Option Pricing with High-Frequency-Based Affine Stochastic Volatility (HEAVY-SV) Models*

Hyojin Han[†]

November 10, 2016

Abstract

One common observation from financial data is a negative relationship between returns and volatility known as a leverage effect. This paper develops an affine discrete-time option pricing model accommodating the leverage effect. We use the general framework of discrete-time affine models by Darolles et al. (2006) for modeling a bivariate process of returns and stochastic volatility (SV). This allows us to share the same advantage of analytical tractability as affine continuous-time models, that is closed-form option pricing formulas, while discrete-time modeling provides a better empirical fit of skewness and kurtosis of asset returns. We exploit information in high-frequency data as summarized by realized variance (RV) which produces dynamics of RV as a SV-type extension of traditional high-frequency-based volatility (HEAVY) models by Shephard and Sheppard (2010) that are of the GARCH type. Such a SV-type extension allows the model to be robust under temporal aggregation as well as to provide option pricing formulas that are homogeneous of degree one with respect to the underlying stock and strike prices. Moreover, the leverage effect is characterized through a time invariant correlation coefficient between returns and SV that is similar to popular modeling in continuous-time models. This enables us to see how the leverage effect affects volatility smile. We use GMM with a large number of moment conditions for estimation and we can provide analytical identification conditions thanks to the affine structure. An empirical illustration is provided with the returns and options data of the S&P500 index.

*I am very grateful to my advisor Eric Renault for his guidance, constant support and encouragement as well as to Adam McCloskey and Susanne Schennach for helpful comments and discussions. I also thank Stanislav Khrapov for his help at several stages of this work. First, the asset pricing model part of this paper owes largely to Khrapov and Renault (2016). Second, the empirical part is the outcome of some tight collaboration.

[†]PhD candidate, Department of Economics, Brown University, hyojin_han@brown.edu

1 Introduction

Since Duffie, Pan, and Singleton (2000), Affine Jump-Diffusion (AJD) models have been popular for derivative pricing. AJD models are continuous-time models and nest in particular the Cox, Ingersoll, and Ross (1985) model for interest rates and the affine stochastic-volatility model for currency and equity returns proposed by Heston (1993) for the case of option pricing. The main advantage of AJD models is their analytical tractability. Duffie et al. (2000) show the existence of semi-closed form expressions for derivative prices through the use of Fourier transform.

While AJD models have their advantages, discrete-time models have attracted considerable attention in derivative pricing as well. Skewness and excess kurtosis of asset returns are commonly observed in financial time-series. However, up to jumps, the AJD models' ability to reproduce higher order moments is limited by the maintained assumption of conditional normality. This is the reason why discrete-time models provide an additional degree of freedom that leads to a better empirical fit of such higher order moments. The fact that only discrete observations are often available for empirical study has also promoted the development of discrete-time modeling. One such class of discrete-time models is GARCH option pricing models that give closed-form pricing formulas (Duan (1995), Heston and Nandi (2000), Christoffersen et al. (2010, 2012)). Another strand of literature that is of interest to this paper concerns a class of discrete-time affine stochastic volatility (SV) models (Darolles et al. (2006)).

This paper develops an affine discrete-time SV option pricing model incorporating the leverage effect, a negative relationship between returns and volatility. We use the general framework of compound autoregressive processes (CAR) of Darolles et al. (2006) for modeling a bivariate process of returns and volatility. With an exponentially affine stochastic discount factor, the CAR framework has the structure-preserving property, meaning that the historical and risk neutral measures share the same dynamics with possibly different parameters. This provides a convenient framework for option pricing since the affine property is maintained under the risk-neutral dynamics which enables us to compute option prices in closed-form.

The stochastic volatility factor is unobservable. We exploit the information in high-frequency data as summarized by "realized variance", which is constructed from intraday price movements, for identification of the latent volatility factor. The model developed in this paper provides dynamics of realized variance that are closely related to linear high-frequency-based volatility (HEAVY) models by Shephard and Sheppard (2010)[1]. For this reason, we dub this model a "HEAVY-SV" model.

The HEAVY-SV model enhances traditional HEAVY models that are of the GARCH type. The HEAVY models that Shephard and Sheppard (2010) focus on have AR(1) dynamics of the conditional

[1]HEAVY models are predictive models of daily asset returns volatility based on realized measures constructed from high-frequency data. Realized variance is one particular example of realized measures.

mean of realized variance that is of the GARCH(1,1) type. We use a general AR(1) model for the conditional mean which encompasses the HEAVY models. More specifically, we add another source of randomness to the HEAVY models by considering SV. In the same way that Meddahi and Renault (2004) have pointed out that affine SV models are an extension of GARCH models, the HEAVY-SV model can be seen as a SV-type extension of the HEAVY models. Due to this extension, the HEAVY-SV model is robust under temporal aggregation[2], which underpins the continuous-time limit of this model.

In addition, with the assumption of no Granger causality from returns to volatility, we get option pricing formulas that are homogeneous of degree one (as in the Black and Scholes case) with respect to the underlying stock and strike prices. This homogeneity property is desirable because it ensures that option prices are convex with respect to underlying asset prices, which is consistent with actual data (Garcia an Renault (1998)). This homogeneity property also allows the volatility smile to be a function of moneyness [3] only. Another weakness of GARCH option pricing models is the absence of such a homogeneity property (see Garcia and Renault (1998)).

A well-documented feature of financial time series is the leverage effect: there tends to be a negative relationship between returns and volatility. Typically, rising asset prices are accompanied by declining volatility, and vice versa. In continuous-time models, the leverage effect is characterized through an instantaneous correlation coefficient (e.g. Heston (1993)). We introduce the leverage effect as a time-invariant correlation coefficient between the contemporary returns and volatility as in continuous-time models. This allows us to see how the shape of volatility smile is affected by the leverage effect (see Khrapov and Renault (2016)[4] for details). However, as discussed by Bollerslev et al. (2006), it is hard to distinguish the leverage effect from volatility feedback effect in discrete-time. The fundamental difference between those two effects lies in the direction of causality[5] between volatility and returns but the discrete time approach complicates the separate identification of those causalities (see Renault et al. (1998)). While Bollerslev et al. (2006) enhance the usefulness of high-frequency data to separate a leverage effect and a volatility feedback effect, we impose a parameterization leaving room for both effects simultaneously.

This paper also addresses a practical identification issue that can arise with the application of the HEAVY-SV model. CAR models characterize the dynamics of returns and volatility through conditional Laplace transforms. This provides us with a continuum set of closed-form conditional moment restrictions that identify parameters. One commonly used estimation method is to choose instruments that are functions of conditioning variables and construct unconditional moments from

[2]In general, we say that a model is closed under temporal aggregation if the model keeps the same structure for any data frequency.

[3]Moneyness is defined as the percentage $x_t = log(S_t/K)$ where S_t and K denote the underlying asset price at time t and the strike price of an option.

[4]They show explicitly how volatility smile is distorted with the leverage effect.

[5]The direction of causality for volatility feedback effect is from volatility to returns (Bollerslev et al. (2006)): An anticipated increase in volatility would lead to a decline in stock prices for higher future returns.

conditional moments with them. However, as discussed in Dominguez and Lobato (2004) and Hsu and Kuan (2011), the global identification assumption may fail in nonlinear models when instruments are chosen in an arbitrary way to induce unconditional moments even if it holds for conditional moments. We demonstrate this point with an example of an affine volatility model and choose an instrument that guarantees identification by following Carrasco et al. (2007) and Hsu and Kuan (2011).

Carrasco and Florens (2000, 2014) and Carrasco et al. (2007) promote GMM with a continuum of moments in order to exploit the full information (C-GMM) by showing that C-GMM can attain MLE efficiency asymptotically. In this paper, we use GMM with a discrete albeit infinite subset of moment conditions (D-GMM). This has less computational burden than C-GMM. Also there is no efficiency loss when a large number of moments is exploited. D-GMM and C-GMM achieve the same asymptotic efficiency when moment conditions are based on characteristic functions.

Another practical issue is that the latent volatility factor is unobservable but we instead observe a series of realized variance. By inverting the ARMA model of realized variance, we approximate the latent AR volatility with a finite series of realized variances. This allows us to treat a realized variance process with an ARMA representation as an approximated AR process (with more lags than the latent volatility factor), which is convenient for inference although it makes the model locally misspecified.

We incorporate the information contained in options data for estimation. They play a role in both identifying the risk prices and choosing the regularization parameter[6] for the construction of the weighting matrix in GMM.

The rest of the paper is organized as follows. Section 2 sets up a HEAVY-SV model while completing HEAVY models for the purpose of option pricing. Section 3 provides the risk neutral dynamics of returns and volatility and an option pricing formula. Section 4 discusses GMM estimation with a large number of moment conditions. Section 5 provides the empirical analysis with the observations of realized variance and delivers the option pricing performance of the model developed in this paper. Section 6 concludes. All figures and proofs are relegated to the appendix.

[6]Due to the singularity of the sample covariance matrix of moment conditions, we adopt the regularization method proposed by Carrasco and Florens (2000). See section 4.2 for more details.

2 Completing a HEAVY model for the purpose of option pricing

2.1 HEAVY-SV model

Following Shephard and Sheppard (2010), our analysis will be based on daily financial returns

$$r_1, r_2, ..., r_T$$

and the corresponding sequence of daily realized variances

$$RV_1, RV_2, ..., RV_T.$$

For sake of notational convenience, the daily returns will be seen as daily continuously compounded rates of returns on some index S_t in excess of the risk free rate $r_{f,t}$ in the period $[t, t+1]$:

$$r_{t+1} = \log\left(S_{t+1}/S_t\right) - r_{f,t}.$$

Note that RV_{t+1} is typically computed from intraday data on the index $S_\tau, \tau \in [t, t+1]$ by summing squared increments of $\log(S)$. The focus of interest of Shephard and Sheppard (2010) was to propose a predictive model of volatility "out of the intellectual insights of the ARCH literature" but offering to "bolster them with high-frequency information". For this purpose, their HEAVY model (see their equations (3)/(4)) was a bivariate GARCH-type model, specifying two dynamic linear equations about the conditional expectation of realized variance and the conditional variance of returns given high frequency information F_t available at time t. They actually interpret the conditional variance $h_t = Var[r_{t+1} \,|F_t]$ as a "close-to-close" conditional variance while the conditional expectation of RV_{t+1}, $\mu_t = E[RV_{t+1} \,|F_t]$ can be interpreted as an "open-to-close" conditional variance of returns.

2.1.1 Model specification for $\mu_t = E[RV_{t+1} \,|F_t]$

Shephard and Sheppard (2010) write down a GARCH(1,1) type equation:

$$\mu_t = \omega_R + \alpha_R RV_t + \beta_R \mu_{t-1}. \tag{2.1}$$

By considering conditional expectation given F_{t-1} on both sides of the equation, we immediately deduce some $AR(1)$ dynamics for $\mu_t = E[RV_{t+1} \,|F_t]$:

$$\mu_t = \omega_R + (\alpha_R + \beta_R)\mu_{t-1} + \eta_t, E[\eta_t \,|F_{t-1}] = 0. \tag{2.2}$$

However, it is important to note that model (2.1) implies a quite constrained $AR(1)$ in (2.2) since the innovation process η_t of μ_t is a linear function of observed RV_t. To put it differently, (2.1) implies that μ_t is a deterministic function of current and past realized variances $RV_\tau, \tau \leq t$. In the same way that Meddahi and Renault (2004) have pointed out that, within the general class of Stochastic Volatility models with $AR(1)$ dynamics for the conditional variance, GARCH(1,1) is a tightly constrained subclass (because it allows only for one dimension of uncertainty), we can similarly put forward a general $AR(1)$ model for $\mu_t = E[RV_{t+1} \,|F_t]$. This means that it encompasses the particular model (2.1) but allows more generally for two sources of uncertainty, one in the innovation η_t of the $AR(1)$ process μ_t, and another in the difference between future realized variance and its current expectation . In other words, the innovation process η_t of μ_t is not necessarily a function of observed RV_t. We end up with the following specification:

$$\begin{aligned} RV_{t+1} &= \mu_t + \nu_{t+1}, E[\nu_{t+1} \,|F_t] = 0 \\ \mu_t &= \omega_R + \gamma_R \mu_{t-1} + \eta_t, E[\eta_t \,|F_{t-1}] = 0. \end{aligned}$$

It is well known that this specification implies $ARMA(1,1)$ dynamics for RV_{t+1} :

$$\begin{aligned} RV_{t+1} - \nu_{t+1} &= \omega_R + \gamma_R \left(RV_t - \nu_t \right) + \eta_t \\ &\Rightarrow \; RV_{t+1} - \gamma_R RV_t = \omega_R - \gamma_R \nu_t + \eta_t + \nu_{t+1}, \\ E[-\gamma_R \nu_t + \eta_t + \nu_{t+1} \,|F_{t-1}] &= 0. \end{aligned}$$

2.1.2 Model specification for $h_t = Var[r_{t+1} \,|F_t]$

As extensively discussed in Andersen, Bollerslev, Diebold and Labys (2003) (see their corollary 1 on page 586), if one thinks that the price process is a continuous time arbitrage-free squared integrable process, random feedback effects from the intraday evolution of the system to the instantaneous mean can be neglected. Thus, realized variance should be an unbiased estimator of the returns variance conditional on past information:

$$Var[r_{t+1} \,|F_t] = E[RV_{t+1} \,|F_t] \,. \tag{2.3}$$

However, due to several kinds of friction including overnight effects, some authors (see e.g. Brownlees and Gallo (2010)) have proposed more generally to link the conditional variance of returns to an affine transform of the predicted realized variance:

$$h_t = \xi + \varkappa \mu_t. \tag{2.4}$$

It is worth knowing that, when Brownlees and Gallo (2010) test for significance of deviations from the null hypothesis $\xi = 0$ and/or $\varkappa = 1$,they do not find compelling evidence against the simple model (2.3). For the purpose of empirical work, I will basically discuss normalizing assumptions allowing us to use the simplified version (2.3) of (2.4). Shephard and Sheppard (2010) use the more complicated model:

$$h_t = \omega + \alpha RV_t + \beta h_{t-1}.$$

However, they acknowledge that “although these models are distinct, they have quite a lot of common thinking in their structure”. Basically, the role of their “momentum parameter β” will be to compensate for the fact that they use the raw measure RV_t instead of its smoothed version μ_t. Note that we arguably need even less to introduce the raw measure RV_t in the equation for h_t that we have allowed for richer dynamics than Shephard and Sheppard (2010) for the smoothed realized variance μ_t. For us, it is of the SV type, that is endowed with an autonomous source of uncertainty. We will show later in section 2.3.2 how μ_t is defined in terms of SV factors. Note that by (2.4), the conditional variance h_t will be $AR(1)$ like μ_t, with the same persistence coefficient γ_R .

2.2 Affine specification for SV factor

For the purpose of option pricing, we will need a parametric model for the SV latent factor. This factor, denoted by σ_t^2, will be an $AR(1)$ process linked to the $AR(1)$ processes $\mu_t = h_t$ defined above by a deterministic affine function. It is only for sake of direct interpretation of the parameters that it is more convenient to specify a driving factor σ_t^2 common to both variance processes μ_t and h_t rather than the equivalent way of viewing one of these two processes as the driving factor. I use the general framework of compound autoregressive processes (CAR) put forward by Darolles et al. (2006) to describe the dynamics of σ_t^2. This framework is particularly well suited for option pricing since it characterizes the probability distribution through a conditional Laplace transform that is exponentially affine with respect to past state variables, and as such, provides a discrete time analog of the affine processes considered by Duffie et al. (2000). Hence I assume that:

$$E[\exp\left(-u\sigma_{t+1}^2\right)\left|\sigma_t^2\right] = \exp\left(-a(u)\sigma_t^2 - b(u)\right)$$

for some deterministic functions $a(.)$ and $b(.)$ defined for all complex numbers $u \in \mathbb{C}$.Then, the conditional expectation and variance can be computed as:

$$\begin{aligned} E[\sigma_{t+1}^2\left|\sigma_t^2\right] &= a'(0)\sigma_t^2 + b'(0) \\ Var[\sigma_{t+1}^2\left|\sigma_t^2\right] &= -a''(0)\sigma_t^2 - b''(0). \end{aligned}$$

Therefore, the computation of these two conditional moments involves the value of four parameters expected to fulfill the following inequality restrictions

$$\begin{aligned} 0 &\leq \rho = a'(0) < 1 \\ c &= -\frac{a''(0)}{2a'(0)} > 0 \\ \delta &= -2\frac{a'(0)b'(0)}{a''(0)} > 0 \\ \omega &= -4\frac{b''(0)[a'(0)]^2}{[a''(0)]^2} > 0 \end{aligned} \tag{2.5}$$

leading to the following affine model

$$\begin{aligned} E[\sigma_{t+1}^2 \left| \sigma_t^2 \right] &= \rho\sigma_t^2 + \delta c \\ Var[\sigma_{t+1}^2 \left| \sigma_t^2 \right] &= 2\rho c\sigma_t^2 + \omega c^2. \end{aligned}$$

Khrapov and Renault (2016) have shown that the continuous time limit of this model is the affine model:

$$d\sigma_t^2 = \kappa(\bar{\sigma}^2 - \sigma_t^2)dt + \sqrt{\nu + \eta\sigma_t^2}dW_t,$$

where the four continuous time parameters $(\kappa, \bar{\sigma}^2, \nu, \eta)$ are known one-to-one functions of the four discrete time parameters $(\rho, \delta, \omega, c)$. In particular, one gets the square root process of Feller (1951) by imposing $\nu = 0$, which is tantamount to $\delta = \omega$. The discrete time parametric model corresponding to the square root process is actually the so-called AutoRegressive-Gamma process (ARG(1)) proposed by Gourieroux and Jasiak (2006). In addition to the constraint $\delta = \omega$, it is characterized by the following parametric specification consistent with (2.5):

$$a(u) = \frac{\rho u}{1 + cu}, \qquad b(u) = \delta \log(1 + cu).$$

2.3 A bivariate CAR model for returns and SV factor

2.3.1 General framework

In order to define the joint probability distribution of the stochastic process $(r_t, \sigma_t^2), t = 1, ..., T$, we specify the conditional distribution of $(r_{t+1}, \sigma_{t+1}^2)$ given $I_t = \left\{(r_\tau, \sigma_\tau^2), \tau \leq t\right\}$ through the following conditional Laplace transform:

$$\begin{aligned} E[\exp\left(-u\sigma_{t+1}^2\right) |I_t] &= \exp\left(-a(u)\sigma_t^2 - b(u)\right), \forall u \in \mathbb{C} \\ E[\exp\left(-vr_{t+1}\right) \left| I_t, \sigma_{t+1}^2 \right] &= \exp\left(-\alpha(v)\sigma_{t+1}^2 - \beta(v)\sigma_t^2 - \gamma(v)\right), \forall v \in \mathbb{C}. \end{aligned}$$

These formulas define a bivariate $CAR(1)$, constrained by the two following maintained hypotheses:

(A1) The returns process r_t does not Granger cause the SV factor process.

(A2) Given the path of of the SV factor process σ_t^2, the consecutive returns are serially independent.

It is shown in Renault (1997) that the conjunction of conditions (A1) and (A2) is necessary and sufficient for the natural homogeneity (of degree one) property of the option pricing formula with respect to the pair (S, K) of underlying stock price and strike price. This homogeneity property, shared with Black and Scholes (BS) pricing, ensures that Black-Scholes (BS) implied volatilities depend only on the moneyness. This is in line with a common tradition of representing the volatility smile.

We will actually assume that the above conditional Laplace transforms are not modified when the information set I_t is augmented by the past of all relevant high frequency information leading to the information set F_t defined in subsection 2.1. Then,

$$\begin{aligned} E[\exp\left(-u\sigma_{t+1}^2\right)|F_t] &= \exp\left(-a(u)\sigma_t^2 - b(u)\right), \forall u \in \mathbb{C} \\ E[\exp\left(-vr_{t+1}\right)|F_t^\sigma] &= \exp\left(-\alpha(v)\sigma_{t+1}^2 - \beta(v)\sigma_t^2 - \gamma(v)\right), \forall v \in \mathbb{C}, \end{aligned} \tag{2.6}$$

where

$$F_t^\sigma = F_t \cup \left\{\sigma_{t+1}^2\right\}.$$

In particular,

$$\begin{aligned} E[r_{t+1}|F_t^\sigma] &= \alpha'(0)\sigma_{t+1}^2 + \beta'(0)\sigma_t^2 + \gamma'(0) \\ Var[r_{t+1}|F_t^\sigma] &= -\alpha''(0)\sigma_{t+1}^2 - \beta''(0)\sigma_t^2 - \gamma''(0). \end{aligned} \tag{2.7}$$

Note that, following the terminology of Darolles et al. (2006), the two equations (2.6) define a constrained bivariate $CAR(1)$ model

$$E[\exp\left(-u\sigma_{t+1}^2 - vr_{t+1}\right)|F_t] = \exp\left\{-l(u,v)\sigma_t^2 - g(u,v)\right\},$$

with the constraints

$$\begin{aligned} l(u,v) &= a[u+\alpha(v)] + \beta(v) \\ g(u,v) &= b[u+\alpha(v)] + \gamma(v). \end{aligned}$$

2.3.2 HEAVY-CAR model

From (2.7), we see that the conditional variance of returns $Var[r_{t+1}\,|F_t]$ involves four parameters in addition to the SV factor parameters defined in (2.5):

$$\begin{aligned} \alpha'(0) &= \psi < 0 \\ -\alpha''(0) &= 1-\phi^2 > 0 \\ e = \frac{\beta''(0)}{\rho\,(1-\phi^2)}, &\quad f = \frac{\gamma''(0)}{\delta c\,(1-\phi^2)} \end{aligned} \tag{2.8}$$

The constraints that ψ and ϕ must fulfill will be discussed shortly. We obtain the conditional variance as an affine function of the SV factor σ_t^2:

$$\begin{aligned} Var[r_{t+1}\,|F_t] &= \psi^2 Var[\sigma_{t+1}^2\,\big|\sigma_t^2] + (1-\phi^2)\left\{E[\sigma_{t+1}^2\,\big|\sigma_t^2] - e\rho\sigma_t^2 - f\delta c\right\} \\ &= \psi^2\left\{2\rho c\sigma_t^2 + \omega c^2\right\} + (1-\phi^2)\left\{\rho\sigma_t^2 + \delta c - e\rho\sigma_t^2 - f\delta c\right\} \\ &= \left[2c\psi^2 + (1-e)(1-\phi^2)\right]\rho\sigma_t^2 + c[\psi^2\omega c + (1-f)(1-\phi^2)\delta] \end{aligned}$$

The HEAVY relationship in (2.4) will then be fulfilled if and only $E[RV_{t+1}\,|F_t]$ is also an affine function of the SV factor σ_t^2, which in turns amounts to assuming that the realized variance is a linear transformation of the SV factor as follows

$$RV_{t+1} = A\sigma_{t+1}^2 - B\sigma_t^2 - D, \tag{2.9}$$

leading to

$$\begin{aligned} E[RV_{t+1}\,|F_t] &= A[\rho\sigma_t^2 + \delta c] - B\sigma_t^2 - D \\ &= [A\rho - B]\sigma_t^2 + [A\delta c - D] \end{aligned}$$

and

$$Var[RV_{t+1}\,|F_t] = 2A^2\rho c\sigma_t^2 + A^2\omega c^2$$

Note that the affine transformation in (2.9) implies that RV_{t+1} is an $ARMA(1,1)$ process, which is consistent with the general HEAVY model discussed above, while the SV factor σ_t^2 is $AR(1)$. Since the SV factor is latent, it takes some identification restrictions to identify all parameters. Imposing the natural condition (2.3) provides two equations about the three parameters A, B, D :

$$\begin{aligned} A\rho - B &= \rho\left[2c\psi^2 + (1-e)(1-\phi^2)\right] \\ A\delta c - D &= c[\psi^2\omega c + (1-f)(1-\phi^2)\delta]. \end{aligned}$$

However, since the SV factor is clearly identified only up to a scale factor, we can impose the normalization condition $A = 1$ and then conclude:

$$\begin{aligned} RV_{t+1} &= \sigma_{t+1}^2 - B\sigma_t^2 - D \\ B &= -\rho\left[2c\psi^2 + (1-e)(1-\phi^2) - 1\right] \\ D &= -c\left[\psi^2\omega c + (1-f)(1-\phi^2)\delta - \delta\right]. \end{aligned} \tag{2.10}$$

2.4 Leverage effect

The normalization condition allows us to characterize the leverage effect through the correlation coefficient

$$L_t = Corr[r_{t+1}, \sigma_{t+1}^2 \,|F_t] = Corr[r_{t+1}, RV_{t+1} \,|F_t].$$

However,

$$\begin{aligned} Cov[r_{t+1}, \sigma_{t+1}^2 \,|F_t] &= Cov[E[r_{t+1} \,|F_t^\sigma], \sigma_{t+1}^2 \,|F_t] \\ &= Cov[\psi\sigma_{t+1}^2, \sigma_{t+1}^2 \,|F_t] = \psi Var[RV_{t+1} \,|F_t] \end{aligned}$$

so that

$$L_t = \psi\left[\frac{Var[RV_{t+1} \,|F_t]}{Var[r_{t+1} \,|F_t]}\right]^{1/2} = \psi\left[\frac{Var[RV_{t+1} \,|F_t]}{E[RV_{t+1} \,|F_t]}\right]^{1/2}.$$

It is common to assume that the leverage effect is a time-invariant correlation coefficient. In our setting, this amounts to assuming that the ratio of $Var[RV_{t+1} \,|F_t]$ and $E[RV_{t+1} \,|F_t]$ is time-invariant. Our empirical study confirms that this assumption is sensible (see figure 1). We thus choose to maintain it, that is to assume that

$$\frac{Var[RV_{t+1} \,|F_t]}{E[RV_{t+1} \,|F_t]} = \frac{2\rho c}{\rho - B} = \frac{\omega c^2}{\delta c - D},$$

that is

$$\frac{Var[RV_{t+1} \,|F_t]}{E[RV_{t+1} \,|F_t]} = \frac{2c}{2c\psi^2 + (1-e)(1-\phi^2)} = \frac{\omega c}{\psi^2\omega c + (1-f)(1-\phi^2)\delta}$$

In other words, we have shown the following proposition.

Proposition 1

The leverage effect L_t is a time-invariant constant L if and only if

$$1 - f = \frac{\omega}{2\delta}(1-e),$$

in which case,

$$L = \psi \left[\frac{2c}{2c\psi^2 + (1-e)(1-\phi^2)} \right]^{1/2}.$$

3 Risk neutral HEAVY-CAR model and option pricing

3.1 Structure-preserving change of measure

Following Khrapov and Renault (2016) (see also Bertholon, Monfort and Pegoraro (2008)), we will define a risk-neutral probability distribution of a bivariate process of returns and volatility that belongs to the same family of constrained bivariate $CAR(1)$:

$$\begin{aligned} E^*[\exp\left(-u\sigma_{t+1}^2 - vr_{t+1}\right)|F_t] &= \exp\left\{-l^*(u,v)\sigma_t^2 - g^*(u,v)\right\} \\ &= \exp(r_{f,t})E[M_{t,t+1}(\varsigma)\exp\left(-u\sigma_{t+1}^2 - vr_{t+1}\right)|F_t], \end{aligned}$$

where $E^*[\cdot]$ denotes an expectation under a risk-neutral distribution, with

$$\begin{aligned} l^*(u,v) &= a^*[u+\alpha^*(v)] + \beta^*(v) \\ g^*(u,v) &= b^*[u+\alpha^*(v)] + \gamma^*(v) \end{aligned}$$

for some functions $a^*(.)$, $b^*(.)$, $\alpha^*(.)$, $\beta^*(.)$, $\gamma^*(.)$ to be characterized below. Here, $M_{t,t+1}(\varsigma)$ is an exponential affine pricing kernel:

$$M_{t,t+1}(\varsigma) = \exp(-r_{f,t})\exp\left\{m_0(\varsigma) + m_1(\varsigma)\sigma_t^2 - \varsigma_1\sigma_{t+1}^2 - \varsigma_2 r_{t+1}\right\},$$

where

(i) ς_1 and ς_2 are the two preference parameters corresponding to the two sources of risk. The parameter ς_1 is expected to be non-positive and characterizes the price of volatility risk while ς_2 is expected to be non-negative and characterizes the price of equity risk.

(ii) The functions $m_0(\varsigma)$ and $m_1(\varsigma)$, with $\varsigma = (\varsigma_1, \varsigma_2)$, are defined in order to match the following no-arbitrage condition

$$E\left[\exp\left\{m_0(\varsigma) + m_1(\varsigma)\sigma_t^2 - \varsigma_1\sigma_{t+1}^2 - \varsigma_2 r_{t+1}\right\}|F_t\right] = 1$$

which, by the law of iterated expectations, implies

$$\begin{aligned} m_0(\varsigma) &= \gamma(\varsigma_2) + b\left[\alpha(\varsigma_2) + \varsigma_1\right] = g(\varsigma_1, \varsigma_2) \\ m_1(\varsigma) &= \beta(\varsigma_2) + a\left[\alpha(\varsigma_2) + \varsigma_1\right] = l(\varsigma_1, \varsigma_2). \end{aligned} \tag{3.1}$$

The risk neutral dynamics defined by functions $a^*(\cdot)$, $b^*(\cdot)$, $\alpha^*(\cdot)$, $\beta^*(\cdot)$, and $\gamma^*(\cdot)$ are fully known when the historical dynamics and the risk price parameters ς_1 and ς_2 are known. Following Khrapov and Renault (2016), it is easy to check the following:

$$\begin{aligned} \alpha^*(v) &= \alpha(\varsigma_2 + v) - \alpha(\varsigma_2) \\ \beta^*(v) &= \beta(\varsigma_2 + v) - \beta(\varsigma_2) \\ \gamma^*(v) &= \gamma(\varsigma_2 + v) - \gamma(\varsigma_2), \end{aligned}$$

and

$$\begin{aligned} a^*(u) &= a\left[u + \varsigma_1 + \alpha(\varsigma_2)\right] - a\left[\varsigma_1 + \alpha(\varsigma_2)\right] \\ b^*(u) &= b\left[u + \varsigma_1 + \alpha(\varsigma_2)\right] - b\left[\varsigma_1 + \alpha(\varsigma_2)\right]. \end{aligned}$$

3.2 Identification of prices of risk

The following no-arbitrage condition contains the identifying information brought by observing underlying asset returns data:

$$E\left[M_{t,t+1}(\varsigma)\exp\left(r_{t+1}\right)|F_t\right] = 1.$$

This can be rewritten as

$$E\left[\exp\left\{-\varsigma_1\sigma_{t+1}^2 - (\varsigma_2 - 1)r_{t+1}\right\}|F_t\right] = \exp\left(r_{f,t}\right)\exp\left\{-m_0(\varsigma) - m_1(\varsigma)\sigma_t^2\right\}$$

with $m_0(\varsigma)$ and $m_1(\varsigma)$ defined in (3.1). This leads us to the following two equations about ς_1 and ς_2 by the law of iterative expectations[7]:

$$\begin{aligned} \beta(\varsigma_2 - 1) + a\left[\alpha(\varsigma_2 - 1) + \varsigma_1\right] &= \beta(\varsigma_2) + a\left[\alpha(\varsigma_2) + \varsigma_1\right] \\ \gamma(\varsigma_2 - 1) + b\left[\alpha(\varsigma_2 - 1) + \varsigma_1\right] &= \gamma(\varsigma_2) + b\left[\alpha(\varsigma_2) + \varsigma_1\right]. \end{aligned} \tag{3.2}$$

[7]We treat $\exp\left(-r_{f,t}\right) = 1$ since $r_{f,t}$ is the daily risk-free rate at time t.

One sufficient (not necessary) condition for the above equation (3.2) to hold is

$$\alpha(\varsigma_2) = \alpha(\varsigma_2 - 1), \qquad \beta(\varsigma_2) = \beta(\varsigma_2 - 1), \qquad \gamma(\varsigma_2) = \gamma(\varsigma_2 - 1) \tag{3.3}$$

Note that this identification scheme leaves the volatility risk price ς_1 completely unidentified from the observations of stock price. However, this scheme does not accomodate a non-zero leverage effect. In order to see this point, consider $\alpha(\cdot)$, $\beta(\cdot)$, and $\gamma(\cdot)$ that are well-approximated by their second order Taylor expansions in a neighborhood of zero[8]. Then

$$\begin{aligned} \alpha(\varsigma_2) - \alpha(\varsigma_2 - 1) &\approx \alpha'(0) + \alpha''(0)\left(\varsigma_2 - \frac{1}{2}\right) \\ &= \psi - (1-\phi^2)\left(\varsigma_2 - \frac{1}{2}\right) \end{aligned}$$

and thus,

$$\alpha(\varsigma_2) = \alpha(\varsigma_2 - 1) \quad \Leftrightarrow \quad \psi = (1-\phi^2)\left(\varsigma_2 - \frac{1}{2}\right). \tag{3.4}$$

We will now argue that this assumption on ψ is unacceptable when there is a non-zero leverage effect. To see that, we note that ψ measures the correlation between an update of the returns forecast and a shock in volatility, which follows directly from the below equation:

$$E\left[r_{t+1}|F_t^\sigma\right] - E\left[r_{t+1}|F_t\right] = \psi\left\{\sigma_{t+1}^2 - E\left[\sigma_{t+1}^2|F_t\right]\right\}.$$

As discussed in Bollerslev et al. (2006), this correlation consists of two different economic phenomena that are impossible to disentangle in discrete time:

i) The volatility feedback effect,

ii) The leverage effect.

The importance of the former effect is directly drawn by the size of the equity risk price ς_2, while the latter should be proportional to a leverage effect parameter. It turns out that ϕ, which by equations (2.7) and (2.8) characterizes the share of variance of returns that is explained by current volatility[9], should be interpreted as a genuine leverage effect parameter. Recall that from Proposition 1 we get

$$L = \psi\left[\frac{2c}{2c\psi^2 + (1-e)(1-\phi^2)}\right]^{1/2}.$$

[8] $\alpha(\cdot)$, $\beta(\cdot)$ and $\gamma(\cdot)$ are quadratic if r_{t+1} given F_t^σ is normally distributed.

[9] Recall that

$$Var\left[r_{t+1}|F_t^\sigma\right] = (1-\phi^2)\sigma_{t+1}^2 - \beta''(0)\sigma_t^2 - \gamma''(0).$$

Those two quantities (ϕ and L) are equal when

$$\psi = \phi \left(\frac{1-e}{2c} \right)^{1/2}. \tag{3.5}$$

We want to maintain this interpretation of ϕ as a leverage effect while understanding that ψ measures both a leverage effect and a volatility feedback effect. This leads us to consider the aggregation formula also used in Khrapov and Renault (2016):

$$\psi = \phi \left(\frac{1-e}{2c} \right)^{1/2} + (1-\phi^2)\left(\varsigma_2 - \frac{1}{2} \right) \tag{3.6}$$

so that ψ is rather an approximation of (3.5) with a small component attributable to the volatility feedback effect that takes the risk premium effect and Jensen effect into account. This parameterization of ψ will be maintained throughout this paper.

Thus, when $\phi \neq 0$ (non-zero leverage effect), the assumption on ψ in (3.4) is clearly unacceptable and the equalities in (3.3) cannot be fulfilled. This motivates a different identification approach: identifying the volatility risk price ς_1 from the observations of stock prices.

See that (3.2) is equivalent to

$$\begin{aligned}
\beta(\varsigma_2 - 1) - \beta(\varsigma_2) &= a\left[\alpha(\varsigma_2) + \varsigma_1\right] - a\left[\alpha(\varsigma_2 - 1) + \varsigma_1\right] \\
\gamma(\varsigma_2 - 1) - \gamma(\varsigma_2) &= b\left[\alpha(\varsigma_2) + \varsigma_1\right] - b\left[\alpha(\varsigma_2 - 1) + \varsigma_1\right].
\end{aligned}$$

From the risk neutral dynamics given in the previous subsection, we can see that this is equivalent to

$$\begin{aligned}
\beta^*(-1) &= -a^*(\alpha^*(-1)) \\
\gamma^*(-1) &= -b^*(\alpha^*(-1))
\end{aligned} \tag{3.7}$$

The restrictions (3.7) will be maintained throughout.

However, as noted by some papers (e.g., Bandi and Reno (2015), Khrapov and Renault (2016)), the identification of volatility risk price from returns data is not strong. We use options data for the identification of ς_1 for the empirical analysis (see section 5.5).

3.3 Risk neutral parameters for ARG(1)-Normal dynamics

The bivariate model of ARG(1) volatility and conditionally Gaussian returns is the leading example of this paper. This is an appealing example since the constraint of structure-preserving change of measure is fulfilled. One thing to note is that this Gaussian assumption does not imply thin tails or symmetry for the marginal and the conditional distribution given F_t of returns.

The risk neutral parameters of this ARG(1)-Normal model are given below in Proposition 2. As we can see later, this allows us to fully specify the functions $\beta(\cdot)$ and $\gamma(\cdot)$ that characterize the returns dynamics.

Proposition 2

Assume that σ_{t+1}^2 is ARG(1) and $r_{t+1}|F_t^{\sigma}$ is normally distributed with the mean and variance specified in Section 2. Then σ_{t+1}^2 and $r_{t+1}|F_t^{\sigma}$ are still ARG(1) and normal under the risk neutral measure with the following risk neutral parameters:

$$\begin{aligned}
&\rho^* = \rho\mathcal{X}(\varsigma)^{-2}, \qquad \delta^* = \delta, \qquad c^* = c\mathcal{X}(\varsigma)^{-1} \\
&e^* = e\left\{\mathcal{X}(\varsigma)\right\}^2, \qquad f^* = f\mathcal{X}(\varsigma)
\end{aligned}$$

and

$$\begin{aligned}
\psi^* &= \psi - (1-\phi^2)\varsigma_2 \\
\phi^* &= \phi \\
(\beta^*)'(0) &= \beta'(0) + e\rho(1-\phi^2)\varsigma_2 \\
(\gamma^*)'(0) &= \gamma'(0) + f\delta c(1-\phi^2)\varsigma_2
\end{aligned}$$

where $\mathcal{X}(\varsigma) = 1 + c\left[\varsigma_1 + \alpha(\varsigma_2)\right]$.

Note that, as implied by the Girsanov theorem in continuous time models, the leverage parameter ϕ is not affected by risk neutralization. By contrast, the volatility feedback effect must be subtracted from ψ to get ψ^* leaving us with

$$\psi^* = \phi\left(\frac{1-e}{2c}\right)^{1/2} - \frac{1}{2}(1-\phi^2),$$

that is a (rescaled) leverage effect and Jensen effect.

I can now specify the functions $\beta(\cdot)$ and $\gamma(\cdot)$ using the results in Proposition 2 and the equations given in (3.7) with $\beta'(0)$ and $\gamma'(0)$ given as follows:

$$\begin{aligned}
\beta'(0) &= a^*(\alpha^*(-1)) - e\rho(1-\phi^2)\left(\varsigma_2 - \frac{1}{2}\right) \\
\gamma'(0) &= b^*(\alpha^*(-1)) - f\delta c(1-\phi^2)\left(\varsigma_2 - \frac{1}{2}\right),
\end{aligned}$$

where

$$\alpha^*(-1) = -\psi^* - \frac{1}{2}(1-\phi^2) = -\phi\left(\frac{1-e}{2c}\right)^{1/2}.$$

3.4 Option pricing

Given the assumption that returns are conditionally normally distributed, Khrapov and Renault (2016) propose the following option price formula for a one-period call option:

$$\begin{aligned} C_t(x_t,\phi) &= \exp(-r_{f,t})E_t^*\left[max\{0, S_{t+1}-K\}\right] \\ &= E_t^*\left[BS(S_t\xi_{t,t+1},(1-\phi^2)\sigma_{t+1}^2,K)\right], \end{aligned}$$

where $x_t = log(K/S_t)$ is the log-moneyness of the option at time t, S_t and K are the asset and the strike price, and price adjustment $\xi_{t,t+1}$ is defined by

$$log\xi_{t,t+1} = E^*[r_{t+1}|I_t^\sigma] + \frac{1}{2}Var^*[r_{t+1}|I_t^\sigma]$$

and $BS(\cdot)$ is the standard Black-Scholes formula.

Note that a multi-period (T) European call option price is

$$\begin{aligned} C_t(x_t,T) &= \exp\left(-\sum_{i=t}^{T-1} r_{f,i}\right)E_t^*\left[max\{0, S_{t+T}-K\}\right] \\ &= S_tE_t^*\left[max\left\{0, \exp\left(\sum_{i=1}^{T} r_{t+i}\right) - \exp\left(x_t\right)\right\}\right], \end{aligned}$$

where $x_t = log(K/S_t) - \sum_{i=t}^{T-1} r_{f,i}$. The closed-form expression of $C_t(x_t,T)$ is given in Proposition 3 below.

Proposition 3

Assume that volatility σ_t^2 and returns are bivariate CAR(1) as defined in section 2.3. Also assume a constant risk-free rate, i.e. $r_{f,t} = r$, $\forall t$. Then the price at time t of a European call option at time $t+T$ is given by

$$C_t(x_t,T) = S_tP_1(x_t,T) - K\exp\left(-rT\right)P_2(x_t,T),$$

where

$$P_1(x_t,T) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left\{\frac{\exp\left(-iux_t f^*_{t,t+T}(iu+1)\right)}{iu}\right\} du$$
$$P_2(x_t,T) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left\{\frac{\exp\left(-iux_t f^*_{t,t+T}(iu)\right)}{iu}\right\} du$$

where $f^*_{t,t+T}(\cdot)$ is the conditional characteristic function of $\sum_{i=1}^T r_{t+i}$ at time t under the risk neutral measure given in the appendix B, and $\Re(\cdot)$ denotes the real part of a complex number.

4 Estimation with Generalized Method of Moments (GMM)

The equations in (2.6) defined in Section 2 provide closed-form conditional moment restrictions that identify $\theta = (\rho, \delta, c, \omega, e, f, \phi, \varsigma_1, \varsigma_2)$. They imply that the following unconditional moment restrictions hold:

$$E\left[A(\sigma_t^2)\left[\exp\left(-u\sigma_{t+1}^2\right) - \exp\left\{-a^\theta(u)\sigma_t^2 - b^\theta(u)\right\}\right]\right] = 0, \quad \forall u \in \mathbb{C}$$
$$E\left[B(\sigma_{t+1}^2,\sigma_t^2)\left[\exp\left(-ur_{t+1}^2\right) - \exp\left\{-\alpha^\theta(u)\sigma_{t+1}^2 - \beta^\theta(u)\sigma_t^2 - \gamma^\theta(u)\right\}\right]\right] = 0, \quad \forall u \in \mathbb{C}$$

where $A(\sigma_t^2)$ and $B(\sigma_{t+1}^2,\sigma_t^2)$ are some instruments and the functions have a superscript θ to highlight that they depend on the unknown parameters θ. Note that even with a small set of instruments $A(\cdot)$ and $B(\cdot,\cdot)$, the above set of moments may be very rich since it is written for any complex number u.

4.1 Identification

4.1.1 Possible identification failure

One of the critical assumptions for GMM estimators to be consistent for true parameters is that the parameters in the conditional moment restrictions are globally identified by the induced unconditional moment conditions. However, the global identification assumption may fail in nonlinear models when the instruments are chosen in an arbitrary way because the unconditional moments implied with the chosen instruments may convey less information than the original conditional moments (Dominguez and Lobato (2004), Hsu and Kuan (2011)).

The example of ARG(1) model, the leading example of this paper for the volatility factor σ_t^2, illustrates this idea with a constant instrument.

Example: ARG(1) model

Assume that the random variable $X_{t+1},\ t = 1, 2, \cdots$, is ARG(1). Recall that when X_{t+1} is ARG(1), we have:

$$E\left[\exp\left(-uX_{t+1}\right) - \exp\left\{-a(u)X_t - b(u)\right\}|X_t\right] = 0, \qquad \forall u \in \mathbb{C}$$

with

$$a(u) = \frac{\rho u}{1 + cu}, \qquad b(u) = log(1 + cu).$$

Let ρ^0, δ^0, and c^0 denote the true values of the parameter ρ, δ, and c, respectively. Consider unconditional moment restrictions implied by the above conditional moment restrictions with a constant instrument. Theorem 1 shows that there are many parameter values including the true ones that satisfy such unconditional moments regardless of the choice of u's.

Theorem 1

Suppose the observations $\{X_t\}_{t=1}^{T}$ follow ARG(1). Consider the below J number of unconditional moments:

$$E\left[\exp\left(-u_jX_{t+1}\right) - \exp\left\{-a(u_j)X_t - b(u_j)\right\}\right] = 0, \qquad u_j \in \mathbb{C}$$

for $j = 1, 2, \cdots, J$ where $J \geq 3$. Then there are infinitely many values of the tuple (ρ, δ, c) such that these moments are satisfied.

The proof of Theorem 1 is given in the appendix B. Figure 2 in the appendix A visualizes a result with Monte Carlo histograms of GMM estimators with the given unconditional moments for some choices of equally spaced u's in $[1i, 10i]$. Figure 3 shows that using an optimal instrument (Hansen, 1985) would not fix the identification problem. The optimal instrument has the closed-form

$$D(X_t)'\Omega(X_t)^{-1}$$

where $D(X_t)$ is a $J \times 3$ matrix of the Jacobian of the given conditional moments such that the j-th row of $D(X_t)$ is given as

$$\begin{aligned} &\exp\left\{-a(u_j)X_t - b(u_j)\right\}\begin{pmatrix} \frac{\partial a(u_j)X_t + b(u_j)}{\partial \rho} & \frac{\partial a(u_j)X_t + b(u_j)}{\partial \delta} & \frac{\partial a(u_j)X_t + b(u_j)}{\partial c} \end{pmatrix} \\ =&\exp\left\{-a(u_j)X_t - b(u_j)\right\}\begin{pmatrix} \frac{u_j}{1+cu_j}X_t & log(1 + cu_j) & -\frac{\rho u_j^2}{(1+cu_j)^2}X_t + \frac{\delta u_j}{1+cu_j} \end{pmatrix} \end{aligned}$$

and $\Omega(X_t)^{-1}$ is a $J \times J$ variance matrix of the given conditional moments such that the (j, s) element of $\Omega(X_t)^{-1}$ is given as

$$\begin{aligned}
&E\left[\{\exp\left(-u_j X_{t+1}\right) - \exp\left\{-a(u_j)X_t - b(u_j)\right\}\}\left\{\exp\left(u_s X_{t+1}\right) - \exp\left\{-a(u_s)X_t - b(u_s)\right\}\right\}|X_t\right] \\
&= \exp\left\{-a(u_j - u_s)X_t - b(u_j - u_s)\right\} - \exp\left\{-\left[a(u_j) + a(-u_s)\right]X_t - \left[b(u_j) + b(-u_s)\right]\right\}.
\end{aligned}$$

4.1.2 Choice of instrument

Dominguez and Lobato (2004) propose to use a complete set of indicator functions as a continuum of identifying instruments. Hsu and Kuan (2011) note that there are many different ways to get a continuum of identifying instruments. Let X be a m-dimensional conditioning variable and consider $G(A(X,\tau))$ with $\tau \in \mathcal{T} \subset \mathbb{C}^{m+1}$ and $A(X,\tau) = \tau_0 + \sum_{j=1}^{m} X_j\tau_j$ for some function G. When G belongs to a class of generically comprehensively revealing (GCR) functions, the continuum of unconditional moments induced by the instruments $G(A(X,\tau))$ contains the same information as the initial conditional moment restrictions (Stinchcombe and White (1998)).

Examples of GCR functions include exponential and logistic functions. With the ARG(1) example, we will see how global identification is maintained for such induced unconditional moments.

Example: ARG(1) model

Since the conditional characteristic function of X_{t+1} at u is in exponentially affine form, the simplest GCR function we can use is an exponential function: $\exp\left(-vX_t\right)$. Hsu and Kuan (2011) suggest that we use a continuum of such instruments (a continuum of v) but we consider both cases with $u = v$ and $u \neq v$.

Theorem 2 below shows that (ρ^0, δ^0, c^0) are identified when $u \neq v$ but not identified when $u = v$. Also, it implies that for identification purposes, only a finite number of instruments (with a finite number of v values) is sufficient rather than a whole continuum of them.

Theorem 2

Suppose the observations $\{X_t\}_{t=1}^{T}$ follow ARG(1). Consider the below J number of unconditional moments:

$$E\left[\exp\left(-u_j X_t\right)\left\{\exp\left(-u_j X_{t+1}\right) - \exp\left\{-a(u_j)X_t - b(u_j)\right\}\right\}\right] = 0$$

for $j = 1, 2, \cdots, J$ where $J \geq 3$. Then the true parameters (ρ^0, δ^0, c^0) are not jointly identified from the above moments regardless of the value of J.

Now consider the below J number of unconditional moments:

$$E\left[\exp\left(-vX_t\right)\left\{\exp\left(-u_jX_{t+1}\right)-\exp\left\{-a(u_j)X_t-b(u_j)\right\}\right\}\right]=0$$

where $(u_j,v)'\in\mathbb{C}^2$ and $v\neq u_j$ for all $j=1,2,\cdots,J$, $J\geq 3$. Then (ρ^0,δ^0,c^0) are jointly identified from these moments for all $J\geq 3$.

The first and second sets of instrument in Theorem 2 are the Single Index (SI) and the Double Index (DI) moment conditions (Carrasco et al., 2007). It has already been mentioned above that the SI moments are not sufficient for identification since they include the one induced with the optimal instrument as a particular case. Theorem 2 states that identification is guaranteed with the DI moments. Figure 4 and 5 present the results of some Monte Carlo experiments about GMM estimators based on the SI and DI moments, respectively.

4.2 GMM with a large number of moment conditions

Let X_t be a Markov process of order 1 with the following conditional characteristic function[10]:

$$E\left[\exp\left(-uX_{t+1}\right)|X_t\right]=g(u|X_t,\theta^0).$$

We consider the DI moments so that the moment function for each observation is

$$\psi_t(\theta)=\exp(-vx_t)\left\{\exp(-ux_{t+1}-g\left(u|x_t;\theta\right)\right\},$$

where both x_t and x_{t+1} are univariate. Note that this function is defined almost everywhere in a closed subset of $\mathbb{C}^2$ which implies that a continuum of unconditional moments exist. Carrasco and Florens (2000, 2014) propose a GMM estimation technique using a whole continuum of moments (C-GMM). In this paper, we consider GMM estimation using a countable number of available moments where the number of moments increases with the sample size. We dub this 'D-GMM' (Discrete GMM) as a counterpart to C-GMM.

Then the GMM estimator is defined as follows.

Definition of GMM estimator:

$$\hat{\theta}_T=\underset{\theta\in\Theta}{argmin}\;\bar{\psi}_T(\theta)'\hat{W}_T\bar{\psi}_T(\theta) \tag{4.1}$$

[10] I assume Markov of order 1 for simplicity. This can be easily generalized to any higher order (see the application in section 5).

where $\bar{\psi}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\psi_{t,T}(\theta)$. Let $J_{1,T}$ and $J_{2,T}$ denote the dimension of u and v, respectively. Then $\psi_{t,T}(\theta)$ is a $J_T = J_{1,T} \times J_{2,T}$ vector that has the moment function defined on different values of v and u, and $\hat{W}_T$ is a sample-based weighting matrix.

Let Ω_T denotes the variance matrix of the moments:

$$\Omega_T = E\left[\psi_{t,T}(\theta^0)\psi_{t,T}(\theta^0)'\right].$$

As noted by Carrasco and Florens (2000) and Carrasco et al. (2007), the minimum eigenvalue of the optimal weighting matrix (Ω_T^{-1}) converges to 0 as we refine the grids of u since the neighboring moments become closely correlated with each other.

In order to address this problem, I consider the Tikhonov method of regularization suggested by Carrasco and Florens (2000). That is, I choose $\alpha > 0$ and let $W_T = (\Omega_T^2 + \alpha I_{J_T})^{-1}\Omega_T$ where I_J is a $J \times J$ identity matrix for all $J \in \mathbb{N}$. By letting $\alpha = \alpha_T \to 0$ as $T \to \infty$, W_T may be a good proxy of Ω_T^{-1}. $\hat{W}_T$ is then the sample analogue of W_T defined as follows.

Definition of the weighting matrix:

$$\hat{W}_T = (\hat{\Omega}_T^2 + \alpha_T I_{J_T})^{-1}\hat{\Omega}_T, \tag{4.2}$$

where $\hat{\Omega}_T$ is the sample analogue of Ω_T and may depend on consistent preliminary parameter estimates, $\tilde{\theta}_T$.

The practical use of this weighting matrix will require the choices of the tuning parameters α_T and J_T. However, D-GMM is simple to implement (it is a standard two-step GMM estimator) and computationally less expensive than the C-GMM estimator as the number of moment conditions, J_T, is usually required to diverge at a much slower rate than the sample size. Also the smoothing parameter α_T allows us to relax the rather restrictive assumption that the minimum eigenvalue of the variance matrix of the moments is bounded away from 0 for any J_T and still implement efficient GMM estimation ($\alpha_T \to 0$ as $T \to \infty$).

Assumption 1

Let $\rho(Y_t, u, \theta) = \rho_t(u, \theta) = \exp(-uX_{t+1}) - g(u|X_t;\theta)$ for a given u. Then

$$E\left[\rho_t(u,\theta)|X_t\right] = 0 \quad \Leftrightarrow \quad \theta = \theta^0 \quad , \forall u \in \mathbb{C}.$$

Assumption 2

1. $\theta^0 \in int(\mathcal{B})$ and $\mathcal{B}$ is compact.

2. $\underset{\theta\in\mathcal{B}}{sup}\, E\left[\|\rho(y,u,\theta)\|^2|X_t\right]$ is bounded for all $u\in\mathbb{C}$.
3. For all $\theta\in\mathcal{B}$, there exists $\delta(y)$ such that $\|\rho(y,u,\theta)-\rho(y,u,\theta^0)\|\le\delta(y)\|\theta-\theta^0\|$ and $E\left[\delta(y)^2|X\right]<\infty$.

Assumption 3

1. $g(u|X,\theta)$ is twice continuously differentiable in a neighborhood $\mathcal{N}$ of θ^0.
2. $\underset{\theta\in\mathcal{B}}{sup}\, E\left[\left\|\frac{\partial g(u|x,\theta)}{\partial\theta'}\right\|^2\right]<\infty$ and $\underset{\theta\in\mathcal{B}}{sup}\, E\left[\left\|\frac{\partial^2 g(u_m|x,\theta)}{\partial\theta\theta'}\right\|^2\right]<\infty$ for each $m=1,2,\cdots,M_T$.
3. Let $D(X)=\frac{\partial g(u|X,\theta^0)}{\partial\theta'}$. Then $E\left[D(X)'D(X)\right]$ is positive definite.

The above assumptions are Assumption 3 and 4 in Donald et al. (2003) that are standard conditions imposed for consistency and asymptotic normality. Assumption 1 is the condition for θ^0 to be identified (for conditional restrictions). Assumption 2 imposes a bounded second conditional moment and Lipschitz condition for uniform convergence. Part (3) of Assumption 3 is a local identification condition required for asymptotic normality. Part (1) and the condition on the first derivative of g in Part (2) are standard smoothness conditions. The assumption of twice differentiability of g in Part (2) is stronger than is usually assumed due to the growing number of moments.

For the results in Theorem 3 and 4, I assume that the dimension of $\rho(y,u,\theta)$ is finite and fixed meaning that $J_{1,T}=J_1<\infty$ but let $J_{2,T}\to\infty$ as $T\to\infty$. Theorem 3 and 4 below show that then the GMM estimator with $\hat{W}_T$ defined above as the weighting matrix is consistent and attains Chamberlain's (1987) semiparametric efficiency bound. Let $\hat{\theta}_T$ denotes such GMM estimator.

The GMM model I study here is a specific example of the models considered in Donald et al. (2003):

$$E\left[\rho(Y,\theta)|X\right]=0\quad\Leftrightarrow\quad\theta=\theta^0.$$

The series of instruments is denoted as, with $J=J_T$,

$$q^J(x)=(q_1^J(x),q_2^J(x),\cdots,q_J^J(x))$$

and the unconditional moment functions are

$$\psi_{t,T}(\theta)=\rho(y_t,\theta)\otimes q^J(x_t).$$

Donald et al. (2003) assume that ρ is finite dimensional and $\Sigma(z)=E\left[\rho(y,\theta^0)\rho(y,\theta^0)'|x\right]$ and $E\left[q^J(x)q^J(x)'\right]$ have the eigenvalues bounded away from zero. However, these assumptions can be unrealistic when infinite dimensionality is involved, as in the example in this paper. I now relax the latter assumption and allow that $\lambda_{min}\left(E\left[q^J(x)q^J(x)'\right]\right)\to 0$.

I show by a slight extension of the proofs of theorem 5.3 and 5.4 in Donald et al. (2003) that the

GMM estimator $\hat{\theta}_T$ defined in (4.1) is consistent and attains Chamberlain's semiparametric efficiency bound.

Theorem 3 (Consistency)

Suppose Assumption 1, 2, and 3 hold. Assume $\underset{x\in\mathcal{X}}{sup} \left\|q^J(x)\right\| = \sqrt{J_T}$. Also assume that $\alpha_T \to 0$ and $J_T^2/T\alpha_T^2 \to 0$ as $T \to \infty$. Then

$$\hat{\theta}_T \xrightarrow{p} \theta^0.$$

Theorem 4 (Efficiency bound)

Assume that the conditions for Theorem 1 are satisfied and $\left\|\tilde{\theta}_T - \theta^0\right\| = O_p(T^{-1/2})$ where $\tilde{\theta}_T$ is a preliminary estimator that enters the weighting matrix. Then

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

where

$$\begin{aligned} V^{-1} &= E\left[D(X)'\Sigma(X)^{-1}D(X)\right] \\ D(X) &= E\left[\frac{\partial\rho(Y,\theta^0)}{\partial\theta'}\bigg|X\right]. \end{aligned}$$

The choice of the instrument in this paper $\exp(-v'X_t)$, $v \in \mathbb{C}^L$ guarantees the identification of θ^0 from the unconditional moment conditions when a continuum of v is used (or $J_T \to \infty$). Then, with Assumption 1, 2, and 3, Lemma 2.1 of Donald et al. (2003) holds. Also, $\|\exp(-v'X_t)\|$ is bounded (by 1) so the condition $\underset{x\in\mathcal{X}}{sup} \left\|q^J(x)\right\| \leq \sqrt{J_T}$ holds. Therefore, the GMM estimator defined in (4.1) with the weighting matrix $\hat{W}_T$ defined in (4.2) is consistent and attains semiparametric efficiency bound for a finite dimensional $\rho(u,\theta)$.

Now let $\hat{\theta}_T$ be defined as in (4.1) with both $J_{1,T} \to \infty$ and $J_{2,T} \to \infty$. Then Corollary 1 below shows that this GMM estimator attains MLE efficiency when we have conditional moments based on characteristic functions. C-GMM also attains MLE efficiency (Carrasco et al. (2007)) and thus, D-GMM and C-GMM have the same limit distribution.

Corollary 1

Suppose Assumption 1, 2, and 3 hold. Suppose that the entries of $u = (u_1, u_2, \cdots, u_{J_{1,T}})'$ are equally spaced (i.e. $u_i - u_{i-1} = (u_{J_{1,T}} - u_1)/(J_{1,T} - 1)$) on the interval $[-R, R] * 1i$ where $R \in \mathbb{R}_{++}$

such that $J_{1,T} \to \infty$ and $R/(J_{1,T}-1) \to \infty$ as $T \to \infty$. Then

$$\sqrt{T}(\hat{\theta}_T - \theta^0) \xrightarrow{d} \mathcal{N}(0, I(\theta^0)^{-1})$$

as both $J_{1,T}$ and $J_{2,T}$ diverge to infinity such that $J_T^2/T\alpha_T^2 \to 0$ as $T \to \infty$, where $I(\theta^0)$ is the Fisher information matrix.

5 Estimation and Empirics with Realized Variance

In this section, we present an empirical application of of the model developed in section 2 applying the results of section 4 for some part (section 5.3). We implement a two-step estimation that the parameters characterizing volatility dynamics are estimated in the first step. The first step uses the observations of realized variance only that is shown to be ARMA(1,1) and thus, is not Markov. However, as section 5.3 shows, we treat realized variance as an AR$(H+1)$ with some $H \geq 1$, which allows us to see the dynamics of realized variance as an application of the setup in section 4: realized variance is CAR$(H+1)$ and an infinitely many number of conditional moment restrictions based on conditional characteristic functions is available. We use the DI moment conditions in order to ensure identification. This AR$(H+1)$ approximation of realized variance generates a misspecification error. However, we assume that this error is close to zero and does not have a significant effect on the estimation. This assumption is supported by the overidentification test performed in section 5.3.

5.1 Data

The dataset was obtained from Oxford-Man Institute[11] and consists of the daily log returns and realized volatilities of the S&P 500 over the period from January 2000 to June 2016. The sample size is 4,121. Variable r_t denotes the daily log returns in excess of the risk-free rate, which is proxied by the yield on a 30-day treasury bill[12]. The realized variance process $\{RV_t\}$ is computed from 5-minute intraday returns.

5.2 ARG(1)-Normal model

The bivariate CAR model of returns and volatility that we use for the empirical analysis is the ARG(1)-Normal model. The volatility σ_{t+1}^2 is ARG(1) and the functional forms of $a(u)$ and

[11]Oxford-Man Institutes realized library, http://realized.oxford-man.ox.ac.uk
[12]This rate is obtained from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data library.html

$b(u)$ are given in Section 4. Returns are assumed to be normally distributed conditional on the contemporaneous and past volatility, i.e. $r_{t+1}|F_t^\sigma$ is normal. As a reminder, the conditional mean and variance of returns are as follows:

$$E\left[r_{t+1}|F_t^\sigma\right] = \mu_{t+1} = \psi\sigma_{t+1}^2 + \beta'(0)\sigma_t^2 + \gamma'(0)$$
$$Var\left[r_{t+1}|F_t^\sigma\right] = \Sigma_{t+1} = (1-\phi^2)RV_{t+1},$$

where

$$\psi = \phi\left(\frac{1-e}{2c}\right)^{1/2} + (1-\phi^2)\left(\varsigma_2 - \frac{1}{2}\right)$$
$$\beta'(0) = a^*\left[\alpha^*(-1)\right] - e\rho(1-\phi^2)\left(\varsigma_2 - \frac{1}{2}\right)$$
$$\gamma'(0) = b^*\left[\alpha^*(-1)\right] - f\delta c(1-\phi^2)\left(\varsigma_2 - \frac{1}{2}\right).$$

As shown in Proposition 3, the risk neutral dynamics remain the same as the historical dynamics with the risk neutral parameters given in the proposition. This model has 4 volatility parameters $\theta_\sigma = (\rho, \delta, c, e)'$ and 3 returns parameters $\theta_r = (\phi, \varsigma_1, \varsigma_2)'$ giving us in total 7 parameters to estimate. I estimate $\theta = (\theta_\sigma', \theta_r')'$ using a two-step estimation procedure. I first estimate θ_σ from the historical data of realized variance only. Then, treating θ_σ as given by its estimate, I estimate θ_r from the observations of returns and realized variance.

5.3 GMM estimation of volatility dynamics with realized variance

The ARG(1) volatility σ_{t+1}^2 is unobservable. What we observe is realized variance RV_{t+1} constructed from high-frequency data that we have defined as

$$RV_{t+1} = \sigma_{t+1}^2 - e\rho\sigma_t^2 - f\delta c, \qquad t = 1, 2, \cdots, T-1.$$

This suggests that the series of realized variances is informative about the path of the volatility factor σ_{t+1}^2. In fact, it is easy to see that the unobservable volatility factor can be represented by an affine form of the infinite series of the contemporaneous and past observable volatility factor:

$$\sigma_{t+1}^2 = \sum_{k=0}^{\infty} (e\rho)^k \left(RV_{t+1-k}^2 + f\delta c\right). \tag{5.1}$$

Then, by assuming $|e\rho| < 1$, σ_{t+1}^2 can be approximated a finite series of observations of realized variance such that for some $H < \infty$,

Approximated volatility:

$$\sigma_{t+1}^2 \approx \left[\sum_{k=0}^{H} (e\rho)^k RV_{t+1-k}\right] + \frac{f\delta c}{1-e\rho}. \tag{5.2}$$

Using the above approximation of the unobservable volatility factor, we have an approximation to the conditional characteristic functions of realized variance as if it were CAR($H+1$). To see this, recall that we have the following:

$$E\left[\exp\left(-u\sigma_{t+1}^2\right)|F_t\right] = \exp\left\{-a(u)\sigma_t^2 - b(u)\right\}, \quad u \in \mathbb{C}.$$

Then plugging the approximation of σ_{t+1}^2 and σ_t^2 given in (5.2) leads us to the following conditional characteristic function of RV

Approximated conditional characteristic function of RV:

$$E\left[\exp\left(-uRV_{t+1}\right)|F_t\right] \approx \exp\left\{-\tilde{a}(u)'\underline{RV}_t - \tilde{b}(u)\right\}, \quad u \in \mathbb{C} \tag{5.3}$$

where $\underline{RV}_t = (RV_t, RV_{t-1}, \cdots, RV_{t-H})'$, $\tilde{a}(u)$ that is a $(H+1)\times 1$ vector with

$$\tilde{a}_i(u) = (e\rho)^{i-1}\left[a(u) - e\rho u\right]$$

as its i-th element and

$$\tilde{b}(u) = b(u) + \frac{f\delta c}{1-e\rho}\left[a(u) - u\right].$$

Note that this is a misspecified model of realized variance since realized variance is not Markov and does not have a closed-form conditional characteristic function. However, the misspecification is local when letting $H = H_T$ diverge as $T \to \infty$.

We estimate θ_σ using D-GMM with the approximate conditional moments given in (5.3). The discussion in Section 4 suggests that for each u we should construct the unconditional moments with the DI instrument $\exp\left(-v'\underline{\tilde{\sigma}}_t^2\right)$ with $v \in \mathbb{C}^{H+1}$. However, using the following instrument

$$Z_t = \exp\left\{-v_1 RV_t - v_2 RV_{t-1}\right\}, \quad (v_1, v_2)' \in \mathbb{C}^2 \tag{5.4}$$

is sufficient for identification since the $i+1$-th element of $\tilde{a}(u)$ denoted as $\tilde{a}_{i+1}(u)$ is $e\rho\tilde{a}_i(u)$ for all $i = 1, 2, \cdots, H$.

Unconditional GMM moment restrictions:

$$E\left[\psi_t(u,\theta_\sigma)\right] = E\left[\begin{pmatrix} \Re\left(Z_t\left\{\exp\left(-uRV_{t+1}\right) - \exp\left(-\tilde{a}(u)'\underline{RV}_t - \tilde{b}(u)\right)\right\}\right) \\ \Im\left(Z_t\left\{\exp\left(-uRV_{t+1}\right) - \exp\left(-\tilde{a}(u)'\underline{RV}_t - \tilde{b}(u)\right)\right\}\right)\end{pmatrix}\right] = 0, \quad u \in \mathbb{C} \quad (5.5)$$

where $\Re[\cdot]$ and $\Im[\cdot]$ denote the real and the imaginary part of a complex number, respectively.

We use 5 equally-spaced u, v_1, and v_2 points on $[1i, 10i]$. This generates $J = 2 \times 5 \times 5 \times 5 = 250$ moment conditions. The GMM estimator is obtained by minimizing the objective function given in (4.1) with $\hat{W}_T$ given in (4.2) for a choice of α_T. The preliminary estimates are obtained with an identity weighting matrix.

We understand that, for empirical analysis, we have a finite sample size and have to choose a finite H. We apply the asymptotic results derived in section 4 pretending that the misspecification error is zero. In order to see whether this assumption is reasonable, we perform an overidentification test with the moment conditions in (5.5). Since we use a large number of moment conditions, we use the test statistic proposed by Carrasco and Florens (2000):

$$\tau_T = \frac{\sqrt{T}\hat{Q}_{T,\alpha_T}(\hat{\theta}_\sigma) - \hat{p}_{T,\alpha_T}}{\sqrt{\hat{q}_{T,\alpha_T}}},$$

where $\hat{Q}_{T,\alpha_T}(\cdot)$ is the GMM objective function given in (4.1) with a chosen α_T and $\hat{\theta}_\sigma$ is the GMM estimator minimizing $\hat{Q}_{T,\alpha_T}(\cdot)$. $\hat{p}_{T,\alpha_T}$ and $\hat{q}_{T,\alpha_T}$ are defined as follows:

$$\hat{p}_{T,\alpha_T} = \sum_{j=1}^{J} \frac{\hat{\lambda}_j^2}{\hat{\lambda}_j^2 + \alpha_T}, \qquad \hat{q}_{T,\alpha_T} = 2\sum_{j=1}^{J} \frac{\hat{\lambda}_j^4}{\left(\hat{\lambda}_j^2 + \alpha_T\right)^2},$$

where $\hat{\lambda}_j$ is the j-th eigenvalue of $\hat{\Omega}_T$ defined in (4.2). Carrasco and Florens (2000) show that , as $T \to \infty$,

$$\tau_T \xrightarrow{d} \mathcal{N}(0,1)$$

as long as α_T does not decrease to 0 too fast. The overidentification test is not rejected when $\alpha_T \geq 0.04$ is used at 5% significance level which supports our assumption of zero misspecification error. Figure 6 shows that $|\tau_T| < 1.96$ for all $\alpha_T \geq 0.04$.

5.4 MLE estimation of returns dynamics

Let $\hat{\theta}_\sigma = (\hat{\rho}, \hat{\delta}, \hat{c}, \hat{e})'$ denote the D-GMM estimate of θ_σ. Once we obtain the GMM estimates of the volatility parameters, we construct the estimates of the unobservable volatility factor σ_{t+1}^2

as

$$\hat{\sigma}_{t+1}^2 = \sum_{l=0}^{H+1} (\hat{e}\hat{\rho})^l \, RV_{t+1-l} + \frac{\hat{f}\hat{\delta}\hat{c}}{1-\hat{e}\hat{\rho}}.$$

The returns parameters to estimate are ϕ, ς_1, and ς_2. The volatility risk premium ς_1 is priced and we will treat it as given (see next subsection). We then estimate the returns parameters ϕ and ς_2 using the series of the estimates of σ_{t+1}^2 and the volatility parameter estimates by MLE for a given ς_1. By the assumption of conditional normality of returns, the conditional likelihood of returns at time $t+1$ is defined as

$$l(r_{t+1}; \hat{\sigma}_{t+1}^2, \hat{\sigma}_t^2, \hat{\theta}_\sigma, \varsigma_1) = \frac{1}{\sqrt{2\pi\hat{\Sigma}_{t+1}}} \exp\left\{ -\frac{1}{2\hat{\Sigma}_{t+1}} \left(r_{t+1} - \hat{\mu}_{t+1} \right)^2 \right\},$$

where $\hat{\Sigma}_{t+1}$ and $\hat{\mu}_{t+1}$ are estimates of Σ_{t+1} and μ_{t+1} with $\hat{\sigma}_{t+1}^2$ and $\hat{\theta}_\sigma$. Then the MLE estimates $\hat{\phi}$ and $\hat{\varsigma}_2$ are obtained by maximizing

$$L^R(\hat{\phi}, \hat{\varsigma}_2) = \sum_{t=1}^{T-1} l(\phi, \varsigma_2; r_{t+1}, \hat{\sigma}_{t+1}^2, \hat{\sigma}_t^2, \hat{\theta}_\sigma, \varsigma_1)$$

5.5 Estimation with options data

One thing to note is that the volatility risk parameter ς_1 is weakly identified (e.g., Bandi and Reno (2015), Khrapov and Renault (2016)). This point is supported empirically. When the volatility risk premium is not identified from the returns and volatility data, options data can be used to price it (e.g., Corsi et al. (2013)).

The options data can play another role, that is to provide a criterion for choosing the regularization parameter α_T[13]. I use IVRMSE put forward by Renault (1997) for the volatility risk premium and α_T:

$$IVRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(IV_i^{hist} - IV_i^{mod}(\varsigma_1, \alpha_T) \right)^2},$$

where IV_i^{hist} and IV_i^{mod} denote the i-th observation of historical implied volatility and the implied volatility generated by the model[14], respectively.

Another choice parameter is H, the number of lags to be included in the estimation. However, the estimates seem stable for all $H \geq 10$ (see table 1 for cases with $H = 5, 10, 15, 20, 30$ with $\alpha_T = 0.1$ and $\varsigma_1 = -10$). I use $H = 10$ for the rest of the empirical analysis.

[13]Another approach for choosing the regularization parameter is to minimize the estimation error such as mean squared error (MSE). However deriving the estimation error is difficult in nonlinear models and there are no sound criteria for choosing α_T.

[14]Options prices are computed using the option pricing formula given in Proposition 3.

Table 1: Estimates of the parameters of ARG(1)-Normal with various H

param.	$H = 5$	$H = 10$	$H = 15$	$H = 20$	$H = 30$
ρ	0.9274	0.9357	0.9358	0.9364	0.9364
	(0.0534)	(0.0535)	(0.0534)	(0.0534)	(0.0532)
δ	0.6299	0.6403	0.6365	0.6327	0.6293
	(0.1991)	(0.2109)	(0.2099)	(0.2102)	(0.2077)
c	1.86e-5	1.67e-5	1.67e-5	1.67e-5	1.68e-5
	(6.49e-6)	(6.50e-6)	(6.52e-6)	(6.55e-6)	(6.56e-6)
e	0.3454	0.3396	0.3394	0.3396	0.3394
		(0.0313)	(0.0312)	(0.0311)	(0.0311)
Persistence	0.9274	0.9357	0.9358	0.9364	0.9364
ϕ	-0.1682	-0.1515	-0.1512	-0.1506	-0.1514
	(0.0131)	(0.0123)	(0.0123)	(0.123)	(0.0123)
ς_2	0.4708	1.0861	1.4700	1.4673	1.4953
	(1.2485)	(1.2503)	(1.2518)	(1.2535)	(1.2547)
ς_1	-10	-10	-10	-10	-10

* The standard errors are given in parentheses.
* $ARMA(1,1)_h$ means estimation with ARMA(1,1) realized variance with $H = h$ for $h = 5, 10, 15, 20, 30$.

Table 2: Estimates of the parameters

param.	ARG(1)-Normal[1]	ARG(1)-Normal[2]	ARV
ρ	0.9632	0.9895	
	(0.0377)	(0.0229)	
δ	0.6469	0.9552	
	(0.2385)	(0.6286)	
c	1.13e-5	3.75e-06	
	(4.73e-6)	(2.69e-6)	
e	0.4013	0.6206	
	(0.0204)	(0.0148)	
α_2			4.99e-6
			(2.07e-7)
β_2			4.76e-5
			(0.0369)
γ_2			438.59
			(14.19)
ω_2			4.98e-6
σ			1.00e-5
			(1.00e-5)
ρ			0.1597
			(0.0275)
Persistence	0.9632	0.9895	0.9595
ϕ	-0.1340	-0.1	
	(0.012)	(0.008)	
ς_2	0.1458		
	(1.2904)		
ς_1	-0.5		
$\chi(\varsigma)$	1	1.003	
λ_2			0.0540
			(0.0405)
γ_2^*			443.16
$\mathcal{X}$			-4.57
IVRMSE	4.7240	4.1959	4.6056

* The standard errors are given in parentheses.
* $H = 10$ for ARG(1)-Normal models.
* $\alpha_T = 0.01$ and $\varsigma_1 = -0.5$ for ARG(1)-Normal[1] model.
* $\alpha_T = 0.0083$ for ARG(1)-Normal[2] model.
* Persistence level of ARV model is computed as $\beta_2 + \alpha_2\gamma_2^2$.
* ω_2 is computed as $E\left[h_t^{RV}\right]\left(1-\beta_2-\alpha_2\gamma_2^2\right)-\alpha_2$ where the unconditional mean $E\left[h_t^{RV}\right]$ is the sample mean of realized variance.

I choose α_T and ς_1 that jointly minimize the IVRMSE. Since there is a continuum of α_T that could produce the same results, I choose α_T from a small grid on $[0.0001, 1]$. The computed values are $\alpha_T = 0.01$ and $\varsigma_1 = -0.5$. With this choice of α_T, the volatility parameters are estimated. Then given those volatility parameters and ς_1, the returns parameters are estimated. The results are presented in table 2. The first column shows the parameter estimates of the ARG(1)-Normal model with $H = 10$. The leverage effect, ϕ, is estimated to be around -0.13.

The second column displays the estimate of $\chi(\varsigma)$ that is determined jointly by ς_1 and ς_2. When the option pricing formula given in Proposition 3 is applied to the ARG(1)-Normal example, an option price is a function of $\chi(\varsigma)$. The estimates of ϕ do not seem to be much affected by the values of ς_1 and, using this fact, we estimate $\chi(\varsigma)$ by minimizing IVRMSE for a given value of ϕ that is estimated from the returns data. The IVRMSE is minimized for $\alpha_T = 0.0083$ with the leverage effect around $\phi = -0.1$. This correlation coefficient estimate is similar to the value reported in Christoffersen et al. (2014).

Note that we implement a two-step estimation (plugging in the estimates from returns data to options data for option pricing) in this paper to simplify the optimization problem but this is inefficient. We can obtain an efficient estimation either by using the iterative method by Fan, Pastorello, and Renault (2015) or the efficient two-step method by Frazier and Renault (2016). Given that the IVRMSE decreases significantly when $\chi(\varsigma)$ is chosen to minimize the IVRMSE directly using options data, a two-step method the uses the information in returns and options simultaneously is expected to decrease option pricing errors.

5.6 Competitor model

The HEAVY-SV model is a discrete-time option pricing model. Thus, the natural competitors are a class of GARCH-type option pricing models. We consider here the ARV model by Christoffersen et al. (2014). It is a GARCH-type option pricing model where only the realized variance component plays a role in the variance dynamic of returns[15]. The ARV model assumes the following dynamic model of daily returns:

$$r_{t+1} = \left(\lambda_2 - \frac{1}{2}\right) h_t^{RV} + \sqrt{h_t^{RV}}\epsilon_{1,t+1},$$

[15]The ARV model is a special type of the GARV model (Christoffersen et al. (2014)) where the variance dynamic of returns depends both on realized variance and returns. Christoffersen et al. (2014) show that the GARV model outperforms the ARV model in terms of option pricing. But we use the ARV model for comparison since only realized variance is used for the variance of returns in the HEAVY-SV model.

where $h_t^{RV} = E\left[RV_{t+1}|F_t\right]$, and $\epsilon_{1,t+1}$ is a standard normal return shock. It also assumes the following affine structure of h_t^{RV}

$$h_{t+1}^{RV} = \omega_2 + \beta_2 h_t^{RV} + \alpha_2 \left(\epsilon_{2,t+1} - \gamma_2 \sqrt{h_t^{RV}}\right)^2,$$

where $\epsilon_{1,t+1}$ and $\epsilon_{2,t+1}$ follow a bivariate standard normal distribution with correlation ρ. Also,

$$RV_{t+1} = h_t^{RV} + \sigma \left[\left(\epsilon_{2,t+1} - \gamma_2 \sqrt{h_t^{RV}}\right)^2 - \left(1 + \gamma_2^2 h_t^{RV}\right)\right].$$

This model is estimated using quisi-maximum likelihood (QMLE) techniques. From the observations of returns and realized variance, the moments used for the estimation in addition to the first moment of realized variance are:

$$\begin{aligned}
E\left[r_{t+1}|F_t\right] &= \left(\lambda_2 - \frac{1}{2}\right) h_t^{RV} \\
Var\left[r_{t+1}|F_t\right] &= h_t^{RV} \\
Var\left[RV_{t+1}|F_t\right] &= 2\sigma^2 \left(1 + 2\gamma_2^2 h_t^{RV}\right) \\
Cov\left[r_{t+1}, RV_{t+1}|F_t\right] &= -2\rho\gamma_2\sigma h_t^{RV}.
\end{aligned}$$

We use the log of the bivariate normal distribution of returns and realized variance given in Christoffersen et al. (2014).

The estimation result is given in the third column of table 2. ω_1 is estimated using the unconditional variance formula[16]:

$$\omega_2 = E\left[h_t^{RV}\right]\left(1 - \beta_2 - \alpha_2\gamma_2^2\right) - \alpha_2.$$

where we first set $E\left[h_t^{RV}\right] = \frac{1}{T}\sum_{t=1}^{T} RV_t$.

It shows a high level of volaitlity persistence, 0.9595, as in the HEAVY-SV model. In the third column, the risk neutral γ_2, denoted by γ_2^*, is estimated from minimizing IVRMSE given other parameter estimates from the historical data of returns[17]. The price of risk of volatlity, $\mathcal{X}$, is then deduced from

$$\gamma_2^* = \gamma_2 - \mathcal{X}.$$

[16]Christoffersen et al. (2014) estimate the unconditonal mean of realized variance $E\left[h_t^{RV}\right]$ jointly with other parameters.

[17]We use a modification of the matlab code downloaded from http://christoffersen.com/cen to compute option prices for the ARV model.

5.7 Option pricing performance

In order see option pricing performance, we use European options written on the S&P500 index. The data were downloaded from Optionmetrics[18] and the observations range from January 3, 2000 to January 4, 2012. Following Barone-Adesi, Engle, and Mancini (2008) and Corsi et al. (2013), the options with time to maturity[19] less than 10 days or more than 360 days are dropped and we only consider the options for Wednesday to ensure that we use the liquid contracts. Also, the observations with option premiums less than \$0.05 and with an implied volatility of more than 70% are discarded. Moreover, we only consider call options. The same analysis can be done for put options as well. The total number of observations is 23,378.

We categorize options according to their time to maturity and moneyness. Following Khrapov and Renault (2016), we use $log(K/S_t)$ as a measure of moneyness where K and S_t denote a strike price and a price of the underlying asset at time t.

I estimate the prices of each option for given K, S_t and time to maturity following the steps described above for both ARG(1)-Normal with $H = 10$ and ARV models. In order to analyze the option pricing performances of each model, we use Root Mean Square Error on option prices ($RMSE_P$) and on the percentage implied volatility (IVRMSE):

$$RMSE_P = \sqrt{\sum_{i=1}^{N} \frac{\left(P_i^{hist} - P_i^{mod}\right)^2}{N}},$$

where N is the number of observations, P_i^{hist} and P_i^{mod} are the historical price and the model estimated price of the i-th option divided by the underlying price, respectively. The results are presented in table 3.

The first and second columns of table 3 present $RMSE_P$ and IVRMSE of each returns-volatility model. The first row shows the results of the ARG(1)-Normal model with $H = 10$ and $\alpha_T = 0.083$. The second row shows the same results excluding the observations between 2008 and 2009 which is the period of the recent recession. As figure 7 presents, volatility during this period dominates the graph. The third and fourth rows present the option pricing errors of the ARV model. The observations during the recent recession are excluded for the fourth row. For both models, the option pricing errors increase during the recent recession when volatility level was exceptionally high. The ARG(1)-Normal model outperforms ARV model in terms of both $RMSE_P$ and IVRMSE.

Table 4 shows the option pricing performances of the ARG(1)-Normal example and the ARV model for different time periods. ARG(1)-Normal[2] stands for the same case as appeared in table 2 and 3. Note that the early 2000s is the period of high volatility while the mid 2000s is the low volatility period (see figure 6). We see that the option pricing performs well for both models during

[18] We use zero-coupon yield curve and the index dividend yield provided by Optionmetrics in the pricing procedure.
[19] Calendar days

Table 3: Option pricing performances

Models	$RMSE_P$	$IVRMSE$
$ARG(1)$-Normal^2	0.009	4.1959
$ARG(1)$-Normal^3	0.0071	3.3065
ARV	0.0104	4.6056
ARV^1	0.0093	3.7089

* The observations between 2008 and 2009 are excluded for the results of $ARG(1)$-Normal^3 and ARV^1.

Table 4: Option pricing performances by dates

Date range	2000 to 2003	2000 to 2004	2000 to 2005	2000 to 2007
$ARG(1)$-Normal^2				
$RMSE_P$	0.0071	0.0066	0.0067	0.0068
IVRMSE	3.4092	3.1650	3.1224	3.1197
ARV				
$RMSE_P$	0.0058	0.0063	0.0080	0.0098
IVRMSE	3.2661	3.1107	3.3240	3.6116

* For each column, the date range is Jan 2000 to Dec of the last year included.
* ARG(1)-Normal model with $H = 10$ is used.

both periods of high and low volatility levels before the recent recession in 2008 and 2009.

Table 5 presents some descriptions of the options data and the $RMSE_P$ and IVRMSE for each maturity and moneyness category of the $ARG(1)$-Normal model and the ARV model. In terms of IVRMSE, the option pricing performance seems to be better for the option contracts in the long maturity groups. For different groups of moneyness, the option pricing seems to perform well for the contracts that are not relatively deep out-of-money (OTM) in terms of both IVRMSE and $RMSE_P$.

Table 5: Option pricing performances by maturity and moneyness

By Maturity	Less than 60	60 to 120	120 to 150	150 to 180	more than 180
No. of obs	5,125	3,264	3,524	3,617	7,848
Ave. premium	14.7429	21.76	26.01	29.33	37.68
Ave. IV(%)	19.81	18.94	18.72	18.70	18.66
$ARG(1)$-Normal^2					
$RMSE_P$	0.0061	0.0075	0.0084	0.0092	0.0112
IVRMSE	4.8084	4.1905	4.0318	3.9622	3.9383
ARV					
$RMSE_P$	0.0061	0.0080	0.0092	0.0104	0.0134
IVRMSE	4.8963	4.5356	4.4848	4.4969	4.5410

By Moneyness	Less than 3%	3% to 4%	4% to 6%	6% to 10%	More than 10%
No. of obs	3,817	3,343	5,821	6,657	3,944
Ave. premium	31.05	30.73	27.38	25.23	24.59
Ave. IV(%)	17.26	17.91	18.08	19.10	22.55
$ARG(1)$-Normal^2					
$RMSE_P$	0.008	0.0085	0.0083	0.0091	0.0112
IVRMSE	3.7124	3.8675	3.9009	4.2053	5.1791
ARV					
$RMSE_P$	0.0091	0.0092	0.01	0.0108	0.0120
IVRMSE	4.0077	4.1522	4.3028	4.6124	5.7678

* ARG(1)-Normal^2 with $H = 10$.
* Moneyness is $log(K/S)$.
* IV stands for Implied Volatility.

6 Conclusion

In this paper, we develop a discrete-time affine option pricing model that exploits the information in high-frequency data. This model has several attractive features such that it is robust to temporal aggregation, accommodates the leverage effect, and leads to homogeneous of degree one option pricing. In addition, the historical dynamics are maintained under the risk neutral measures with the exponentially affine stochastic discount factor which makes the model tractable. We provide a closed-form option pricing formula that is easy to compute. This model provides the dynamics of realized variance that nests GARCH-type HEAVY models as a special example.

This model is easy to estimate since conditional characteristic functions exist in closed-form. The empirical results show a high level of persistence of volatility, as in the HEAVY models. The leverage effect parameter is estimated to be negative as expected. The empirical results also show that the model performs relatively well even with the inclusion of the periods of the most recent recession in 2008 and 2009.

7 References

Andersen, Bollerslev, Diebold, and Labys, 2003, Modeling and forecasting realized volatility, Econometrica, 71, 579-625

Bandi and Reno, 2016, Price and volatility co-jumps, Journal of Financial Economics, 119 (1), 107-146

Barndorff-Nielsen and Shephard, 2002, Econometric analysis of realised volatility and its use in estimating stochastic volatility models, Journal of the Royal Statistical Society, 64 (B), 253-280

Barone-Adesi, Engle, and Mancini, 2008, A GARCH option pricing with filtered historical simulation, Review of Financial Studies, 21, 1223-1258

Bertholon, Monfort, and Pegoraro, 2008, Econometric asset pricing modelling, Journal of Financial Econometrics, 6, 407-458

Brockwell and Davis, 1991, Time series: theory and methods, second edition, Springer-Verlag, New York

Brownlees and Gallo, 2010, Comparison of volatility measures: a risk management perspective, Journal of Financial Econometrics, 2010, 8 (1), 29-56

Carrasco and Florens, 2000, Generalization of GMM to a continuum of moment conditions, Econometric Theory, 16, 797-834

Carrasco, Chernov, Florens, and Ghysels, 2007, Efficient estimation of general dynamic models with a continuum of moment conditions, Journal of Econometrics, 140, 529-573

Carrasco and Florens, 2014, On the asymptotic efficiency of GMM, Econometric Theory, 30, 372-406

Chamberlain, 1987, Asymptotic efficiency in estimation with conditional moment restrictions, Journal of Econometrics, 34, 305-334

Christoffersen, Elkamhi, Feunou, and Jacobs, 2010, Option valuation with conditional heteroskedasticity and nonnormality, Review of Financial Studies, 23 (5), 2139-2183

Christoffersen, Jacobs, and Ornthanalai, 2012, GARCH option valuation: theory and evidence, CREATES Research Paper No. 2012-50

Christoffersen, Feunou, Jacobs, and Meddahi, 2014, The Economic Value of Realized Volatility: Using High-Frequency returns for Option Valuation, Journal of Financial and Quantitative Analysis, 49 (3), 663-697

Corsi, Fusari, and La Vecchia, 2013, Realizing smiles: Options pricing with realized volatility, Journal of Financial Economics, 107, 284-304

Cox, Ingersoll, and Ross, 1985, A theory of the term structure of interest rates, Econometrica, 53 (2), 385-407

Darolles, Gourieroux, and Jasiak, 2006, Structural Laplace transform and compound autoregressive models, Journal of Times Series Analysis, 27 (4), 477-503

Dominguez and Lobato, 2004, Consistent estimation of models defined by conditional moment restrictions, Econometrica, 72, 1601-1615

Donald, Imbens, and Newey, 2003, Empirical likelihood estimation and consistent tests with conditional moment restrictions, Journal of Econometrics, 117, 55-93

Duan, 1995, The GARCH option pricing model, Mathematical Finance, 5 (1), 13-32

Duffie, Pan, and Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, Econometrica, 68 (6), 1343-1376

Fan, Pastorello, and Renault, 2015, Maximization by parts in extremum estimation, The Econometrics Journal, 18, 147-171

Feller, 1951, Two Singular Diffusion Problems, Annals of Mathematics, 54 (1), 173-182

Feuerverger and McDunnough, 1981, On the efficiency of empirical characteristic function procedures, Journal of the Royal Statistical Society, 43 (1), 20-27

Feunou and Tedongap, 2012, A stochastic volatility model with conditional skewness, Journal of Business and Economic Statistics, 30(4), 576-591

Frazier and Renault, 2016, Efficient two-step estimation via targeting, Forthcoming in Journal of Econometrics

Garcia and Renault, 1998, A note on hedging in ARCH and stochastic volatility option pricing models, Mathematical Finance, 8 (2), 153-161

Giot and Laurent, 2004, Modeling daily Value-at-Risk using realized volatility and ARCH type models, Journal of Empirical Finance, 11, 379-398

Gourieroux and Jasiak, 2006, Autoregressive gamma processes, Journal of Forecasting, 25 (2), 129-152

Han and Renault, 2016, Identification strength with a large number of moments, Working paper

Heston and Nandi, 2000, A closed-form GARCH option valuation model, The Review of Financial Studies, 13 (3), 585-625

Hsu and Kuan, 2011, Estimation of conditional moment restrictions without assuming parameter identifiability in the implied unconditional moments, Journal of Econometrics, 165, 87-99

Khrapov and Renault, 2016, Affine option pricing model in discrete time, working paper

Maheu and McCurdy, 2011, Do high-frequency measures of volatility improve forecasts of returns distributions?, Journal of Econometrics, 160, 69-76

Meddahi, 2003, ARMA representation of integrated and realized variances, The Econometrics Journal, 6 (2), 335-356

Meddahi and Renault, 2004, Temporal aggregation of volatility models, Journal of Econometrics, 119, 355-379

Noureldin, Shephard, and Sheppard, 2012, Multivariate high-frequency-based volatility (HEAVY) models, Journal of Applied Econometrics, 27, 907-933

Renault, 1997, Econometric Models of Option Pricing Errors. In D. M. Kreps & K. F. Wallis (Eds.), Advances in Economics and Econometrics: Theory and Applications, volume 3 chapter 8, 223-278, Cambridge University Press

Renault, Sekkat, and Szafarz, 1988, Testing for spurious causality in exchange rates, Journal of Empirical Finance, 5 (1), 47-66

Shephard and Sheppard, 2010, Realising the future: forecasting with high-frequency-based volatility (HEAVY) models, Journal of Applied Econometrics, 25, 197-231

Singleton, 2001, Estimation of affine asset pricing models using the empirical characteristic function, Journal of Econometrics, 102, 111-141

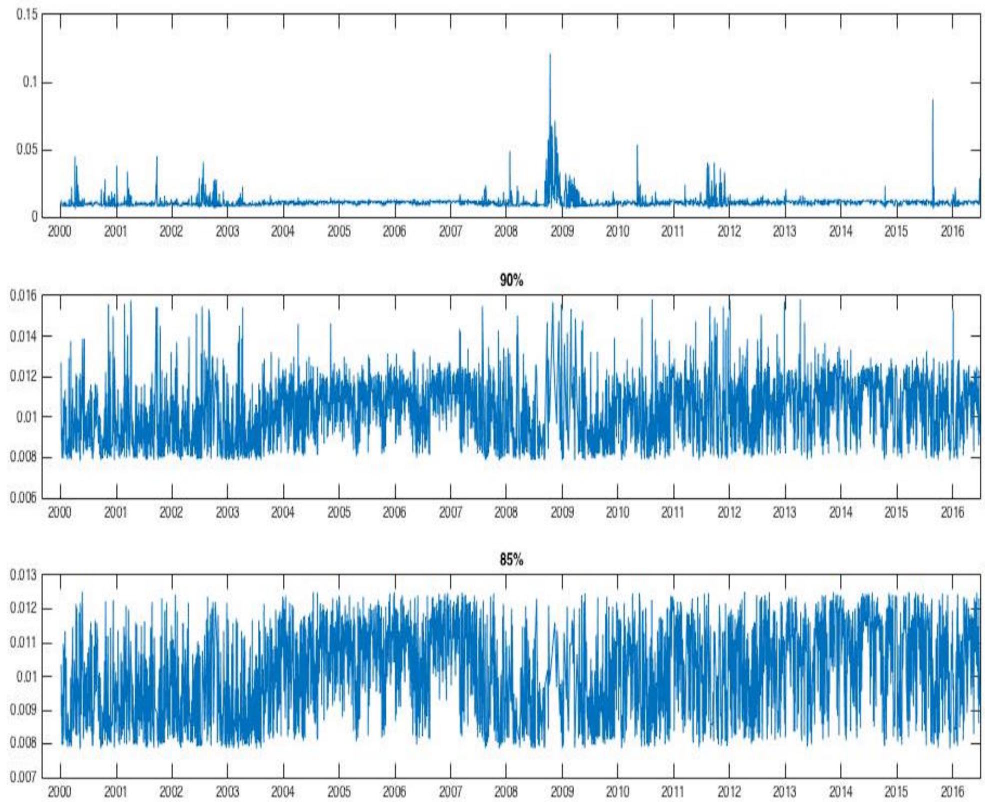
Stinchcombe and White, 1998, Consistent specification testing with nuisance parameters present only under the alternative, Econometric Theory, 14, 295-325

Vasicek, 1977, An equilibrium characterization of the term structure, Journal of Financial Economics, 5 (2), 177188

Appendix

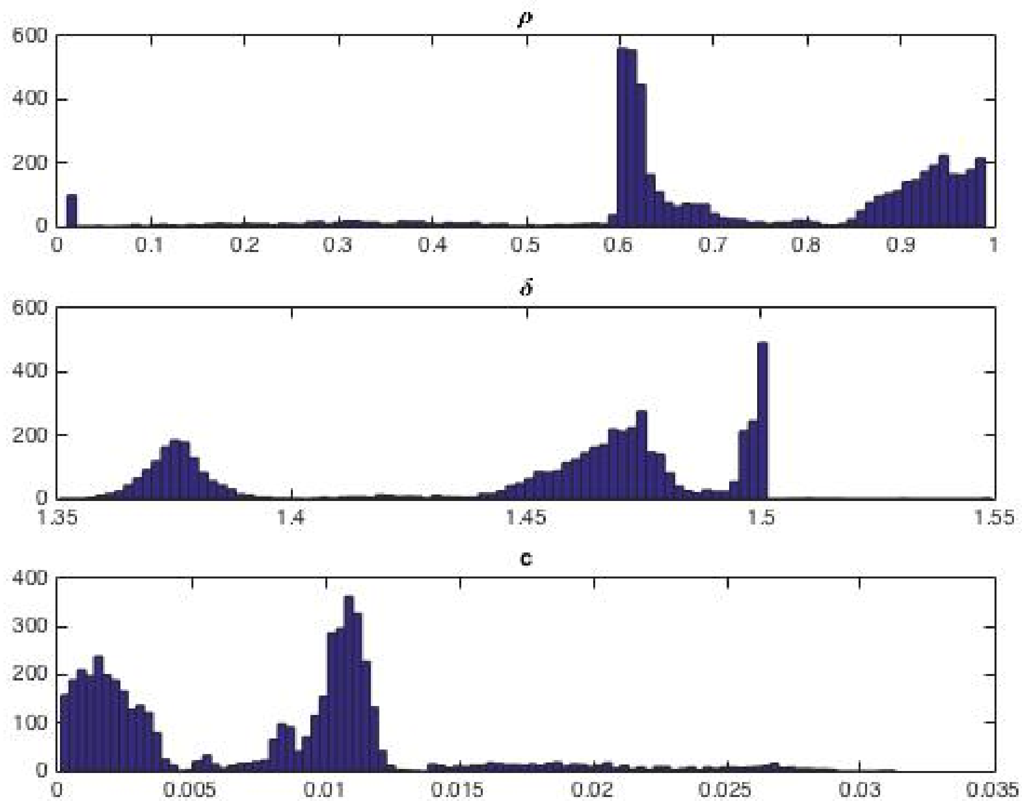
Appendix A

Figure 1: $\sqrt{Var\left[RV_{t+1}|F_t\right]/E\left[RV_{t+1}|F_t\right]}$



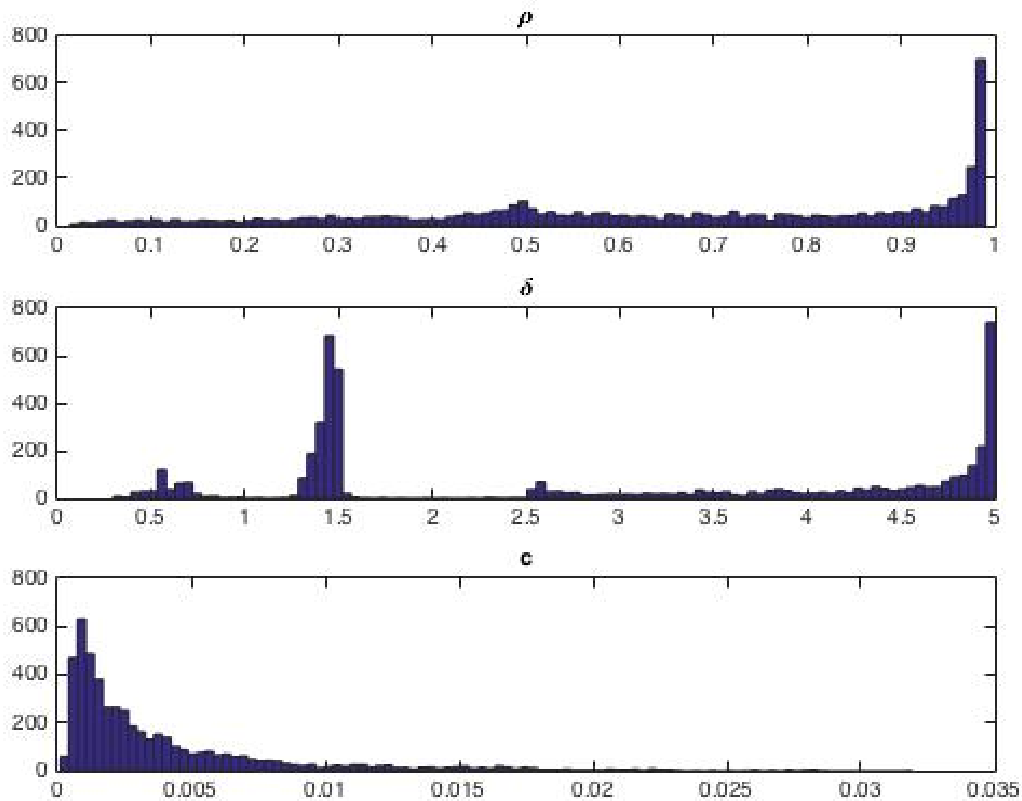
* $\sqrt{Var\left[RV_{t+1}|F_t\right]/E\left[RV_{t+1}|F_t\right]}$ is calculated by fitting AR(1) realized variance with ARCH(1).
* The first one includes the whole sample between Jan 2000 and Jun 2016.
* The second one excludes the 5% largest and 5% smallest values.
* The last one excludes the 10% largest and 5% smallest values.

Figure 2: Distribution of GMM estimators for ARG(1) volatility model with a constant instrument



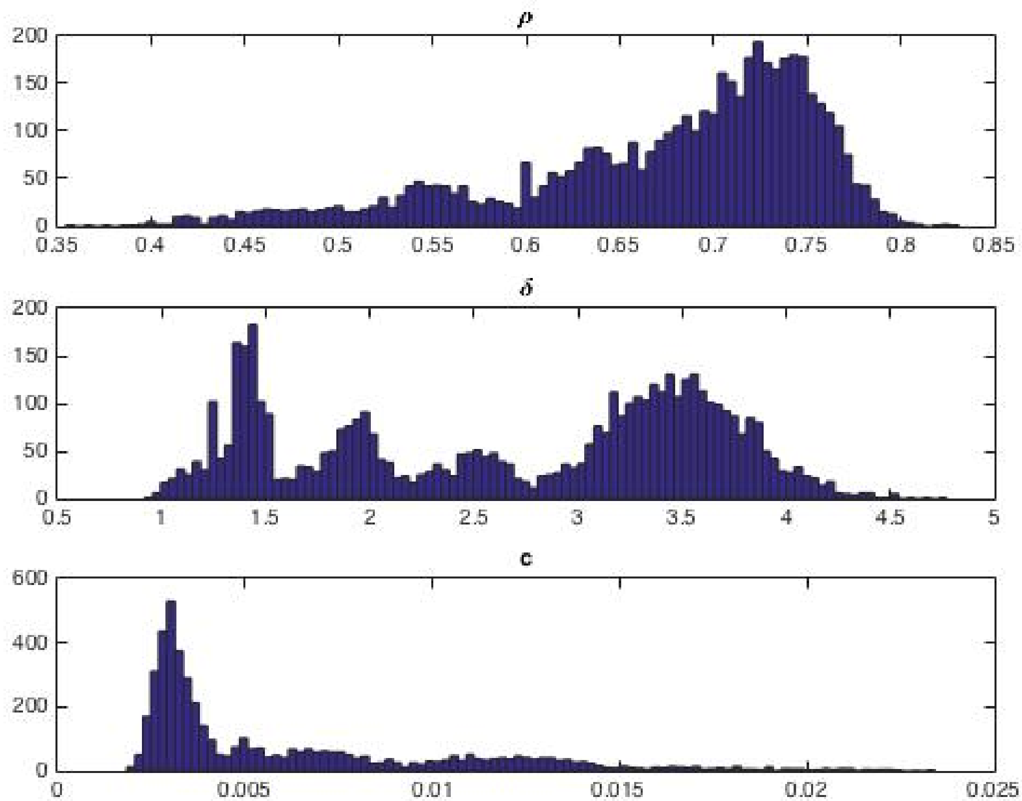
* The true values are: $(\rho^0 = 0.6, \delta^0 = 1.5, c^0 = 0.0106)$.
* We used 5 equally spaced u's on $[1i, 10i]$.
* An identity weighting matrix is used.
* 10 randomly generated values were used as initial values for each ρ, δ, and c
* 5000 replications

Figure 3: Distribution of GMM estimators for ARG(1) volatility model with the optimal instrument



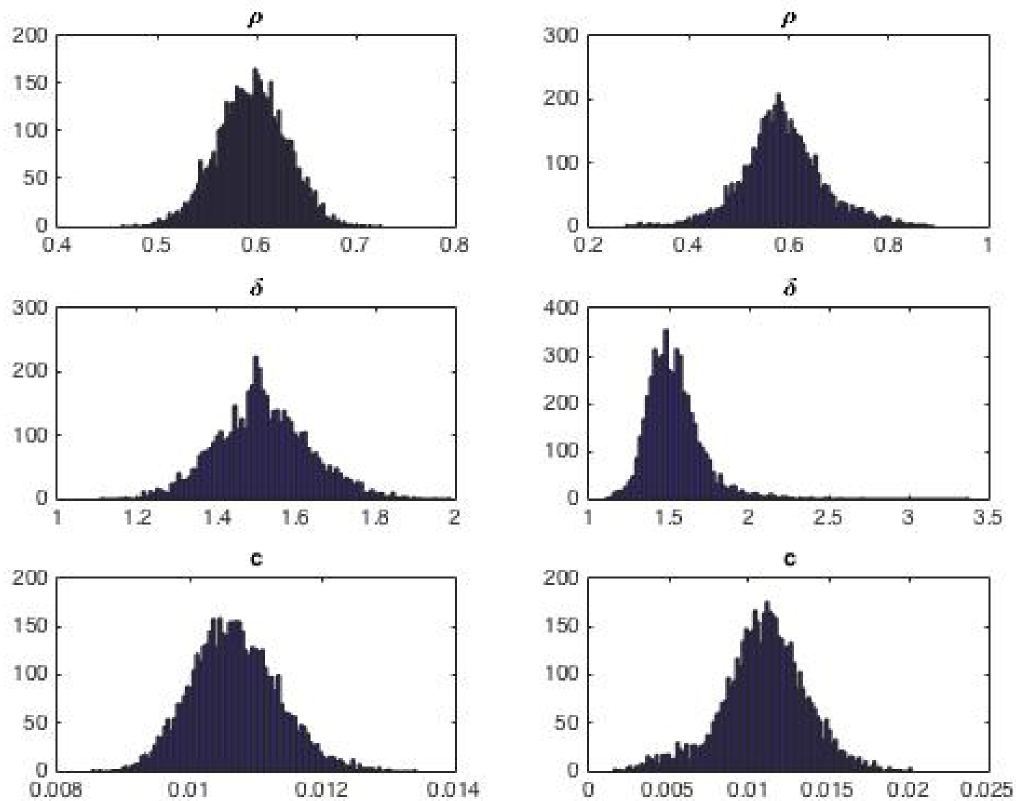
* The true values are: $(\rho^0 = 0.6, \delta^0 = 1.5, c^0 = 0.0106)$.
* We used 10 equally spaced u's on $[1i, 10i]$.
* Due to the singularity of the conditional variance, the Tikhonov regularization method is used with the regularization parameter equal to 0.01.
* δ is restricted to be between $(0, 5]$ to prevent it from getting too big.
* 10 randomly generated values were used as initial values for each ρ, δ, and c
* 5000 replications

Figure 4: Distribution of GMM estimators for ARG(1) volatility model with the SI instrument



* The true values are: $(\rho^0 = 0.6, \delta^0 = 1.5, c^0 = 0.0106)$.
* We used 5 equally spaced u's on $[1i, 10i]$.
* An identity weighting matrix is used.
* 10 randomly generated values were used as initial values for each ρ, δ, and c
* 5000 replications

Figure 5: Distributions of GMM estimators for ARG(1) volatility model with the DI instrument



* The true values are: $(\rho^0 = 0.6, \delta^0 = 1.5, c^0 = 0.0106)$.
* We used 5 equally spaced u's on $[1i, 10i]$.
* 5 equally spaced v's on $[1i, 10i]$ for the first column.
* $v = 2i$ for the second column.
* An identity weighting matrix is used.
* 10 randomly generated values were used as initial values for each ρ, δ, and c
* 5000 replications

Figure 6: Overidentification test

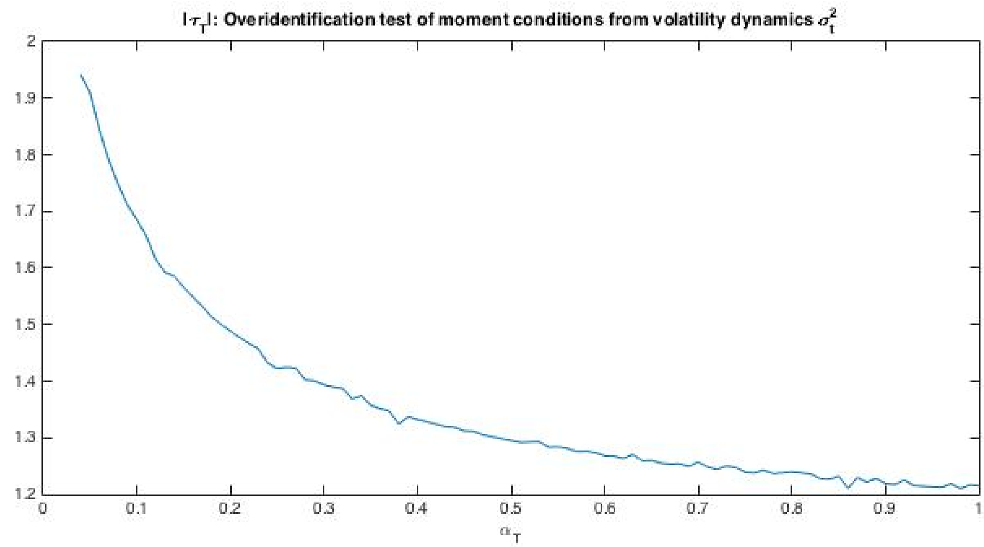
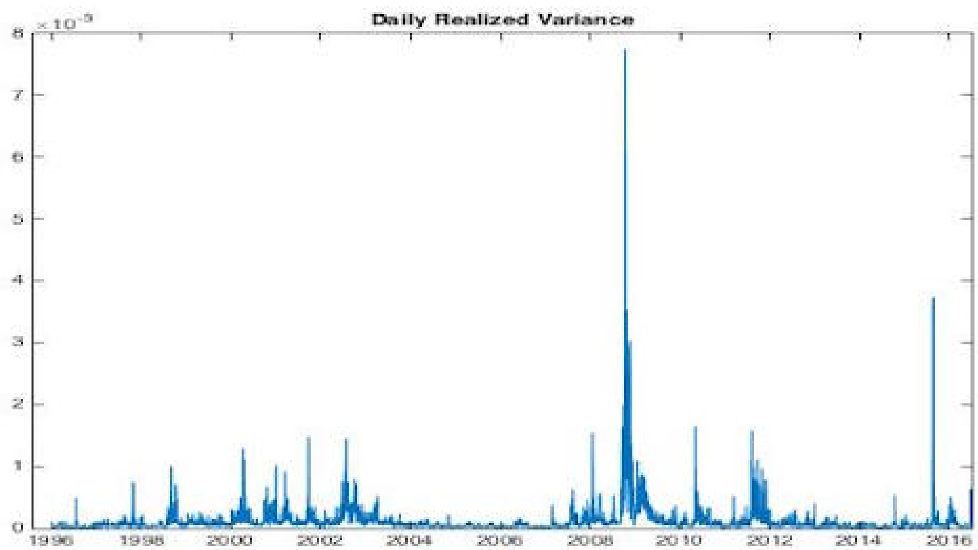


Figure 7: Daily realized variance



* I use the data used in Shephard and Sheppard (2010) to plot the RV from Jan 1996 to Dec 1999.
* I use the live data that is updated daily in the realized library to plot the RV from Jan 2000 to Jun 2012.

Appendix B

Proof of Proposition 1

Note that

$$\begin{aligned}
Corr\left[r_{t+1}, RV_{t+1}|F_t\right] &= \frac{Cov\left[r_{t+1}, RV_{t+1}|F_t\right]}{\sqrt{Var\left[r_{t+1}|F_t\right]}\sqrt{Var\left[RV_{t+1}|F_t\right]}} \\
&= \frac{\psi Var\left[\sigma_{t+1}^2|F_t\right]}{\sqrt{Var\left[r_{t+1}|F_t\right]}\sqrt{Var\left[\sigma_{t+1}^2|F_t\right]}} \\
&= \psi\frac{\sqrt{Var\left[\sigma_{t+1}^2|F_t\right]}}{\sqrt{E\left[RV_{t+1}|F_t\right]}} \\
&= \psi\frac{\sqrt{2\rho c\sigma_t^2 + \omega c^2}}{\sqrt{(\rho - B)\sigma_t^2 + (\delta c - D)}}
\end{aligned}$$

Then we get the constant leverage effect if and only if $\frac{2\rho c\sigma_t^2+\omega c^2}{(\rho-B)\sigma_t^2+(\delta c-D)}$ is constant over time. It is easy to see that this is equivalent to:

$$\frac{2\rho c}{\rho - B} = \frac{\omega c^2}{\delta c - D} \quad \Leftrightarrow \quad 2\rho(\delta c - D) = \omega c(\rho - B)$$

which is equivalent to:

$$2(1-f)\delta = \omega(1-e)$$

using B and D given in (2.10).

Then the leverage effect L is:

$$\begin{aligned}
L = \psi\left[\frac{2\rho c}{\rho - B}\right]^{1/2} &= \psi\left[\frac{\omega c^2}{\delta c - D}\right]^{1/2} \\
&= \psi\left[\frac{2c}{2c\psi^2 + (1-e)(1-\phi^2)}\right]^{1/2}
\end{aligned}$$

QED

Proof of Proposition 2

We know from the subsection 3.2:

$$\alpha^*(v) = \alpha(\varsigma_2 + v) - \alpha(\varsigma_2)$$
$$\beta^*(v) = \beta(\varsigma_2 + v) - \beta(\varsigma_2)$$
$$\gamma^*(v) = \gamma(\varsigma_2 + v) - \gamma(\varsigma_2)$$

and:

$$a^*(u) = a\left[u + \varsigma_1 + \alpha(\varsigma_2)\right] - a\left[\varsigma_1 + \alpha(\varsigma_2)\right]$$
$$b^*(u) = b\left[u + \varsigma_1 + \alpha(\varsigma_2)\right] - b\left[\varsigma_1 + \alpha(\varsigma_2)\right]$$

Since the returns dynamics functions are quadratic:

$$\begin{aligned}\psi^* v^2 - \frac{1}{2}(1-(\phi^*)^2)v^2 &= \psi v - (1-\phi^2)\varsigma_2 v + (1-\phi^2)v^2 \\ &= (\psi - (1-\phi^2)\varsigma_2)v + (1-\phi^2)v^2\end{aligned}$$

Then we can deduce that:

$$\psi^* = \psi - (1-\phi^2)\varsigma_2$$
$$\phi^* = \phi$$

Recall that $\beta''(0) = e\rho(1-\phi^2)$ and $\gamma''(0) = f\delta c(1-\phi^2)$. Then similarly we have:

$$(\beta^*)'(0) = \beta'(0) + e\rho(1-\phi^2)\varsigma_2$$
$$(e\rho)^* = e^*\rho^* = e\rho$$
$$(\gamma^*)'(0) = \gamma'(0) + f\delta c(1-\phi^2)\varsigma_2$$
$$(f\delta c)^* = f^*\delta^* c^* = f\delta c$$

We now move to the volatility dynamics. Note that

$$a^*(u) = \frac{\rho^* u}{1 + c^* u}, \qquad b^*(u) = \delta^* log(1 + c^* u)$$

Using the above specification of $b^*(u)$ we have:

$$\begin{aligned}
\delta^* log(1+c^*u) &= \delta log(1+c\left[u+\varsigma_1+\alpha(\varsigma_2)\right]) - \delta log(1+c\left[\varsigma_1+\alpha(\varsigma_2)\right]) \\
&= \delta log\left(\frac{1+c\left[u+\varsigma_1+\alpha(\varsigma_2)\right]}{1+c\left[\varsigma_1+\alpha(\varsigma_2)\right]}\right) \\
&= \delta log\left(1+\frac{cu}{1+c\left[\varsigma_1+\alpha(\varsigma_2)\right]}\right)
\end{aligned}$$

so we can deduce that:

$$\begin{aligned}
\delta^* &= \delta \\
c^* &= \frac{c}{1+c\left[\varsigma_1+\alpha(\varsigma_2)\right]} = c\mathcal{X}(\varsigma)^{-1}
\end{aligned}$$

Also using the above specification of $a^*(u)$, we have:

$$\begin{aligned}
\frac{\rho^* u}{1+c^*u} &= \frac{\rho\left[u+\varsigma_1+\alpha(\varsigma_2)\right]}{1+c\left[u+\varsigma_1+\alpha(\varsigma_2)\right]} - \frac{\rho\left[\varsigma_1+\alpha(\varsigma_2)\right]}{1+c\left[\varsigma_1+\alpha(\varsigma_2)\right]} \\
&= \frac{\rho u}{\left\{1+c\left[u+\varsigma_1+\alpha(\varsigma_2)\right]\right\}\left\{1+c\left[\varsigma_1+\alpha(\varsigma_2)\right]\right\}} \\
&= \frac{\rho u}{\left\{1+c\left[\varsigma_1+\alpha(\varsigma_2)\right]\right\}^2}\frac{1}{1+c^*u}
\end{aligned}$$

meaning:

$$\rho^* = \rho\mathcal{X}(\varsigma)^{-2}$$

From the risk neutral parameter specifications of ρ and c we can now also deduce that:

$$\begin{aligned}
e^* &= e\left\{\mathcal{X}(\varsigma)\right\}^2 \\
f^* &= f\mathcal{X}(\varsigma)
\end{aligned}$$

QED

Conditional characteristic function of $\sum_{i=1}^{T} r_{t+i}$ at time t

In this section, we derive the conditional moment generating function of $\sum_{i=1}^{T} r_{t+i}$ at time t,

$$f_{t,t+1}(u) = E\left[\exp\left(u\sum_{i=1}^{T} r_{t+i}\right)|F_t\right], \ \ \forall u \in \mathbb{C}.$$

for the leading example that is ARG(1) volatility and conditionally normally distributed returns.

By the law of iterated expectations, we get

$$f_{t,t+T}(u) = exp\left\{-G(u,T) - L(u,T)\sigma_t^2\right\}$$

where $G(u,T)$ and $L(u,T)$ are defined as follows.

1. $L(u,T)$

$$L(u,1) = l(0,-u) \text{ and } L(u,i) = l(L(u,i-1),-u) \text{ for all } i = 2,3,\cdots,T.$$

2. $G(u,T)$

$$G(u,T) = \sum_{i=1}^{T} G(u,i)$$

where

$$G(u,1) = g(0,-u) \text{ and } G(u,i) = g(L(u,i-1),-u) \text{ for all } i = 2,3,\cdots,T.$$

Proof of Proposition 3

We start from

$$\begin{aligned} C_t(x_t,T) &= \exp(-rT)\, E_t^*\left[max\{0, S_{t+T} - K\}\right] \\ &= S_t E_t^*\left[max\left\{0, \exp\left(\sum_{i=1}^{T} r_{t+i}\right) - \exp(x_t)\right\}\right], \end{aligned}$$

with

$$x_t = log(K/S_t) - rT.$$

Then

$$C_t(x_t,T) = S_t E_t^*\left[\exp\left(\sum_{i=1}^{T-t} r_{t+i}\right)\mathbb{1}\left\{\sum_{i=1}^{T-t} r_{t+i} > x_t\right\}\right] - K\exp(-rT)\, E_t^*\left[\mathbb{1}\left\{\sum_{i=1}^{T-t} r_{t+i} > x_t\right\}\right]$$

Let $P_t^*[\cdot]$ denote the probability under the risk-neutral measure. Note that

$$E_t^*\left[\mathbb{1}\left\{\sum_{i=1}^{T-t} r_{t+i} > x_t\right\}\right] = P_t^*\left[\sum_{i=1}^{T-t} r_{t+i} > x_t\right].$$

Then by the proof of Proposition 3 in Heston and Nandi (2000),

$$P_t^*\left[\sum_{i=1}^{T-t} r_{t+i} > x_t\right] = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{exp\left(-iux_t\right) f_{t,t+T}^*(iu)}{iu}\right] du$$

where $f_{t,t+T}^*(iu) = E^*\left[\exp\left(iu\sum_{i=1}^T r_{t+i}\right)|F_t\right]$. Also following the proof of Proposition 3 in Heston and Nandi (2000),

$$E_t^*\left[\exp\left(\sum_{i=1}^{T-t} r_{t+i}\right)\mathbb{1}\left\{\sum_{i=1}^{T-t} r_{t+i} > x_t\right\}\right] = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{\exp\left(-iux_t\right) f_{t,t+T}^*(iu+1)}{iuf_{t,t+1}^*(1)}\right] du.$$

Note that

$$\begin{aligned} f_{t,t+1}^*(1) &= E^*\left[\exp\left(\sum_{i=1}^{T} r_{t+i}\right) F_t\right] \\ &= E^*\left[\exp\left\{log\left(\frac{S_{t+T}}{S_t}\right) - rT\right\}|F_t\right] \\ &= \frac{1}{S_t}\exp\left(-rT\right)E^*\left[S_{t+T}|F_t\right] = 1, \end{aligned}$$

using the fact that $S_t = \exp\left(-rT\right)E^*\left[S_{t+T}|F_t\right]$ under the risk neutral distribution.

QED

Proof of Theorem 1

We will show that $\theta^0 = (\rho^0, \delta^0, c^0)'$ is not identified from the following equalities.

$$E\left[\exp(-u\sigma_{t+1}^2)\right] = E\left[\exp\left\{-a(u_1)^\theta\sigma_t^2 - b(u_1)^\theta\right\}\right] \tag{7.1}$$

for any $u \in \mathbb{C}$ and for any number of u's used.

Using the fact that the unconditional distribution of σ_t^2 is gamma with the shape parameter, δ^0, and the scale parameter, $\frac{c^0}{1-\rho^0}$, (7.1) is equivalent to:

$$\left(1 + \frac{c^0}{1-\rho^0}u_1\right)^{-\delta^0} = (1+cu_1)^{-\delta}\left\{1 + \frac{c^0}{1-\rho^0}\frac{\rho u_1}{1+cu_1}\right\}^{-\delta^0}$$

and this is rearranged to:

$$(1+cu)^{-\frac{(\delta-\delta^0)}{\delta^0}} = \left\{1 + \frac{(c-c^0)(1-\rho^0) + c^0(\rho-\rho^0)}{1-\rho^0+c^0u}u\right\} \tag{7.2}$$

for all $u \in \mathbb{C}$.

In order to see how identification fails, we will check that the above equation does not imply

$\theta = \theta^0$ even when $\delta = \delta^0$. When $\delta = \delta^0$, the equation (7.2) is:

$$1 = \left\{1 + \frac{(c-c^0)(1-\rho^0) + c^0(\rho - \rho^0)}{1-\rho^0 + c^0 u} u\right\} \Leftrightarrow (c-c^0)(1-\rho^0) + c^0(\rho-\rho^0) = 0$$

We can find multiple combinations of (ρ, c) that are the solutions to the RHS equations including the true one, (ρ^0, c^0). Since these (ρ, c) do not depend on u, having a multiple of u's does not help and the parameters are unidentified. This result is also shown in figure 1 in the appendix A.

QED

Proof of Theorem 2

Consider the following unconditional moment conditions:

$$E\left[\exp\left(-vX_t - uX_{t+1}\right)\right] = E\left[\exp\left\{-\left[a^\theta(u) + v\right]X_t - b^\theta(u)\right\}\right] \tag{7.3}$$

for $u, v \in \mathbb{C}$. By using the marginal distribution of X_t, (7.3) is equivalent to:

$$\left(1 + c^0 u\right)^{-\delta^0}\left(1 + \frac{c^0}{1-\rho^0}\left[\frac{\rho^0 u}{1+c^0 u} + v\right]\right)^{-\delta^0} = (1+cu)^{-\delta}\left(1 + \frac{c^0}{1-\rho^0}\left[\frac{\rho u}{1+cu} + v\right]\right)^{-\delta^0} \tag{7.4}$$

Using the Taylor expansion, for each j the LHS is:

$$\left(1 + c^0 u + \frac{c^0}{1-\rho^0}\left[\rho^0 u + v + c^0 uv\right]\right)^{-\delta^0} = \left(1 + \frac{c^0(u+v) + \left(c^0\right)^2 uv}{1-\rho^0}\right)^{-\delta^0}$$

since

$$\begin{aligned} \frac{\rho u}{1+cu} &= \rho u - \rho c u^2 + \rho c^2 u^3 - \rho c^3 u^4 + \cdots \\ &= \rho u \sum_{i=1}^{\infty} (-1)^{i-1} (cu)^{i-1} \end{aligned}$$

Case 1: $\mathbf{u} = \mathbf{v}$

First, let's consider the case where $u = v$ for each $u \in \mathcal{U} \in \mathbb{C}$ chosen to construct the moments. Then the LHS is equal to:

$$\left(1 + \frac{2c^0 u + \left(c^0\right)^2 u^2}{1-\rho^0}\right)^{-\delta^0}$$

Choose $\delta = 2\delta^0$. Also choose a small c so that $c^2u^2 \approx 0$ and $c^2u^3 \approx 0$. Then the RHS is (approximately) equal to:

$$\left(1+2cu+\frac{c^0}{1-\rho^0}\left[\rho u+u+2cu^2+\rho cu^2+c^2u^3\right]\right)^{-\delta^0}=\left(1+\frac{(2c-2c\rho^0+c^0+c^0\rho)u+cc^0(2+\rho)u^2}{1-\rho^0}\right)^{-\delta^0}$$

Then LHS equals RHS for all u if

$$\begin{aligned}2c-2c\rho^0+c^0+c^0\rho&=2c^0\\ cc^0(2+\rho)&=\left(c^0\right)^2\end{aligned}$$

We will now show that there exists $(\rho,c)\neq(\rho^0,c^0)$ such that the above equalities hold. From the first equality, we have:

$$\rho=1+\frac{2c(\rho^0-1)}{c^0}$$

Given that $\rho^0<0$, $\rho<1$. Also, $\rho>0$ if $c<\frac{c^0}{2(1-\rho^0)}$ which does not violate c being a very small positive number. Then from the second equality we have:

$$\left(c^0\right)^2=cc^0\left(3+\frac{2c(\rho^0-1)}{c^0}\right)\quad\Leftrightarrow\quad 2c^2(1-\rho^0)-3c^0c+\left(c^0\right)^2=0$$

Using the quadratic formula, we get the solution for c:

$$c=\frac{3c^0\pm\sqrt{9\left(c^0\right)^2-8(1-\rho^0)\left(c^0\right)^2}}{4(1-\rho^0)}=\frac{3c^0\pm c^0\sqrt{9-8(1-\rho^0)}}{4(1-\rho^0)}$$

We know that this solution of c exists since $9-8(1-\rho^0)>9-8=1>0$. If this c satisfies the condition that $c^2u^2\approx0$ and $c^2u^3\approx0$, then the unconditional moments (7.3) hold for $(\rho,\delta,c)\neq(\rho^0,\delta^0,c^0)$ for all $u\in\mathcal{U}\subset\mathbb{C}$.

Case 2: $\mathbf{u}\neq\mathbf{v}$

The RHS of (7.4) is:

$$\left((1+cu)^{\delta/\delta^0}\left\{1+\frac{c^0}{1-\rho^0}\left[\frac{\rho u}{1+cu}+v\right]\right\}\right)^{-\delta^0}\tag{7.5}$$

By the Taylor expansion at around $cu=0$, we have:

$$(1+cu)^{\delta/\delta^0}=1+\frac{\delta}{\delta^0}cu+\frac{1}{2!}\frac{\delta}{\delta^0}\left(\frac{\delta}{\delta^0}-1\right)(cu)^2+\frac{1}{3!}\frac{\delta}{\delta^0}\left(\frac{\delta}{\delta^0}-1\right)\left(\frac{\delta}{\delta^0}-2\right)(cu)^3+\cdots$$

which shows that (7.5) will have the terms involving u^i, $i \geq 1$. Especially, see that:

$$(1+cu)^{\delta/\delta^0} v = v + \frac{\delta}{\delta^0} cuv + \frac{1}{2!}\frac{\delta}{\delta^0}\left(\frac{\delta}{\delta^0} - 1\right)(cu)^2 v + \frac{1}{3!}\frac{\delta}{\delta^0}\left(\frac{\delta}{\delta^0} - 1\right)\left(\frac{\delta}{\delta^0} - 2\right)(cu)^3 v + \cdots$$

However, as shown before, the only terms of v that appear in the LHS of (7.4) are v and uv meaning that $\delta/\delta^0 = 1$ unless $(cu)^i = 0$ (or very close to 0) for all $i \geq 2$.

We first consider the case with $\delta/\delta^0 = 1$. Then $(1+cu)^{\delta/\delta^0} = 1 + cu$ and (7.5) is:

$$\left(1 + cu + \frac{c^0}{1-\rho^0}\left[\rho u + (1+cu)v\right]\right)^{-\delta^0} = \left(1 + \frac{(c - c\rho^0 + c^0\rho)u + c^0 v + c^0 cuv}{1-\rho^0}\right)^{-\delta^0}$$

which is equivalent to:

$$\begin{aligned} c - c\rho^0 + c^0\rho &= c^0 \\ c^0 c &= \left(c^0\right)^2 \end{aligned}$$

The only solution is that c and ρ equal to their true values:

$$c = c^0, \qquad \rho = \rho^0$$

Now we consider the case with a choice of c such that $(cu)^i \approx 0$ and $u(cu)^i \approx 0$ for all $i \geq 2$ for all u. Then (7.5) is (approximately):

$$\left(1 + \frac{\frac{\delta}{\delta^0} cu(1-\rho^0) + c^0\left[v + \frac{\delta}{\delta^0} cuv + \rho u + \left(\frac{\delta}{\delta^0} - 1\right)\rho c u^2\right]}{1-\rho^0}\right)^{-\delta^0}$$

Then we know $\delta = \delta^0$ is the only choice for the moments to hold for any u and v. Also,

$$\begin{aligned} c(1-\rho^0) + c^0\rho &= c^0 \\ cc^0 &= \left(c^0\right)^2 \end{aligned}$$

so that the only solution is:

$$c = c^0, \qquad \rho = \rho^0$$

QED

Proof of Theorem 3

We have $\zeta(J_T) = \sqrt{J_T}$. Following the notations in Donald et al (2003):

$$\tilde{\Omega}_T = \frac{1}{T}\sum_{t=1}^{T}\psi_{t,T}(\theta^0)\psi_{t,T}(\theta^0)', \quad \bar{\Omega}_T = \sum_{t=1}^{T}\Sigma(x_t)\otimes q^J(x_t)q^J(x_t)', \quad \Omega_T = E\left[\psi_{t,T}(\theta^0)\psi_{t,T}(\theta^0)'\right]$$

where

$$\Sigma(x_t) = E\left[\rho_t(u,\theta^0)\rho_t(u,\theta^0)'|x_t\right]$$

By Lemma A.6 in Donald et al (2003),

$$\left\|\hat{\Omega}_T - \tilde{\Omega}_T\right\| = O_p\left(\frac{J_T}{\sqrt{T}}\right), \quad \left\|\tilde{\Omega}_T - \bar{\Omega}_T\right\| = O_p\left(\frac{J_T}{\sqrt{T}}\right), \quad \left\|\bar{\Omega}_T - \Omega_T\right\| = O_p\left(\frac{J_T}{\sqrt{T}}\right)$$

Let

$$\tilde{W}_T = (\tilde{\Omega}_T^2 + \alpha_T I_J)^{-1}\tilde{\Omega}_T, \quad \bar{W}_T(\bar{\Omega}_T^2 + \alpha_T I_J)^{-1}\bar{\Omega}_T, \quad W_T = (\Omega_T^2 + \alpha_T I_J)^{-1}\Omega_T$$

Recall that $J = J_T = J_1 \times J_{2,T}$ where J_1 is the dimension of $\rho(\cdot,\theta)$. Then it suffices to show that[20]

$$\left\|\hat{W}_T - W_T\right\| = O_p\left(J_T/\sqrt{T\alpha_T}\right)$$

We now proceed to prove this. We know that

$$\begin{aligned} W_T^{1/2} &= (\Omega_T^2 + \alpha_T I_J)^{-1/2}\Omega_T^{1/2} \\ \hat{W}_T^{1/2} &= (\hat{\Omega}_T^2 + \alpha_T I_J)^{-1/2}\hat{\Omega}_T^{1/2} \end{aligned}$$

with $W_T = W_T^{1/2}W_T^{1/2}$ and $\hat{W}_T = \hat{W}_T^{1/2}\hat{W}_T^{1/2}$. Then

$$\begin{aligned} \left\|\hat{W}_T - W_T\right\| &= \left\|W_T^{1/2}W_T^{1/2} - \hat{W}_T^{1/2}\hat{W}_T^{1/2}\right\| \\ &\le \left\|\left(\hat{W}_T - W_T\right)\hat{W}_T^{1/2}\right\| + \left\|W_T^{1/2}\left(\hat{W}_T - W_T\right)\right\| \\ &\le \left\|\hat{W}_T - W_T\right\|\left\|\hat{W}_T^{1/2}\right\| + \left\|W_T^{1/2}\right\|\left\|\hat{W}_T - W_T\right\| \end{aligned}$$

[20] By the same argument given in the proof, $\left\|\hat{W}_T - \tilde{W}_T\right\| = O_p\left(J_T/\sqrt{T\alpha_T}\right)$, $\left\|\tilde{W}_T - \bar{W}_T\right\| = O_p\left(J_T/\sqrt{T\alpha_T}\right)$, $\left\|\bar{W}_T - W_T\right\| = O_p\left(J_T/\sqrt{T\alpha_T}\right)$.

First,

$$\left\| W_T^{1/2} \right\|^2 = \lambda_{max}\left((\Omega_T^2 + \alpha_T I_J)^{-1/2} \Omega_T^{1/2} (\Omega_T^2 + \alpha_T I_J)^{-1/2} \Omega_T^{1/2} \right)$$
$$\leq \lambda_{max}\left((\Omega_T^2 + \alpha_T I_J)^{-1/2} \right) \lambda_{max}\left(\Omega_T^{1/2} (\Omega_T^2 + \alpha_T I_J)^{-1/2} \Omega_T^{1/2} \right)$$
$$\leq \frac{1}{\sqrt{\alpha_T}}$$

since $\lambda_{max}\left(\Omega_T^{1/2} (\Omega_T^2 + \alpha_T I_J)^{-1/2} \Omega_T^{1/2} \right) \leq 1$. Also,

$$\left\| \hat{W}_T^{1/2} - W_T^{1/2} \right\| = O_p\left(\frac{J_T}{\sqrt{T}} \frac{1}{\alpha^{3/4}} \right)$$

by proof of Theorem 7 in Carrasco and Florens (2000) and $\left\| \hat{W}_T - W_T \right\| = O_p(J_T/\sqrt{T})$ by Lemma A.6 in Donald et al (2003). Then

$$\left\| \hat{W}_T - W_T \right\| = O_p\left(\frac{J_T}{\sqrt{T}} \frac{1}{\alpha_T} \right) = o_p(1)$$

by the assumption. Then we have the desired result.

QED

Proof of Theorem 4

Let

$$\hat{G}_T = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \rho(y_t, \hat{\theta}_T)}{\partial \theta'} \otimes q^J(x_t), \quad \bar{G}_T = \frac{1}{T} \sum_{t=1}^{T} D(x_t) \otimes q^J(x_t), \quad G_T = E\left[D(x) \otimes q^K(x) \right]$$

Since Theorem 5.4 in Donald et al (2003) presents the semiparametric efficiency result, we just need to modify the proof of Theorem 5.4 using $\left\| \hat{W}_T^{1/2} \right\| = O_p(\alpha_T^{-1/4})$, $\left\| \tilde{W}_T^{1/2} \right\| = O_p(\alpha_T^{-1/4})$, $\left\| \bar{W}_T^{1/2} \right\| = O_p(\alpha_T^{-1/4})$ and $\left\| W_T^{1/2} \right\| = O_p(\alpha_T^{-1/4})$.

We first need to show

$$\left\| \hat{G}_T' \hat{W}_T \hat{G}_T - \hat{G}_T' \bar{W}_T \hat{G}_T \right\| = o_p(1)$$

This holds since

$$\begin{aligned}\left\|\hat{G}_T'\hat{W}_T\hat{G}_T - \hat{G}_T'\bar{W}_T\hat{G}_T\right\| &= \left\|\hat{G}_T'\hat{W}_T^{1/2}\hat{W}_T^{1/2}\hat{G}_T - \hat{G}_T'\bar{W}_T^{1/2}\bar{W}_T^{1/2}\hat{G}_T\right\| \\ &\leq \left\|\hat{G}_T'\hat{W}_T^{1/2}\right\|\left\|(\hat{W}_T^{1/2} - \bar{W}_T^{1/2})\hat{G}_T\right\| + \left\|\hat{G}_T'(\hat{W}_T^{1/2} - \bar{W}_T^{1/2})\right\|\left\|\bar{W}_T^{1/2}\hat{G}_T\right\| \\ &= o_p(1)\end{aligned}$$

by $\left\|\hat{W}_T^{1/2} - \bar{W}_T^{1/2}\right\| = O_p(J_T/\sqrt{T}\alpha_T^{3/4})$, $\left\|\hat{W}_T^{1/2}\right\| = O_p(\alpha_T^{-1/4})$, and $\left\|\bar{W}_T^{1/2}\hat{G}_T\right\| = O_p(1)$.

Then $\hat{G}_T'\hat{W}_T\hat{G}_T \xrightarrow{p} V^{-1}$.

We also need to show

$$\left\|(\tilde{G}_T'\hat{W}_T - \bar{G}_T'\bar{W}_T)\bar{\psi}_T(\theta^0)\right\| = o_p(1)$$

where $\tilde{G}_T = \frac{1}{T}\sum_{t=1}^T \frac{\partial\rho(y_t,\theta^0)}{\partial\theta'} \otimes q^J(x_t)$. By Donald et al (2003), $\left\|\tilde{G}_T - \bar{G}_T\right\| = o_p(1)$. Also $\left\|\hat{W}_T\bar{\psi}_T(\theta^0)\right\| \leq \frac{1}{\alpha_T^{1/2}}\left\|\bar{\psi}_T(\theta^0)\right\| = O_p\left(\frac{1}{\sqrt{\alpha_T}}\frac{\sqrt{J_T}}{\sqrt{T}}\right)$. Then

$$\begin{aligned}\left\|(\tilde{G}_T'\hat{W}_T - \bar{G}_T'\bar{W}_T)\bar{\psi}_T(\theta^0)\right\| &\leq (\left\|\tilde{G}_T - \bar{G}_T\right\| + \left\|\bar{G}_T'\bar{W}_T(\hat{W}_T - \bar{W}_T)\right\|)\left\|\hat{W}_T\bar{\psi}_T(\theta^0)\right\| \\ &= O_p(J_T/\sqrt{T}\alpha_T)O_p(\sqrt{J_T}/\sqrt{T\alpha_T}) = o_p(1)\end{aligned}$$

Then the result holds by Theorem 5.4 of Donald et al (2003).

QED

Proof of Corollary 1

From section 5.2 in Singleton (2001)[21],

$$V^{-1} = E\left[D(X)'\Sigma(X)^{-1}D(X)\right] \to I(\theta^0)^{-1}$$

as $J_{1,T} \to \infty$. Then, by Brockwell and Davis (1991), we may conclude that:

$$\sqrt{T}(\hat{\theta}_T - \theta^0) \xrightarrow{d} \mathcal{N}(0, I(\theta^0)^{-1})$$

QED

[21] The main idea of the proof in Singleton (2001) is that there exists a continuum (in u) of instrument:

$$\frac{1}{2\pi}\int \exp\left(-ux\right)\frac{\partial ln f_{\theta^0}}{\partial\theta}\left(x|X_t\right)dx$$

that leads to the MLE efficiency (Feuerverger and McDunnough (1981)). Then by using the optimal instrument by Hansen (1985) with an increasing number of u's, the same efficiency result can be attained.