (a $t$-statistic of 1.53), while the point estimate for $c_3$ is 0.0006 and is highly insignificant. For the RRA step function model, the point estimate for $c_2$ is 0.0126, while the point estimate for $c_3$ is 0.0779. While the standard errors for the step function model are considerably larger than those of the CTP model, this suggests that the step

Table 3
GARCH test results

| Model | No. of regimes | $c_0$ | $c_1$ | $c_2$ | $c_3$ | Engle (1lag) | No GARCH LR test |
|---|---|---|---|---|---|---|---|
| S&P 500 | – | 0.1117 | 0.0002 | 0.8423 | 0.0577 | 3.351 | 6.468 |
| | | (5.085)** | (1.328) | (9.886)** | (2.004)* | (0.0672) | (0.0394) |
| CRRA | 1 | 0.0109 | 0.0065 | 0.8647 | 0.0504 | 3.470 | 5.504 |
| | | (0.724) | (1.241) | (10.995)** | (1.863)* | (0.0625) | (0.0638) |
| RRA | 5 | 0.0078 | 0.0986 | 0.0126 | 0.0779 | 0.338 | 4.330 |
| | | (0.383) | (1.062) | (0.014) | (1.072) | (0.5609) | (0.1147) |
| RRA CTP | 3 | 0.0062 | 0.0217 | 0.8644 | 0.0006 | 0.508 | 4.983 |
| | | (0.251) | (0.241) | (1.530) | (0.016) | (0.4761) | (0.0828) |
| CARA | 1 | 0.0220 | 0.0002 | 0.8754 | 0.1402 | 11.849 | 297.269 |
| | | (3.235)** | (1.019) | (24.613)** | (3.382)** | (0.0006) | (0.0000) |
| ARA | 4 | 0.0211 | 0.0047 | 0.8886 | 0.0534 | 2.483 | 22.584 |
| | | (1.421) | (1.762)* | (18.247)** | (2.337)** | (0.1151) | (0.0000) |
| ARA CTP | 3 | 0.0141 | 0.0094 | 0.8489 | 0.0061 | 0.187 | 8.976 |
| | | (1.095) | (0.480) | (2.788)** | (0.289) | (0.6650) | (0.0112) |

A GARCH(1,1) model is fit to the actual S&P 500 data (top row) and the maximum likelihood (ML) estimation errors for the given models (other rows). For the CTP models, the ML error is constructed by assuming that the regime with the largest smoothed probability is the true regime. The GARCH(1,1) model is $v_t$=$c_0$+$\varepsilon_t$, where $\varepsilon_t$ is normally distributed with zero mean and variance $h_t$=$c_1$+$c_2h_{t-1}$+$c_3\varepsilon_{t-1}^2$. $t$-statistics are shown in parentheses, and * indicates significance at the 5% level and ** at the 1% level. The Engle test for ARCH disturbances is asymptotically distributed $\chi^2(1)$. The last column is a likelihood ratio (LR) test statistic for $c_2$=$c_3$=0, which is asymptotically distributed $\chi^2(2)$. For the last two columns, the $p$-values are in parentheses.

function model is better at capturing volatility persistence. Likewise, the CTP model appears to be better at capturing the sensitivity of volatility to the squared innovations, $\varepsilon_t^2$.

These results are also illustrated in Fig. 6, which shows the sample autocorrelations of the squared errors for lags 1 through 20. The dashed lines in each panel correspond to the 95% confidence band. In the case of i.i.d. normal errors, we should expect to see no more than one autocorrelation coefficient that lies outside of the confidence band. For the CRRA model, there are two values that are very near the upper band, while for the CARA model all but one value lies outside the band. Similarly, the ARA step function model and the two CTP models all have at least one autocorrelation coefficient that lies outside of the confidence band. In contrast, the five-regime RRA step function model offers a considerably improved fit over both the CRRA model and the RRA CTP model. For the RRA step function model, none of the autocorrelation coefficients lie outside the confidence band. This latter result is consistent with Diebold et al. (1994) who show that time varying transition probabilities offer a better fit to the data relative to CTP models.

### 4.4. Option pricing

As noted earlier (see Section 3.2 and Fig. 4), the step function equilibrium models are capable of generating an implied volatility skew and a non-flat term structure of implied volatility. These topics are now revisited with an emphasis on assessing the ability of the step function approach to empirically match the options data. For brevity, since option

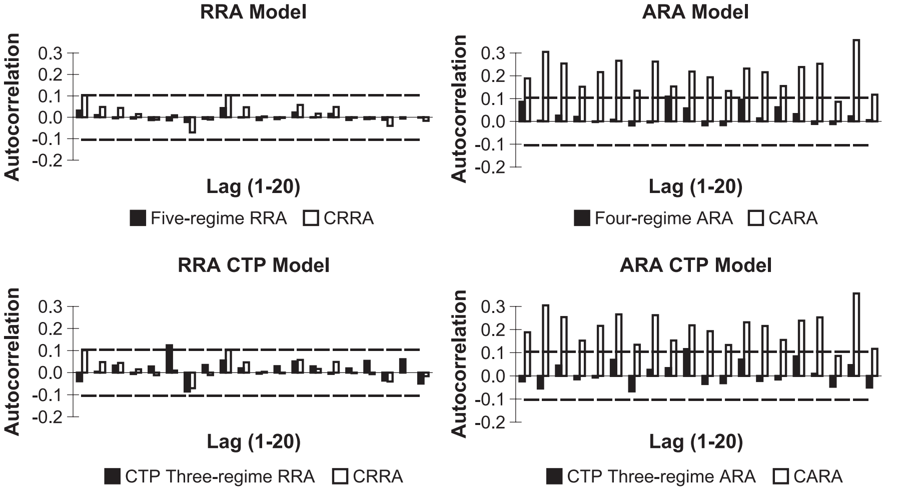


Fig. 6. The sample autocorrelations for the squared maximum likelihood (ML) estimation errors are shown for lags 1–20. The upper left panel shows the five-regime RRA model and the CRRA benchmark, while the upper right panel shows the four-regime ARA model and the CARA benchmark. The two lower panels show the three-regime RRA and ARA CTP models. For the CTP models, the ML error is constructed by assuming that the regime with the largest smoothed probability is the true regime. The dashed lines illustrate the 95% confidence region.

pricing in CTP regime-switching models is examined elsewhere,[15] the analysis here focuses solely on the step function models. The data for the analysis consists of daily S&P 500 index option settlement prices for June 1992–November 1998. The June 1992 start date is chosen to approximately coincide with the point in time that the Chicago Board Options Exchange (CBOE) first listed the three "near term" expiration months for the S&P 500 contract. Specifically, at the end of month $m$, data for option contracts expiring in months $m$+1, $m$+2 and $m$+3 are available. These are the "near term" contracts.[16] The existence of these three contracts greatly increases our ability to precisely control for both expiration effects and moneyness effects when calculating implied volatilities. It is also worth noting that the June 1992–November 1998 sample period includes the volatility "spikes" that are evident in Fig. 5.

For each month during the period June 1992–November 1998, nine different option prices are extracted from the options settlement data. Specifically, given the index level at the beginning of month $m$, option prices are recorded for contracts expiring in month $m$+1 whose strike prices are $\pm 1\%$, $\pm 2\%$, $\pm 3\%$ and $\pm 4\%$ away from the index level. Focusing on options whose strike prices are within 4% of the index level is consistent with Buraschi and Jackwerth (2001), who claim that options outside of this range may have large pricing

[15] See Bollen (1998), Bollen et al. (2000) and Campbell and Li (2003). In addition, the CTP regime-switching model is nested within the framework of Duan et al. (2002).

[16] Data for the standard quarterly expiration cycle are also available. The exact expiration date is the Saturday following the third Friday of the expiration month. In addition, the S&P 500 option contract is European and hence early exercise is not an issue.

errors. These eight option prices are then inverted to obtain the implied volatilities. Plotting these volatilities as a function of option moneyness produces the implied volatility skew for month $m$. In addition to these eight options, the nearest-the-money contract at the beginning of month $m$ that expires in month $m$+2 is extracted from the data. The price of this option contract is paired up with the price for an otherwise similar contract that expires in month $m$+1, which has already been extracted as part of the volatility skew data. The prices for the pair of nearest-the-money options (expiring in months $m$+1 and $m$+2) are then inverted to obtain an estimate of the term structure of implied volatility for month $m$. For reference, the average time to expiration for the options that expire in month $m$+1 is 35 days, while the average time to expiration for the options that expire in month $m$+2 is 60 days. The short-dated nature of these options helps to ensure that stale prices are avoided. Indeed, as reported by Buraschi and Jackwerth (2001), most of the trading activity in S&P 500 options is concentrated in option contracts that have no more than 60 days until expiration.

Given the contractual specifications of the nine option contracts in each month, the options are valued using the fitted five-regime RRA step function model. Specifically, the theoretical option prices are obtained by using Hull and White's (1990) explicit finite difference scheme. Their technique, as applied to our problem, involves transforming log $x(t)$ in expression (30) into a new stochastic process that has constant volatility.[17] A trinomial tree for the risk-adjusted transformed process is then constructed and the transformed option payoffs are valued by taking an expectation and discounting at the riskless rate. The theoretical option prices are then used to construct an implied volatility skew and a term structure of implied volatility that are consistent with the fitted five-regime RRA step function model.

Fig. 7 shows both the theoretical volatility skew (top panel) and the actual volatility skew (bottom panel) for the entire 78-month sample period. While Fig. 4 characterized the volatility skew for a simple two-regime model, the top panel of Fig. 7 allows for a more insightful analysis. During several months in 1992, the skew is "hump"-shaped. This reflects the fact that the current regime, which is regime 3, has a higher volatility level than the two adjoining regimes. Note that the "hump" is not very pronounced since the probability of transitioning from regime 3 to a different regime is very high. From November 1992 to June 1996, the skew is mostly flat with an implied volatility level of about 7.8%. This reflects the fact that the S&P index is in regime 4 during this period. Periodically, the index nears the boundary between regime 4, which is a low volatility regime, and regime 3, which has higher volatility. This first happens in March 1994 and then again in June and December of 1994. For all 3 months, the skew is decreasing as a function of moneyness, which is consistent with the skew in the right-hand panel of Fig. 4. The end of 1995 and the beginning of 1996 is a transition period during which the S&P index moves from regimes 4 to 5, then back to 4, and finally settles in regime 5. Note that, during March 1996, the skew is "U-shaped". This reflects the fact that the index is in regime 4, which has lower volatility than the two adjoining regimes.

[17] The transformation captures the fact that the volatility is state dependent while at the same time allowing us to work in a constant volatility setting, which is numerically efficient. For the exact details of the procedure, see Hull and White (1990, Section III).

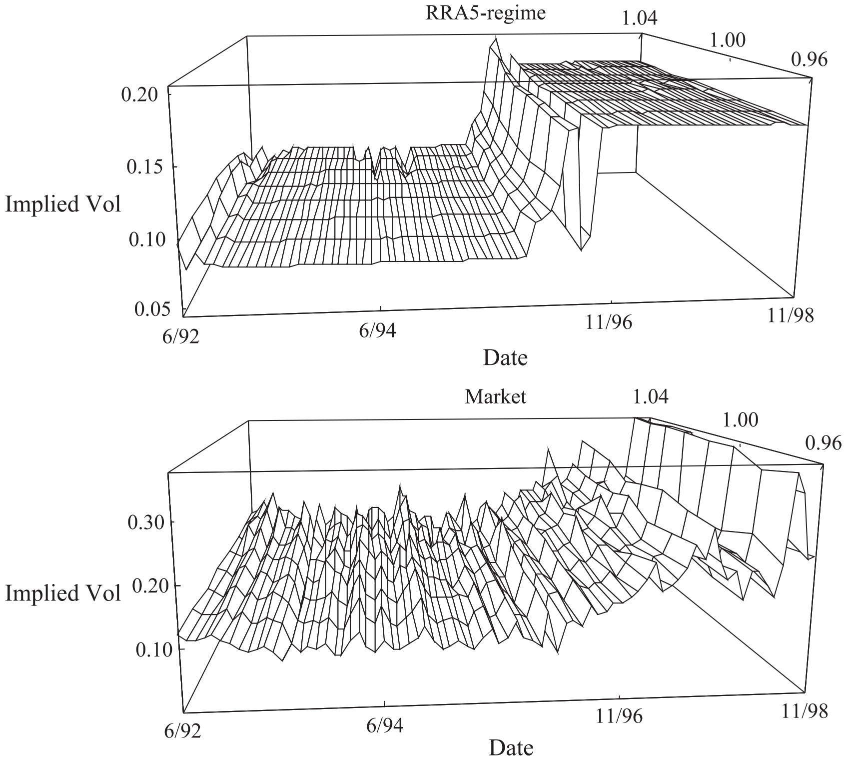


Fig. 7. The figure shows the S&P 500 implied volatility skew for the 78-month period from June 1992 through November 1998. The top panel uses the five-regime RRA model to generate the option prices, while the bottom panel uses the actual market data. The options' moneyness, which is calculated by dividing the index level by the strike price, varies from 0.96 to 1.04.

Essentially, this is the opposite of the hump shape mentioned earlier. After November 1996, the skew is mostly flat since the index is in regime 5 and is not near any of the switching boundaries.

If the two panels of Fig. 7 are compared, it is evident that the behavior of the actual volatility skew is much more dynamic than that of the theoretical volatility skew. The transition from one volatility level to the next is not nearly as distinct as in the regime-switching model. In addition, for the latter part of the sample period, the actual implied volatility levels are considerably higher than those produced by the model. Both of these facts indicate that there are large pricing errors associated with the fitted model. Indeed, a comparison of the theoretical values to their actual counterparts reveals that nearly 78% of the options have pricing errors in excess of 10%. These large, economically significant errors lead us to conclude that the fitted regime-switching model does not adequately capture the dynamics of the S&P 500 index options market. Alternatively, option prices do not fully reflect the regime-switching information that is present in the underlying asset market. Bollen et al. (2000) find a similar result for currency options.

Fig. 8 shows the theoretical (top panel) and actual (bottom panel) term structures of implied volatility for the entire sample period. As the figure illustrates, the term structure dynamics are quite different for the two panels. For the top panel, the term structure is always upward sloping, i.e., the 60-day option in every month has a higher implied volatility than the 35-day option. However, the bottom panel reveals that the actual term structure slopes downward for 29 of the 78 months. In addition, the theoretical model's implied volatilities are considerably higher than those that are calculated from actual market prices, especially for the 60-day option. This appears to reflect the fact that, for distant expiration dates, there is more time for the underlying asset to visit a high volatility regime (e.g., regime 3). As in the case of the volatility skew, the differences in the implied volatilities between the two panels suggest that the

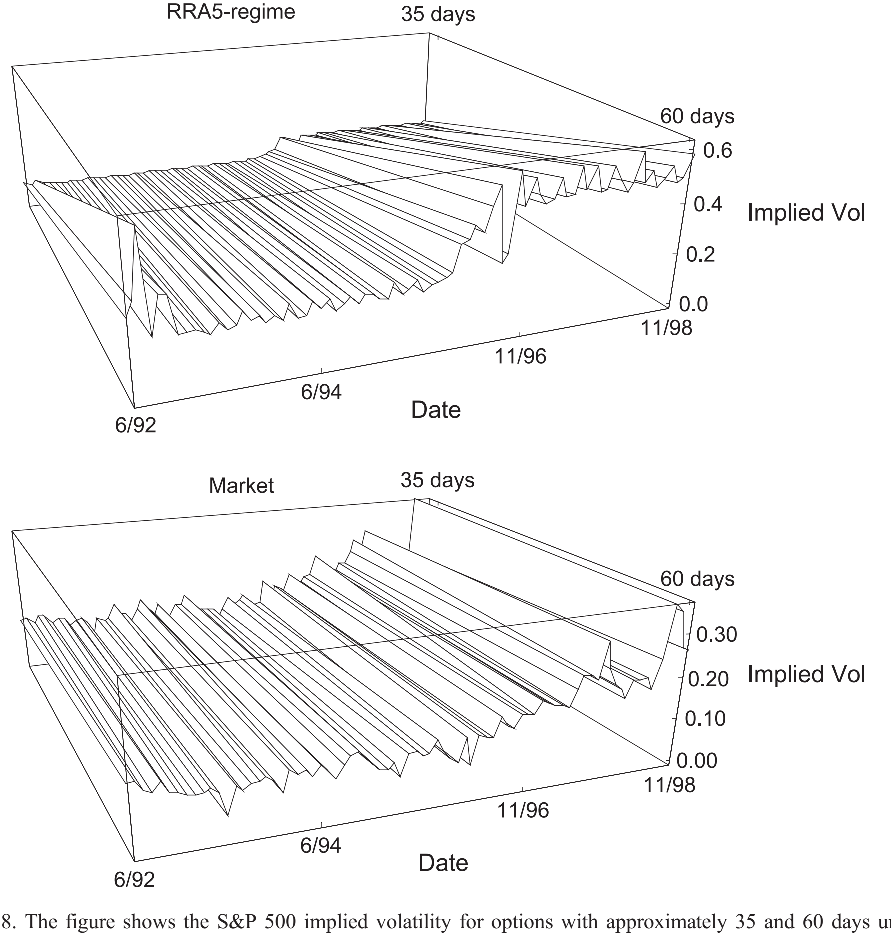


Fig. 8. The figure shows the S&P 500 implied volatility for options with approximately 35 and 60 days until expiration over the 78-month period from June 1992 to November 1998. The top panel uses the five-regime RRA model to generate the option prices, while the bottom panel uses the actual market data. The nearest-the-money options are used for each month.

fitted model is generating economically significant pricing errors. Indeed, 84% of the options used to construct the term structure of implied volatility have pricing errors in excess of 10%. This again provides evidence that the options market is not fully incorporating the regime-switching information that is present in the underlying asset market.

The results are not improved if instead the four-regime ARA model is used to generate the theoretical option prices. The pricing errors are at least as large as those of the five-regime RRA model. Furthermore, recall from Section 4.3 that the ARA step function model does not fully capture the GARCH effects in the data, i.e., the persistence in the volatility of the ARA model's errors is quite strong. Hence, while the inability of the five-regime RRA model to explain the options data can be attributed to the possibility that option traders do not fully account for regime-switching information when valuing options, this same argument does not apply to the ARA step function model. The ARA step function model does not fully capture the dynamics of the underlying asset market and it should be rejected.

Lastly, Fig. 9 compares the expected volatility level from the step function models to two different measures of the actual volatility level. The expected volatility is calculated by using the estimated volatilities and the smoothed probabilities for each model. The first measure of the actual volatility level is obtained by calculating the sample standard deviation of the S&P 500 return using a centered 13-month rolling window. This is shown in the top two panels of Fig. 9 along with the expected volatilities for the full 336-month sample period. The second measure of the actual volatility level is obtained from the CBOE's VIX implied volatility index. The VIX index is a measure of market

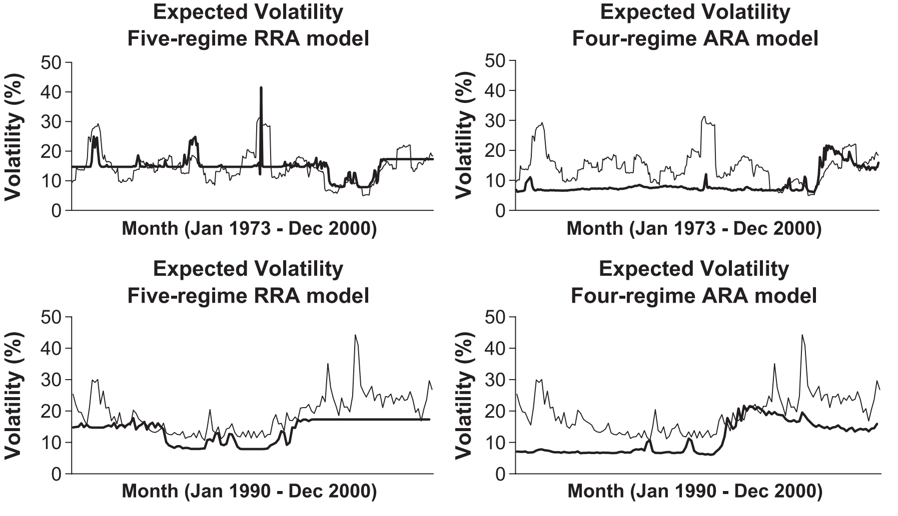


Fig. 9. The figure shows the expected volatility for the five-regime RRA model and the four-regime ARA model using the maximum likelihood volatility estimates and the smoothed probabilities. The lightly shaded line in the top two panels is the sample volatility of the S&P total return using a centered 13-month rolling window. The lightly shaded line in the bottom two panels is the VIX implied volatility index of the Chicago Board Options Exchange.

expectations of near-term volatility as conveyed by the prices of S&P 500 index options.[18] This is shown in the bottom two panels of Fig. 9 along with the expected volatilities for the period January 1990–December 2000. As Fig. 9 illustrates, the five-regime RRA model tracks the rolling S&P 500 standard deviation quite well, but does not capture the dynamics of the VIX implied volatility. On the other hand, the ARA step function model does not track either volatility measure very well. This evidence is consistent with the above conclusions.

### 4.5. GMM tests

To formally assess the specification of the two step function models, Hansen's (1982) GMM is applied to the pricing errors that are derived from each model's Euler equation. Since the above estimation results indicate that five regimes should be enough to explain the data, the GMM analysis also focuses on five regimes or less. This conveniently allows for decile portfolios to be used in the analysis while still maintaining a model that is overidentified. To see this, recall from Table 1 that three parameters must be estimated for the two-regime model, five parameters for the three-regime model, seven for the four-regime model and nine for the five-regime model. Hence, there is one degree of freedom for the five-regime model.

The data used in the GMM analysis are value weighted, size sorted decile portfolios formed from the NYSE/AMEX/NASDAQ universe of equities. The monthly data is obtained from the CRSP database for the sample period 1926–2000 and the standard labeling convention is followed so that small=1 and big=10. The main reason for using size as the sorting variable is to facilitate an analysis of model specification. In particular, as discussed in Berk (1995), if an asset pricing model is misspecified then there should be some variation between the model's unexplained pricing errors and portfolio size. Hence, in addition to the standard test of model specification based on the overidentifying restrictions, particular attention is paid to how the magnitude of the average pricing error varies with size.

To describe the GMM procedure, fix the number of volatility regimes. The Euler equation for either model can be written as

$$E_t\left[\frac{\xi_{t+1}(\theta)}{\xi_t(\theta)}R_{t+1}\right] = 1 \tag{41}$$

where $\xi_t$ is the appropriate pricing kernel at date $t$, $\theta$ is a vector of model parameters and $R_{t+1}$ is a vector of gross portfolio returns from $t$ to $t$+1. The exact functional form for each model's pricing kernel is given in Appendix A. With a slight abuse of notation, the vector of observed pricing errors at date $t$+1 is defined as

$$u_{t+1} \equiv \frac{\xi_{t+1}(\theta)}{\xi_t(\theta)}R_{t+1} - 1 \tag{42}$$

[18] Fig. 9 uses historical data for the new VIX index that is based on S&P 500 index options. The old VIX index, which now has the ticker VXO, is based on S&P 100 index options. These changes to the implied volatility index were implemented by the CBOE on September 22, 2003.

and hence the average pricing error vector, $g(\theta)$, is

$$g(\theta) = \frac{1}{T}\sum_{t=1}^{T} u_t \tag{43}$$

where $T$ is the sample size. For the first stage, the identity matrix $I$ is used as the weighting matrix and the following problem is solved

$$SSE \equiv \min_{\theta} Tg(\theta)' I g(\theta) \tag{44}$$

where $g(\theta)'$ denotes the transpose of Eq. (43). Solving Eq. (44) produces an initial estimate of the parameters, $\theta_1$. Using $\theta_1$ along with the null hypothesis that the model is correctly specified, an estimate of the optimal weighting matrix is formed via the expression

$$S(\theta_1) = \frac{1}{T}\sum_{t=1}^{T}\left[\left(u_t - g(\theta_1)\right)\left(u_t - g(\theta_1)\right)'\right] \tag{45}$$

The second stage minimization problem

$$TJ_T \equiv \min_{\theta} Tg(\theta)' S(\theta_1)^{-1} g(\theta) \tag{46}$$

is then solved which produces an updated parameter estimate, $\theta_2$. Steps (45)–(46) are then repeated until the solution converges. As discussed in Ferson and Foerster (1994) and Hansen et al. (1996), the technique of iterating until convergence may improve the small sample properties of the results.

The GMM results for the RRA step function model are shown in Table 4. As the first stage results in the top panel illustrate, the SSE and the vector of average pricing errors for the four-regime and five-regime models are almost identical. This is consistent with the earlier evidence that a four-regime or perhaps five-regime representative agent model best fits the data. However, there is a systematic pattern across portfolio size, i.e., small stocks have positive average pricing errors, while big stocks have negative average pricing errors. In addition, the average pricing error for the smallest stock decile is relatively large and economically significant, while the average pricing errors for bigger stocks are much smaller in magnitude. Using Berk's (1995) critique, this indicates that the RRA step function model is misspecified.

The bottom panel of Table 4 shows the iterated GMM results for the RRA model. As the two far right columns show, the four-regime model and five-regime model have almost identical average pricing errors and objective function values. While the $TJ_T$ statistics fail to reject several of the models at conventional significance levels, the statistics' point estimates are quite large. Hence, even though an iterated procedure is used to improve the small sample performance, it is probably the case that the chi-squared asymptotic distribution for these statistics is not a reasonable approximation to their small sample distributions (a similar point is raised in Chapman, 1997). The average pricing errors in the bottom panel also vary in magnitude with portfolio size and there is clearly no appreciable

Table 4
GMM results for the RRA model

First-stage GMM results for the RRA model

Average pricing errors

| Portfolio | 1-regime | 2-regimes | 3-regimes | 4-regimes | 5-regimes |
|---|---|---|---|---|---|
| 1 | 0.00456 | 0.00448 | 0.00438 | 0.00401 | 0.00400 |
| 2 | 0.00088 | 0.00080 | 0.00077 | 0.00046 | 0.00046 |
| 3 | −0.00023 | −0.00030 | −0.00028 | −0.00047 | −0.00047 |
| 4 | −0.00057 | −0.00064 | −0.00059 | −0.00072 | −0.00072 |
| 5 | −0.00046 | −0.00053 | −0.00045 | −0.00054 | −0.00054 |
| 6 | −0.00038 | −0.00046 | −0.00037 | −0.00044 | −0.00044 |
| 7 | −0.00025 | −0.00032 | −0.00022 | −0.00025 | −0.00025 |
| 8 | −0.00084 | −0.00089 | −0.00078 | −0.00076 | −0.00077 |
| 9 | −0.00015 | −0.00018 | −0.00010 | −0.00003 | −0.00003 |
| 10 | −0.00042 | −0.00040 | −0.00038 | −0.00021 | −0.00022 |

Sum of squared pricing errors

| | 1-regime | 2-regimes | 3-regimes | 4-regimes | 5-regimes |
|---|---|---|---|---|---|
| SSE | 0.02095 | 0.02051 | 0.01918 | 0.01633 | 0.01632 |

Iterated GMM results for the RRA model

Average pricing errors

| Portfolio | 1-regime | 2-regimes | 3-regimes | 4-regimes | 5-regimes |
|---|---|---|---|---|---|
| 1 | 0.00457 | 0.00514 | 0.00559 | 0.00568 | 0.00568 |
| 2 | 0.00085 | 0.00143 | 0.00207 | 0.00218 | 0.00218 |
| 3 | −0.00037 | 0.00028 | 0.00110 | 0.00127 | 0.00127 |
| 4 | −0.00078 | −0.00012 | 0.00082 | 0.00097 | 0.00097 |
| 5 | −0.00070 | −0.00002 | 0.00100 | 0.00113 | 0.00113 |
| 6 | −0.00065 | 0.00005 | 0.00109 | 0.00119 | 0.00119 |
| 7 | −0.00059 | 0.00013 | 0.00126 | 0.00136 | 0.00136 |
| 8 | −0.00127 | −0.00053 | 0.00068 | 0.00077 | 0.00077 |
| 9 | −0.00066 | 0.00010 | 0.00132 | 0.00142 | 0.00142 |
| 10 | −0.00117 | −0.00036 | 0.00091 | 0.00100 | 0.00100 |

GMM objective function

| | 1-regime | 2-regimes | 3-regimes | 4-regimes | 5-regimes |
|---|---|---|---|---|---|
| $TJ_T$ | 12.09598 | 11.71843 | 11.48032 | 10.95816 | 10.95722 |
| $p$ | 0.20795 | 0.11021 | 0.04265 | 0.01195 | 0.00093 |

The top panel shows the first-stage GMM results for the RRA model while the bottom panel shows the iterated GMM results. Included are the sum of squared pricing errors (SSE), the converged objective function value $TJ_T$ and the average pricing error vectors. The $p$-values for the $TJ_T$ statistic, shown in the last row of the table, are decreasing in the number of regimes since the loss in degrees of freedom more than offsets the decrease in the value of $TJ_T$.

reduction in the magnitude as the number of regimes increases. In fact, for most of the portfolios, the magnitude is increasing in the number of regimes. Since the point estimates of the $TJ_T$ statistic are decreasing in the number of regimes, this suggests that the variance–covariance matrix is increasing in the number of regimes. The fact that a low $TJ_T$ statistic is achievable by blowing up the $S$ matrix is discussed in Cochrane (1996) and Chapman (1997).

The empirical conclusions are robust with respect to the choice of the GMM weighting matrix. While not included in the tables, the results do not change if the inverse of an estimate of the second moment matrix of returns is used as the weighting matrix. As

Table 5
GMM results for the ARA model

First-stage GMM results for the ARA model

Average pricing errors

| Portfolio | 1-regime | 2-regimes | 3-regimes | 4-regimes | 5-regimes |
|---|---|---|---|---|---|
| 1 | 0.00518 | 0.00458 | 0.00423 | 0.00401 | 0.00386 |
| 2 | 0.00151 | 0.00086 | 0.00059 | 0.00043 | 0.00031 |
| 3 | 0.00041 | −0.00018 | −0.00027 | −0.00038 | −0.00050 |
| 4 | 0.00005 | −0.00055 | −0.00061 | −0.00067 | −0.00077 |
| 5 | 0.00016 | −0.00046 | −0.00047 | −0.00052 | −0.00055 |
| 6 | 0.00023 | −0.00035 | −0.00030 | −0.00035 | −0.00036 |
| 7 | 0.00036 | −0.00025 | −0.00018 | −0.00020 | −0.00022 |
| 8 | −0.00023 | −0.00080 | −0.00071 | −0.00071 | −0.00068 |
| 9 | 0.00045 | −0.00006 | 0.00004 | 0.00005 | 0.00007 |
| 10 | 0.00013 | −0.00027 | −0.00018 | −0.00015 | −0.00005 |

Sum of squared pricing errors

| | 1-regime | 2-regimes | 3-regimes | 4-regimes | 5-regimes |
|---|---|---|---|---|---|
| SSE | 0.02680 | 0.02086 | 0.01759 | 0.01603 | 0.01508 |

Iterated GMM results for the ARA model

Average pricing errors

| Portfolio | 1-regime | 2-regimes | 3-regimes | 4-regimes | 5-regimes |
|---|---|---|---|---|---|
| 1 | 0.00528 | 0.00505 | 0.00647 | 0.00623 | 0.00681 |
| 2 | 0.00159 | 0.00135 | 0.00307 | 0.00288 | 0.00348 |
| 3 | 0.00045 | 0.00019 | 0.00231 | 0.00212 | 0.00271 |
| 4 | 0.00007 | −0.00021 | 0.00201 | 0.00185 | 0.00244 |
| 5 | 0.00017 | −0.00012 | 0.00221 | 0.00204 | 0.00263 |
| 6 | 0.00023 | −0.00006 | 0.00242 | 0.00223 | 0.00283 |
| 7 | 0.00034 | 0.00002 | 0.00249 | 0.00230 | 0.00288 |
| 8 | −0.00029 | −0.00062 | 0.00199 | 0.00181 | 0.00240 |
| 9 | 0.00036 | 0.00003 | 0.00267 | 0.00250 | 0.00309 |
| 10 | −0.00005 | −0.00039 | 0.00227 | 0.00214 | 0.00272 |

GMM objective function

| | 1-regime | 2-regimes | 3-regimes | 4-regimes | 5-regimes |
|---|---|---|---|---|---|
| $TJ_T$ | 11.99866 | 11.95208 | 11.14321 | 10.89560 | 10.72520 |
| $p$ | 0.21338 | 0.10215 | 0.04861 | 0.01230 | 0.00106 |

The top panel shows the first-stage GMM results for the ARA model while the bottom panel shows the iterated GMM results. Included are the sum of squared pricing errors (SSE), the converged objective function value $TJ_T$ and the average pricing error vectors. The $p$-values for the $TJ_T$ statistic, shown in the last row of the table, are decreasing in the number of regimes since the loss in degrees of freedom more than offsets the decrease in the value of $TJ_T$.

shown in Hansen and Jagannathan (1997), minimizing an objective function using this weighting matrix is equivalent to minimizing the distance between the model's stochastic discount factor and the space of true stochastic discount factors. In the current setup, if the Hansen-Jagannathan distance is minimized, the magnitude of the average pricing errors still varies considerably with portfolio size. Small stocks have large average pricing errors and big stocks have smaller average errors. Hence, the conclusion that the RRA step function model is likely misspecified is not altered.

The results for the ARA step function case, presented in Table 5, lead to a similar set of conclusions. The four-regime and five-regime models have similar SSEs and average pricing errors, although not as similar as in the RRA step function case. For all of the results shown in Table 5, small stocks have average pricing errors that are much larger in magnitude than the average pricing errors for big stocks. Again, this can be interpreted as evidence of model misspecification.

## 5. Concluding remarks

Since the RRA and ARA step function models can be used to closely approximate any state-independent utility function, the above results suggest that it is doubtful that any representative agent equilibrium (with state-independent utility) is consistent with the data. An equivalent interpretation is that the market portfolio price process is not a diffusion process of the type studied by Bick (1990) and He and Leland (1993). While many individual models that fall into this class have been previously tested and rejected (Dumas et al., 1998; Brandt and Wu, 2002), this article rejects the entire class of economies. In addition, even though the tests are based on a pair of representative agent models, it is worth pointing out that many multi-agent models are also members of this class. The reason is simple—many multi-agent economies can be represented by a single agent for asset pricing purposes (Constantinides, 1982).

While a large category of models is rejected by the data, excluded from this category are many of the "habit formation" models found in the literature (Campbell and Cochrane, 1999; Chan and Kogan, 2002) as well as the class of conditional "scaled factor pricing models" in which the stochastic discount factor depends on a set of instrumental variables (Cochrane, 1996). These models can be interpreted as models with state-dependent utility functions. Building state dependency into the current model might be possible by analyzing a more general setup with a multivariate step function. The agent's risk aversion in this case would vary in a dynamic fashion not only with the market portfolio level, but also with the levels of other economic variables. The variables of Rosenberg and Engle (2002, Table 7) might be a good place to start.

## Acknowledgements

I thank Franz Palm, two anonymous referees and my colleagues at the Tuck School for helpful feedback. This research was partially funded by a NASDAQ Foundation Fellowship. Research support from the Tuck School is gratefully acknowledged. This

paper was previously titled “Volatility Clustering, Time-Varying Risk Aversion and the Pricing of Options”. Any errors are mine.

## Appendix A

**Proof of Proposition 1.** The proof is carried out in the following manner. First, a path-independent pricing kernel for the economy is introduced and it is verified that this pricing kernel correctly prices the riskless bond and the risky asset. Since the market structure is dynamically complete, the representative agent’s dynamic investment problem in Eqs. (9)–(11) is then recast into a static problem using the technique of Cox and Huang (1989). Upon solving the agent’s problem it is found that $W(T)=S(T)$, i.e., the agent optimally consumes the date $T$ value of the market portfolio. Since the agent is endowed with one unit of the market at date 0 and optimally consumes one unit of the market at date $T$, the agent’s optimal self-financing investment policy is to hold one unit of the market portfolio at all dates $t\in[0,T]$. Lastly, it is verified that the agent’s indirect marginal utility coincides with the aforementioned path-independent pricing kernel.[19]

To proceed with the proof, define the function $g(t)$ as

$$g(t) = \mathrm{e}^{\left(r+\frac{\lambda}{2}\right)(T-t)} \tag{47}$$

Next, introduce the pricing kernel $\xi(t)$ defined as

$$\xi(t) = \begin{cases} (S(t))^{-\gamma_1}(g(t))^{1-\gamma_1}, & S(t)\leq b_1(t) \\ k_2(S(t))^{-\gamma_2}(g(t))^{1-\gamma_2}, & b_1(t)<S(t)\leq b_2(t) \\ \vdots & \vdots \\ k_{N+1}(S(t))^{-\gamma_{N+1}}(g(t))^{1-\gamma_{N+1}}, & b_N(t)<S(t) \end{cases} \tag{48}$$

where the switching boundaries are given by Eq. (7) and the constants $\{k_n\}$ are defined in Eq. (4). According to Bick (1990), a necessary condition for an equilibrium is that $S(t)\xi(t)$ and $\mathrm{e}^{-r(T-t)}\xi(t)$ are martingales, i.e., the Euler equations for the stock and the bond are both satisfied.

Letting $x(t)=S(t)g(t)$ and applying Itô’s lemma reveals that $\mathrm{d}x(t)=\lambda/2x(t)\mathrm{d}t+\sigma(x,t)x(t)\mathrm{d}z(t)$ where $\sigma(x,t)$ equals $\sqrt{\lambda\gamma_1^{-1}}$ for $0<x(t)\leq\eta_1$, equals $\sqrt{\lambda\gamma_2^{-1}}$ for $\eta_1<x(t)\leq\eta_2$, etc. In terms of the variable $x(t)$, the quantity $S(t)\xi(t)$ can be written as

$$f(x(t))\equiv 1_{(0,\eta_1]}\left[x(t)\right]^{1-\gamma_1} + 1_{(\eta_1,\eta_2]}k_2\left[x(t)\right]^{1-\gamma_2} + \cdots + k_{N+1}1_{(\eta_N,\infty)}\left[x(t)\right]^{1-\gamma_{N+1}} \tag{49}$$

where $1_{(a,b]}$ is the indicator function for the interval $a<x(t)\leq b$. Note that Eq. (49) can always be expressed as a linear combination of convex functions of $x(t)$. Hence, even

[19] The path independence of both the pricing kernel and the representative agent’s optimal policy is discussed in more detail in Cox and Leland (1982) and He and Leland (1993).

though the first derivative of Eq. (49) is discontinuous at the switching boundaries, a generalized version of Itô's lemma can be applied as long as we interpret the second derivative of Eq. (49) in the sense of a distribution (Karatzas and Shreve, 1991, pp. 218–219). In this case, Itô's lemma is

$$f(x(t)) = f(x(0)) + \int_0^t D^- f(x(s))\mathrm{d}x(s) + \int_{-\infty}^{+\infty} \Lambda(t;a)m(\mathrm{d}a) \tag{50}$$

where $D^-f$ is the left hand derivative of $f$, $m$ is the second derivative measure given by $m([a,b)) \equiv D^-f(b) - D^-f(a)$, and $\Lambda(t;a)$ is the local time of $x(t)$ at the point $a$. Using Theorem 7.1(iii) from Karatzas and Shreve (1991, p. 218) and noting that the first derivative of $f$ has a jump at $x(t) = \eta_n$ whose size is proportional to $(\gamma_n - \gamma_{n+1})$, the right hand side of expression (50) reduces to

$$\begin{aligned} &f(x(0)) + \int_0^t D^- f(x(s))\mathrm{d}x(s) + \sum_{n=1}^{N} \left[\gamma_n - \gamma_{n+1}\right] k_n(\eta_n)^{-\gamma_n} \Lambda(t;\eta_n) \\ &\quad + \frac{\lambda}{2}\int_0^t \left[1_{(0,\eta_1)}(\gamma_1 - 1)\left[x(s)\right]^{1-\gamma_1} + \cdots + k_{N+1} 1_{(\eta_N,\infty)}(\gamma_{N+1} - 1)\left[x(s)\right]^{1-\gamma_{N+1}}\right]\mathrm{d}s \end{aligned}$$

From Karatzas and Shreve (1991, p. 223) note that

$$\frac{\lambda}{2}\int_0^t 1_{\{x(s)=\eta_n\}} x(s)\mathrm{d}s = \Lambda(t;\eta_n) - \Lambda(t;\eta_n -)$$

Hence, the right hand side of Eq. (50) further simplifies to

$$\begin{aligned} &f(x(0)) + \frac{\lambda}{2}\int_0^t f(x(s))\mathrm{d}s + \int_0^t D^- f(x(s))\sigma(x,s)x(s)\mathrm{d}z(s) - \frac{\lambda}{2}\int_0^t \left[1_{(0,\eta_1)}\gamma_1 [x(s)]^{1-\gamma_1}\right. \\ &\quad \left. + \cdots + 1_{(\eta_N,\infty)} k_{N+1}\gamma_{N+1}[x(s)]^{1-\gamma_{N+1}}\right]\mathrm{d}s + \frac{\lambda}{2}\int_0^t \left[1_{(0,\eta_1)}(\gamma_1 - 1)[x(s)]^{1-\gamma_1}\right. \\ &\quad \left. + \cdots + k_{N+1} 1_{(\eta_N,\infty)}(\gamma_{N+1} - 1)[x(s)]^{1-\gamma_{N+1}}\right]\mathrm{d}s + \sum_{n=1}^{N}[\gamma_n - \gamma_{n+1}]k_n(\eta_n)^{-\gamma_n}\Lambda(t;\eta_n) \\ &\quad - \sum_{n=1}^{N}\gamma_n k_n(\eta_n)^{-\gamma_n}[\Lambda(t;\eta_n) - \Lambda(t;\eta_n -)] \end{aligned} \tag{51}$$

Using the expressions in Protter (1990, p. 178) to directly evaluate $\Lambda(t;\eta_n)$ and $\Lambda(t;\eta_n -)$ reveals that

$$\Lambda(t;\eta_n) = \frac{\lambda}{\gamma_{n+1}}\int_0^t \delta(x(s) - \eta_n)x^2(s)\mathrm{d}s \tag{52}$$

$$\Lambda(t;\eta_n-)=\frac{\lambda}{\gamma_n}\int_0^t\delta(x(s)-\eta_n)x^2(s)\mathrm{d}s \tag{53}$$

where $\delta(\cdot)$ is Dirac's delta function. After substituting Eqs. (52)–(53) into Eq. (51) and canceling appropriate terms, the expression for $f(x(t))$ reduces to

$$\begin{aligned}f(x(t))=&f(x(0))+\int_0^t D^{-}f(x(s))\sigma(x,s)x(s)\mathrm{d}z(s)+\frac{\lambda}{2}\int_0^t f(x(s))\mathrm{d}s\\&-\frac{\lambda}{2}\int_0^t\left[1_{(0,\eta_1)}[x(s)]^{1-\gamma_1}+\cdots+k_{N+1}1_{(\eta_N,\infty)}[x(s)]^{1-\gamma_{N+1}}\right]\mathrm{d}s\end{aligned} \tag{54}$$

Taking expectations of both sides of Eq. (54) produces the required result, i.e., $S(t)\xi(t)$ is a martingale. To show that $\mathrm{e}^{-r(T-t)}\xi(t)$ is a martingale, apply Itô's lemma to $h(x(t))\equiv f(x(t))(x(t))^{-1}\mathrm{e}^{\lambda/2(T-t)}$, integrate from 0 to $t$, and take expectations of both sides. Hence, the Euler equations for the market portfolio and the riskless bond are both satisfied.[20]

Next, the representative agent's dynamic investment problem is recast into a static utility optimization problem of the following form:

$$\max_{W(T)} E_0[U(W(T))] \tag{55}$$

subject to the budget constraint

$$E_0\left[\frac{\xi(T)}{\xi(0)}W(T)\right]\leq S(0) \tag{56}$$

Standard optimization techniques reveal that the optimal wealth, $\hat{W}(T)$, is given by

$$\hat{W}(T)=U'^{-1}\left(\mu\frac{\xi(T)}{\xi(0)}\right) \tag{57}$$

where $\mu$, the multiplier for the budget constraint, satisfies the equality

$$E_0\left[\frac{\xi(T)}{\xi(0)}U'^{-1}\left(\mu\frac{\xi(T)}{\xi(0)}\right)\right]=S(0) \tag{58}$$

It is straightforward to verify that $\mu=\xi(0)$ and $\hat{W}(T)=S(T)$. From the Euler equation, this optimal consumption pattern is financed by holding 1 unit of the market and none of the bond at each date $t$. Lastly, the agent's indirect utility function, $E_t[U(S(T))]$, is calculated.

[20] Alternatively, the Euler equations can be directly verified by relying on the Markov property of $S(t)$. For example, within the interval $b_n(t)<S(t)<b_{n+1}(t)$, the expectation $E_t[\xi(T)S(T)]$ can be decomposed into three parts: (i) the expected value given that neither boundary is hit before $T$; (ii) the expected value given that the upper boundary is hit before the lower boundary; and (iii) the expected value given that the lower boundary is hit before the upper boundary. A similar decomposition can be applied to $E_t[\xi(T)]$. The densities for these types of calculations are known and are used for valuing double barrier options (see Pelsser, 2000).

Letting $E_t[U(S(T))]=J(S(t),t)$ and analyzing the interval $b_n(t)<S(t)\leq b_{n+1}(t)$, the function $J(S(t),t)$ satisfies the partial differential equation

$$J_t+(r+\lambda)SJ_S+\frac{1}{2}S^2\frac{\lambda}{\gamma_{n+1}}J_{SS}=0 \tag{59}$$

subject to the conditions that $J$ and $J_S$ are continuous at the switching boundaries. This latter condition is Samuelson's (1965) "high contact" condition for optimality. The solution for the interval $b_n(t)<S(t)\leq b_{n+1}(t)$ is

$$J(S(t),t)=k_{n+1}\frac{(S(t))^{1-\gamma_{n+1}}}{1-\gamma_{n+1}}(g(t))^{1-\gamma_{n+1}}+d_{n+1} \tag{60}$$

where the constant $d_{n+1}$ is given by Eq. (5). It can be verified directly that $J_S$ coincides with $\xi(t)$ in Eq. (48). This completes the proof. □

**Proof of Proposition 2.** First, recall that the bond is taken as numeraire so that the interest rate is zero. Next, define the function $g(t)$ as

$$g(t)=\mathrm{e}^{-\frac{1}{2}\lambda(T-t)} \tag{61}$$

and introduce the pricing kernel $\xi(t)$ defined as

$$\xi(t)=\begin{cases} \mathrm{e}^{-\gamma_1 S(t)}(g(t))^{\gamma_1}, & S(t)\leq b_1(t) \\ k_2\mathrm{e}^{-\gamma_2 S(t)}(g(t))^{\gamma_2}, & b_1(t)<S(t)\leq b_2(t) \\ \vdots & \vdots \\ k_{N+1}\mathrm{e}^{-\gamma_{N+1}S(t)}(g(t))^{\gamma_{N+1}}, & b_N(t)<S(t) \end{cases} \tag{62}$$

where the switching boundaries are from Eq. (20) and the constants $\{k_n\}$ from Eq. (16).

Letting $x(t)=S(t)+\lambda/2(T-t)$ and applying Itô's lemma reveals that $\mathrm{d}x(t)=\lambda/2\mathrm{d}t+\sigma(x,t)\mathrm{d}z(t)$, where $\sigma(x,t)$ equals $\sqrt{\lambda\gamma_1^{-1}}$ for $x(t)\leq\eta_1$, equals $\sqrt{\lambda\gamma_2^{-1}}$ for $\eta_1<x(t)\leq\eta_2$, etc. In terms of the variable $x(t)$, $\xi(t)$ in Eq. (62) can be written as

$$f(x(t))\equiv 1_{(-\infty,\eta_1]}\mathrm{e}^{-\gamma_1 x(t)}+1_{(\eta_1,\eta_2]}k_2\mathrm{e}^{-\gamma_2 x(t)}+\cdots+k_{N+1}1_{(\eta_N,\infty)}\mathrm{e}^{-\gamma_{N+1}x(t)} \tag{63}$$

Noting that the first derivative of $f$ has a jump at $x(t)=\eta_n$ whose size is proportional to $(\gamma_n-\gamma_{n+1})$, the generalized version of Itô's lemma in Eq. (50) reduces to

$$\begin{aligned} f(x(t))=&f(x(0))+\int_0^t D^-f(x(s))\mathrm{d}x(s)+\sum_{n=1}^{N}[\gamma_n-\gamma_{n+1}]k_n\mathrm{e}^{-\gamma_n\eta_n}\Lambda(t;\eta_n) \\ &+\frac{\lambda}{2}\int_0^t[\gamma_1 1_{(0,\eta_1)}\mathrm{e}^{-\gamma_1 x(s)}+\cdots+k_{N+1}\gamma_{N+1}1_{(\eta_N,\infty)}\mathrm{e}^{-\gamma_{N+1}x(s)}]\mathrm{d}s \end{aligned}$$

From Karatzas and Shreve (1991, p. 223), note that

$$\frac{\lambda}{2}\int_0^t 1_{\{x(s)=\eta_n\}}\mathrm{d}s=\Lambda(t;\eta_n)-\Lambda(t;\eta_n-)$$

Hence, the expression for $f(x(t))$ further simplifies to

$$
\begin{aligned}
& f(x(0)) + \int_0^t D^- f(x(s))\sigma(x,s)\mathrm{d}z(s) + \sum_{n=1}^{N} [\gamma_n - \gamma_{n+1}] k_n \mathrm{e}^{-\gamma_n \eta_n} \Lambda(t;\eta_n) \\
& \quad - \sum_{n=1}^{N} \gamma_n k_n \mathrm{e}^{-\gamma_n \eta_n} [\Lambda(t;\eta_n) - \Lambda(t;\eta_n -)] - \frac{\lambda}{2} \int_0^t \left[\gamma_1 1_{(0,\eta_1)} \mathrm{e}^{-\gamma_1 x(s)}\right. \\
& \quad \left. + \cdots + k_{N+1}\gamma_{N+1} 1_{(\eta_N,\infty)} \mathrm{e}^{-\gamma_{N+1} x(s)}\right] \mathrm{d}s + \frac{\lambda}{2} \int_0^t \left[\gamma_1 1_{(0,\eta_1)} \mathrm{e}^{-\gamma_1 x(s)}\right. \\
& \quad \left. + \cdots + k_{N+1}\gamma_{N+1} 1_{(\eta_N,\infty)} \mathrm{e}^{-\gamma_{N+1} x(s)}\right] \mathrm{d}s
\end{aligned} \tag{64}
$$

where the last two terms exactly cancel. Using the expressions in Protter (1990, p. 178) to directly evaluate $\Lambda(t;\eta_n)$ and $\Lambda(t;\eta_n-)$ reveals that

$$
\Lambda(t;\eta_n) = \frac{\lambda}{\gamma_{n+1}} \int_0^t \delta(x(s) - \eta_n)\mathrm{d}s \tag{65}
$$

$$
\Lambda(t;\eta_n -) = \frac{\lambda}{\gamma_n} \int_0^t \delta(x(s) - \eta_n)\mathrm{d}s \tag{66}
$$

where $\delta(\cdot)$ is Dirac's delta function. Substituting Eqs. (65)–(66) into Eq. (64) and canceling appropriate terms gives

$$
f(x(t)) = f(x(0)) + \int_0^t D^- f(x(s))\sigma(x,s)\mathrm{d}z(s) \tag{67}
$$

Taking expectations of both sides of Eq. (67) produces the required result, i.e., $\xi(t)$ is a martingale. To show that $S(t)\xi(t)$ is a martingale, apply Itô's lemma to $h(x(t)) \equiv f(x(t))[x(t) - \lambda/2(T-t)]$, integrate from 0 to $t$, and take expectations of both sides. Note that the drift of $\mathrm{d}h$ is identically zero in this case.

The agent's optimization problem is described by expressions (55)–(56) and again (as in the Proof of Proposition 1) it can be verified that $\mu = \xi(0)$ and $\hat{W}(T) = S(T)$. Hence, the agent optimally consumes the value of the market portfolio at date $T$ and this consumption is financed by holding 1 unit of the market and none of the bond at each date $t$. Lastly, for the interval $b_n(t) < S(t) \le b_{n+1}(t)$, the agent's indirect utility function $J(S(t),t)$ satisfies the partial differential equation

$$
J_t + \lambda J_S + \frac{1}{2} \frac{\lambda}{\gamma_{n+1}} J_{SS} = 0 \tag{68}
$$

where $J$ and $J_S$ are continuous at the switching boundaries. The solution for the interval $b_n(t)<S(t)\leq b_{n+1}(t)$ is

$$J(S(t),t) = -\frac{k_{n+1}}{\gamma_{n+1}} \mathrm{e}^{-\gamma_{n+1}S(t)}(g(t))^{\gamma_{n+1}} + d_{n+1} \tag{69}$$

where the constant $d_{n+1}$ is given by Eq. (17). It can be verified directly that $J_S$ coincides with $\xi(t)$ in Eq. (62). This completes the proof. □

## References


Aït-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica 70, 223–262.

Aït-Sahalia, Y., Lo, A., 2000. Nonparametric risk management and implied risk aversion. Journal of Econometrics 94, 9–51.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

Akaike, H., 1981. Likelihood of a model and information criteria. Journal of Econometrics 16, 3–14.

Ang, A., Bekaert, G., 2002a. International asset allocation with regime shifts. Review of Financial Studies 15, 1137–1187.

Ang, A., Bekaert, G., 2002b. Regime switches in interest rates. Journal of Business and Economic Statistics 20, 163–182.

Ang, J., Schwarz, T., 1985. Risk aversion and information structure: an experimental study of price variability in the securities markets. Journal of Finance 40, 825–844.

Bekaert, G., Wu, G., 2000. Asymmetric volatility and risk in equity markets. Review of Financial Studies 13, 1–42.

Benninga, S., Mayshar, J., 2000. Heterogeneity and option pricing. Review of Derivatives Research 4, 7–27.

Berk, J., 1995. A critique of size-related anomalies. Review of Financial Studies 8, 275–286.

Bick, A., 1987. On the consistency of the Black-Scholes model with a general equilibrium framework. Journal of Financial and Quantitative Analysis 22, 259–275.

Bick, A., 1990. On viable diffusion price processes of the market portfolio. Journal of Finance 45, 673–689.

Black, F., 1976. Studies of stock price volatility changes. Proceedings of the American Statistical Association, Business and Economics Statistics Section Meetings, pp. 177–181.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–659.

Bliss, R., Panigirtzoglou, N., 2004. Option-implied risk aversion estimates. Journal of Finance 59, 407–446.

Bollen, N., 1998. Valuing options in regime switching models. Journal of Derivatives 6, 38–49.

Bollen, N., Gray, S., Whaley, R., 2000. Regime switching in foreign exchange rates: evidence from currency option prices. Journal of Econometrics 94, 239–276.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., Chou, R., Kroner, K., 1992. ARCH modeling in finance. Journal of Econometrics 52, 5–59.

Brandt, M., Wu, T., 2002. Cross-sectional tests of deterministic volatility functions. Journal of Empirical Finance 9, 525–550.

Breeden, D., Litzenberger, R., 1978. Prices of state-contingent claims implicit in option prices. Journal of Business 51, 621–651.

Brennan, M., 1979. The pricing of contingent claims in discrete time models. Journal of Finance 34, 53–68.

Buraschi, A., Jackwerth, J., 2001. The price of a smile: hedging and spanning in option markets. Review of Financial Studies 14, 495–527.

Campbell, J., Cochrane, J., 1999. By force of habit: a consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205–251.

Campbell, J., Hentschel, L., 1992. No news is good news: an asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31, 281–318.

Campbell, S., Li, C., 2003. Option Pricing in Regime-Switching Environments. Working paper, Graduate School of Management, University of California at Riverside.

Campbell, J., Lo, A., MacKinlay, C., 1997. The Econometrics of Financial Markets. Princeton University Press, New Jersey.

Chan, Y., Kogan, L., 2002. Catching up with the Joneses: heterogeneous preferences and the dynamics of asset prices. Journal of Political Economy 110, 1255–1285.

Chapman, D., 1997. Approximating the asset pricing kernel. Journal of Finance 52, 1383–1410.

Chauvet, M., Potter, S., 2000. Coincident and leading indicators of the stock market. Journal of Empirical Finance 7, 87–111.

Christie, A., 1982. The stochastic behavior of common stock variances—value, leverage, and interest rate effects. Journal of Financial Economics 10, 407–432.

Cochrane, J., 1996. A cross-sectional test of an investment-based asset pricing model. Journal of Political Economy 104, 572–621.

Constantinides, G., 1982. Intertemporal asset pricing with heterogeneous consumers and without demand aggregation. Journal of Business 55, 253–267.

Cox, J., Huang, C., 1989. Optimal consumption and portfolio processes when asset prices follow a diffusion process. Journal of Economic Theory 49, 33–83.

Cox, J., Leland, H., 1982. On dynamic investment strategies. Proceedings of the Seminar on the Analysis of Security Prices, Center for Research in Security Prices, Graduate School of Business, University of Chicago, pp. 139–173.

Diebold, F., 1986. Modeling the persistence of conditional variances: a comment. Econometric Reviews 5, 51–56.

Diebold, F., Lee, J., Weinbach, G., 1994. Regime switching with time-varying transition probabilities. In: Hargreaves, C. (Ed.), Nonstationary Time Series Analysis and Cointegration, Chap. 10. Oxford University Press, Oxford.

Duan, J., Popova, I., Ritchken, P., 2002. Option pricing under regime switching. Quantitative Finance 2, 116–132.

Duffee, G., 1995. Stock returns and volatility: a firm-level analysis. Journal of Financial Economics 37, 399–420.

Duffie, D., Singleton, K., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61, 929–952.

Dumas, B., Fleming, J., Whaley, R., 1998. Implied volatility functions: empirical tests. Journal of Finance 53, 2059–2106.

Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. Econometrica 50, 987–1006.

Engle, R., Mustafa, C., 1992. Implied ARCH models from option prices. Journal of Econometrics 52, 289–311.

Ferson, W., Foerster, S., 1994. Finite sample properties of the generalized method of moments in tests of conditional asset pricing models. Journal of Financial Economics 36, 29–55.

French, M., Sichel, D., 1993. Cyclical patterns in the variance of economic activity. Journal of Business and Economic Statistics 11, 113–119.

French, K., Schwert, G., Stambaugh, R., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Friedman, M., Savage, L., 1948. The utility analysis of choices involving risk. Journal of Political Economy 56, 279–304.

Gaunersdorfer, A., Hommes, C., 2002. A Nonlinear Structural Model for Volatility Clustering. Working paper, Center for Nonlinear Dynamics in Economics and Finance, University of Amsterdam.

Genton, M., Ronchetti, E., 2003. Robust indirect inference. Journal of the American Statistical Association 98, 67–76.

Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, S85–S118.

Gray, S., 1996. Modeling the conditional distribution of interest rates as a regime switching process. Journal of Financial Economics 42, 27–62.
Hamilton, J., 1990. Analysis of time series subject to changes in regime. Journal of Econometrics 45, 39–70.
Hamilton, J., 1994. Time Series Analysis. Princeton University Press, Princeton, New Jersey.
Hamilton, J., Lin, G., 1996. Stock market volatility and the business cycle. Journal of Applied Econometrics 11, 573–593.
Hamilton, J., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 64, 307–333.
Hansen, L., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.
Hansen, B., 1992. The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. Journal of Applied Econometrics 7, S61–S82.
Hansen, B., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.
Hansen, L., Jagannathan, R., 1997. Assessing specification errors in stochastic discount factor models. Journal of Finance 52, 557–590.
Hansen, L., Heaton, J., Yaron, A., 1996. Finite sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14, 262–280.
He, H., Leland, H., 1993. On equilibrium asset price processes. Review of Financial Studies 6, 593–617.
Hull, J., White, A., 1990. Valuing derivative securities using the explicit finite difference method. Journal of Financial and Quantitative Analysis 25, 87–100.
Jackwerth, J., 2000. Recovering risk aversion from option prices and realized returns. Review of Financial Studies 13, 433–451.
Jacobsen, B., Dannenburg, D., 2003. Volatility clustering in monthly stock returns. Journal of Empirical Finance 10, 479–503.
Karatzas, I., Shreve, S., 1991. Brownian Motion and Stochastic Calculus, 2nd Edition. Springer-Verlag, New York.
Lamoureux, C., Lastrapes, W., 1990. Persistence in variance, structural change, and the GARCH model. Journal of Business and Economic Statistics 8, 225–234.
Lee, J., 1991. A Lagrange multiplier test for GARCH models. Economics Letters 37, 265–271.
Levy, H., 1994. Absolute and relative risk aversion: an experimental study. Journal of Risk and Uncertainty 8, 289–307.
Lo, A., 1988. Maximum likelihood estimation of generalized Itô processes with discretely sampled data. Econometric Theory 4, 231–247.
Lux, T., Marchesi, M., 2000. Volatility clustering in financial markets: a microsimulation of interacting agents. International Journal of Theoretical and Applied Finance 3, 675–702.
Mehra, R., Prescott, E., 1985. The equity premium: a puzzle. Journal of Monetary Economics 15, 145–161.
Nakao, S., 1972. On the pathwise uniqueness of solutions of one-dimensional stochastic differential equations. Osaka Journal of Mathematics 9, 513–518.
Ogden, J., 1987. An analysis of yield curve notes. Journal of Finance 42, 99–110.
Pagan, A., Schwert, G., 1990. Alternative models for conditional stock volatility. Journal of Econometrics 45, 267–290.
Palm, F., 1996. GARCH models of volatility. In: Maddala, G., Rao, C. (Eds.), Handbook of Statistics, vol. 14. Elsevier, New York, pp. 209–240.
Pelsser, A., 2000. Pricing double barrier options using Laplace transforms. Finance and Stochastics 4, 95–104.
Perez-Quiros, G., Timmermann, A., 2001. Business cycle asymmetries in stock returns: evidence from higher order moments and conditional densities. Journal of Econometrics 103, 259–306.
Pindyck, R., 1984. Risk, inflation, and the stock market. American Economic Review 74, 334–351.
Poterba, J., Summers, L., 1986. The persistence of volatility and stock market fluctuations. American Economic Review 76, 1142–1151.
Protter, P., 1990. Stochastic Integration and Differential Equations. Springer-Verlag, New York.
Rosenberg, J., Engle, R., 2002. Empirical pricing kernels. Journal of Financial Economics 64, 341–372.
Rubinstein, M., 1976. The valuation of uncertain income streams and the pricing of options. Bell Journal of Economics 7, 407–425.

Samuelson, P., 1965. Rational theory of warrant pricing. Industrial Management Review 6, 3–31.
Schaller, H., van Norden, S., 1997. Regime switching in stock market returns. Applied Financial Economics 7, 177–191.
Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.
Schwert, G., 1989. Why does stock market volatility change over time? Journal of Finance 44, 1115–1153.
Stapleton, R., Subrahmanyam, M., 1984a. The valuation of multivariate contingent claims in discrete time models. Journal of Finance 39, 207–228.
Stapleton, R., Subrahmanyam, M., 1984b. The valuation of options when asset returns are generated by a binomial process. Journal of Finance 39, 1525–1539.
Turner, C., Startz, R., Nelson, C., 1989. A Markov model of heteroskedasticity, risk, and learning in the stock market. Journal of Financial Economics 25, 3–22.
Wu, G., 2001. The determinants of asymmetric volatility. Review of Financial Studies 14, 837–859.