The leverage effect puzzle revisited: identification in discrete time

Hyojin Han* Stanislav Khrapov† Eric Renault‡

July 31, 2018

Abstract

The term "leverage effect", as coined by Black (1976), refers to the tendency of an asset's volatility to be negatively correlated with the asset's return. Ait-Sahalia, Fan, and Li (2013) refer to the "leverage effect puzzle" as the fact that, in spite of a broad agreement that the effect should be present, it is hard to identify empirically. For this purpose, we propose an extension with leverage effect of the discrete time stochastic volatility model of Darolles, Gourieroux, and Jasiak (2006). This extension is shown to be the natural discrete time analog of the Heston (1993) option pricing model. It shares with Heston (1993) the advantage of structure preserving change of measure: with an exponentially affine stochastic discount factor, the historical and the risk neutral models belong to the same family of joint probability distributions for return and volatility processes. This allows computing option prices in semi-closed form through Fourier transform. The discrete time approach has several advantages. First, it allows relaxing the constraints on higher order moments implied by the affine specification of a diffusion process. Second, it makes more transparent the role of various parameters: leverage versus volatility feedback effect, connection with daily realized volatility measures on high-frequency intraday returns and with HEAVY-GARCH models, impact of leverage on the volatility smile, etc. Even more importantly it sheds some new light on the identification issue of the various risk premium parameters. The price of volatility risk is identified from underlying asset return data, even without option price data, if and only if leverage effect is present.

1 Introduction

The term "leverage effect", as coined by Black (1976), refers to the tendency of an asset's volatility to be negatively correlated with the asset's return. Ait-Sahalia et al. (2013) (ASFL henceforth) refer to the "leverage effect puzzle" as the fact that, in spite of a broad agreement that the effect should

*Chow Center, Department of Statistics at SOE, Xiamen University, hyojin_han@xmu.edu.cn

†New Economic School, khrapovs@gmail.com

‡Department of Economics, Brown University, eric_renault@brown.edu

be present, “at high frequency and over short horizons, the estimated correlation between the asset returns and changes in its volatility is close to zero, instead of the strong negative value that we have come to expect”. Several authors, including not only ASFL but also Bollerslev, Litvinova, and Tauchen (2006) (BLT henceforth) as well as Bandi and Reno (2016) have argued that it takes inference on continuous time models with high frequency data for a proper identification of the leverage effect. As stressed by ASFL, ”the latency of the volatility variable is partly responsible for the observed puzzle”. Otherwise, with discrete time observations, using option-implied volatility in place of historical volatilities may not be an answer since these implied volatilities have a complicated relationship (involving averaging over the lifetime of the option, risk premium, market expectations, etc.) with actual latent stochastic volatility. The starting point of this paper is to use instead the theory of the volatility smile to identify a significant leverage effect through the shape of this smile (and not through the actual level of implied volatility). Following an argument formerly sketched by Renault and Touzi (1996), Renault (1997) and Garcia, Luger, and Renault (2003, 2005), we start from a very general model for risk neutral distribution of underlying asset return on short horizons that allows us to accommodate asymmetric volatility smiles that are the signal of leverage effect. The discrete time framework may look similar to the popular GARCH option pricing models (Duan (1995), Garcia and Renault (1998), Heston and Nandi (2000)) but the key difference is precisely that we must ensure ”latency of the volatility variable”.

For this purpose, our modelling strategy must rather be seen as a discrete time extension of affine diffusion models. Affine Jump-Diffusion models have been put forward by Duffie, Pan, and Singleton (2000) as a convenient model for state variables to get closed- or nearly-closed form expressions for derivative asset prices. Their model nests in particular the popular Cox, Ingersoll, and Ross (1985) model for interest rates as well as Heston (1993) stochastic volatility model for currency and equity prices for the purpose of option pricing.

Since then, Affine Jump-Diffusion models have often been criticized for their poor empirical fit. The key intuition is that they maintain an assumption of local conditional normality, up to jumps. Jumps are to some extent the only degree of freedom to reproduce the pattern of time-varying skewness and excess kurtosis commonly observed in asset returns. As a response to this criticism, at least two strands of literature have promoted specifications of discrete time models that remain true as much as possible to the affine structure. The goal is to use the additional degree of freedom provided by discrete time modeling to get a better empirical fit of higher order moments while keeping closed- or nearly-closed form expressions for securities prices. While Duan (1995), Heston and Nandi (2000) have initiated a strand of literature on closed-form GARCH option pricing (see Christoffersen, Elkamhi, Feunou, and Jacobs (2010); Christoffersen, Jacobs, and Ornthanalai (2013), and references therein for the most recent contributions), the paper by Darolles et al. (2006) has been seminal to provide a class of discrete time affine stochastic volatility models that nests the class of

Affine Jump-Diffusion models. The so-called "Compound Autoregressive" (CAR henceforth) model is defined from conditional moment generating functions that, in the continuous time limit, are consistent with affine diffusion models.

The stochastic volatility model provides a versatile framework to capture asymmetric volatility dynamics with possibly different parameters for historical and risk-neutral dynamics. While a similar exercise has been performed by Barone-Adesi, Engle, and Mancini (2008) in a GARCH framework (thanks to calibration of option prices data), Meddahi and Renault (2004) have shown that affine discrete-time volatility dynamics may be seen as a relevant weakening of the GARCH restrictions. This weakening restores robustness to temporal aggregation, at least for the affine specification of the first two moments.

However, Meddahi and Renault (2004) approach is only semi-parametric while a complete specification of the conditional probability distributions is called for option pricing. CAR models of Darolles et al. (2006) provide exactly the relevant framework for doing so. However, the focus is only on volatility dynamics and there is no attempt to specify a joint model for volatility and return process, incorporating the leverage effect as in particular in Heston (1993) model. Bertholon, Monfort, and Pegoraro (2008) move in the direction of joint return and volatility modeling within CAR-type framework. As an example, they develop the model with asymmetric GARCH volatility to produce the leverage effect. Our modelling approach pertains more generally to the Factorial Hidden Markov paradigm to accommodate different components of volatility dynamics. Ideally, one would like to follow the guidance of Augustyniak, Bauwens, and Dufays (2018) to capture both jumps in volatility and its predictive behavior through leverage and volatility feedback effects.

The focus of interest of this paper is to extend the framework of Darolles et al. (2006) to a bivariate model of return and volatility that allows for leverage effect and volatility feedback as well. This provides a convenient large class of affine models for option pricing, nesting Heston (1993) model as a particular continuous time limit. Moreover, by contrast with the debates about the right way to define continuous time limits of GARCH models, our limit arguments are underpinned by temporal aggregation formulas and as such, are immune to the criticism of ad hoc specification.

The challenge to provide a versatile discrete time extension of Heston (1993) option pricing with stochastic volatility and leverage effect is twofold:

First, the discrete time approach complicates the separate identification of Granger causality and instantaneous causality (see e.g. Renault, Sekkat, and Szafarz (1998)). This is especially important in the context of stochastic volatility models since, as documented by Bollerslev et al. (2006), the only way to disentangle leverage effect (as defined by Black (1976) from volatility feedback due to risk premium, is to assess the direction of causality between volatility and return. While Bollerslev et al. (2006) enhanced the usefulness of high frequency data to do so, our parametric modeling must carefully leave room for a mixture of these two effects in discrete time. Note that, on the other hand, we maintain the assumption that returns do not Granger cause volatility. This assumption is key (see Renault (1997)) to get option pricing formulas which, like Black and Scholes are homogeneous of degree one with respect to underlying stock price and strike price and as a result, allow us to see the volatility smile as a function of moneyness. The lack of such homogeneity property is another weakness of GARCH option pricing (see Garcia and Renault (1998)).

Second, we want to keep in discrete time the main features of Heston (1993), namely volatility dynamics that are affine for both the historical and the risk-neutral distribution, while keeping the same leverage effect. To the best of our knowledge, the only attempt to do so in the extant literature has been recently proposed by Feunou and Tedongap (2012). However, we note that their affine specification with leverage effect cannot work simultaneously for the historical and the risk neutral distribution, More precisely, a general exponentially affine pricing kernel is not structure preserving in their context. They can use their model either for risk neutral distribution or for the historical one, but not both. Our specification is structure preserving (while keeping the same leverage effect) with a general exponential affine stochastic discount factor. While the shape of volatility smile without leverage effect is well-known (see Renault and Touzi (1996)) our closed form expressions allow us to give new insights on distortions of volatility smiles produced by leverage. Moreover, these formulas also provide conditional moment restrictions for econometric inference as a discrete time extension of the work of Pan (2002). This also paves the way for an extension of Gagliardini, Gourieroux, and Renault (2011) to general discrete time affine models with leverage effect.

The rest of the paper is organized as follows. Section 2 proposes a very general characterization of the shape of volatility smiles through a risk neutral distribution seen as mixture of log-normal distributions. It is shown that only latent state variables may accommodate non-flat smiles while the skewness of the smile is tantamount to leverage effect. Section 3 introduces an exponential affine stochastic discount factor and characterizes the shape preserving transformation from risk neutral distribution to historical distribution. This leads us to address the tightly related identification issues of prices to risk (return risk and volatility risk) and of leverage effect. In particular, it leads to a generalized Black-Scholes option price formula, and analyzes the effect of leverage on implied volatilities. Section 4 proposes a fully parametric model for the distributions described above. This model is shown to be a discrete time version of Heston model and we show how intraday data on realized variance must be used for its statistical identification. Our use of realized variance to filter conditional variance can be seen as a stochastic volatility extension of the HEAVY-GARCH model of Shephard and Sheppard (2010). From these return data and the filtered volatilities, we devise in Section 5 a general two-step GMM estimation strategy based on the conditional moment generating functions. Section 6 provides several numerical illustrations, first on simulated data and second on daily log returns and realized volatilities of the S&P500 over 16 years starting in January 2000. Monte Carlo experiments confirm that our model captures the stylized facts about volatility smile and make a comparison with estimation of Heston Nandi's GARCH model. We use the S&P500 data to assess the accuracy of our model specification and to check that it delivers sensible values of estimated parameters. Section 7 concludes. Two sections of appendix provide the mathematical proofs of theoretical results and the numerical evidence, both through tables and graphical representations.

2 Volatility smile and latency of the volatility variable

2.1 A conditionally log-normal risk-neutral model

Let S_t stand for the time t price of the underlying asset, say a stock, of the option contracts of interest. The observed time series will be the continuously compounded rate of returns $r_t, t = 1, ..., T$ in excess of the risk free rate $r_{f,t}$ over the period $[t, t+1]$:

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

The maturity $t+1$ for investments at time t must typically be understood as a short horizon, say a day. For a short enough horizon, it is then very little restrictive to assume that given some possibly latent information set $J(t)$, the (log) return r_{t+1} is Gaussian. In other words, the conditional distribution of log-returns given some possibly unobserved mixture components is Gaussian. This can be seen as the discrete time implication of the local Gaussianity of diffusion processes with continuous paths. In particular, Mykland and Zhang (2009) show that an insightful way of thinking about inference in the context of high frequency data is to consider that returns have a constant variance and are conditionally Gaussian over small blocks of consecutive observations. Irrespective of this inference strategy, it is well known (see e.g. Garcia et al. (2005) and references therein) that mixtures of Gaussian distributions are a very versatile way to accommodate any observed pattern of time varying conditional variance, skewness, kurtosis and any other distributional characteristics of interest. We will actually be even less restrictive, at least in this section, by only assuming that this convenient conditional Gaussianity is fulfilled by the risk neutral distribution $\pounds^*(r_{t+1} \left| J(t)\right.)$ at stake for the purpose of option pricing. Therefore, with obvious notations, we maintain throughout the following assumption:

$$\pounds^*(r_{t+1} \left| J(t)\right.) = \aleph\left(\mu\left[J(t)\right], \sigma^2\left[J(t)\right]\right)$$

Note that we will use throughout the subscript * to mean that (conditional) probability distributions , their expectations, variances, etc., are computed with the risk neutral distribution.

A maintained assumption throughout the paper will be that past and current returns $r_\tau, \tau \leq t$, belong to the information $I(t)$ observed at time t, with $I(t) \subset J(t)$. By contrast, the (risk neutral) conditional distribution of $J(t)$ given $I(t)$ is assumed to be independent of the value of past and current returns . Therefore, as in common stochastic volatility models, all the serial dependence between consecutive returns goes through some state variables while returns are serially independent given these state variables (see Renault (1997) and Section 3 below for a more formal set up). We actually show in Section 4 that the standard continuous time option pricing model with stochastic volatility Heston (1993) can be seen as a continuous time limit of our setup.

2.2 Short maturity options

The key message of this section is that volatility smiles (for short horizon options) cannot be accommodated if the conditional information $J(t)$ is available to the representative investor on the market. To see that, note that if $J(t)$ is fully observed at time t, we should have:

$$E^{*}[\exp(r_{t+1})\,|J(t)] = \exp\left(\mu\left[J(t)\right] + \frac{\sigma^{2}\left[J(t)\right]}{2}\right) = 1$$

and then, option prices a time t are conformable to the Black and Scholes option pricing formula. For instance, for an European call with strike price K, the option price is nothing but:

$$C_t(K) = BS_{(t)}\left(K, S_t, \sigma^2\left[J(t)\right]\right) \tag{2.1}$$

with:

$$\begin{aligned}
BS_{(t)}\left(K,\sigma^2\right) &= S_t\Phi\left[d_{1,t}\left(K,\sigma^2\right)\right] - Ke^{-r_{f,t}}\Phi\left[d_{2,t}\left(K,\sigma^2\right)\right] \\
d_{1,t}\left(K,\sigma^2\right) &= \frac{1}{\sigma}\left[\log\left(S_t/K\right) + r_{f,t}\right] + \frac{\sigma}{2} \\
d_{2,t}\left(K,\sigma^2\right) &= \frac{1}{\sigma}\left[\log\left(S_t/K\right) + r_{f,t}\right] - \frac{\sigma}{2}
\end{aligned}$$

where Φ stands for the cumulative distribution function of the standard normal. Generally speaking, when a call option price $C_t(K)$ is observed at time t, for an European call maturing at time $(t+1)$ and written on the underlying asset with time-t price S_t, and with strike price K, the Black-Scholes implied volatility for this call is defined as unique solution $\sigma_{imp,t}(K)$ of the equation:

$$C_t(K) = BS_{(t)}\left(K, S_t, \sigma_{imp,t}(K)\right)$$

Therefore, if observed option prices are conformable to the option pricing model (2.1), we will have for any possible value K of the strike price :

$$\sigma_{imp,t}(K) = \sigma\left[J(t)\right]$$

In this case, the volatility smile is flat: the implied volatility may be stochastically time-varying but does not depend on the moneyness. This is the reason why GARCH option pricing cannot produce a volatility smile for short maturity options. It takes the necessity to forecast the variables at stake on the option lifetime to explain the volatility smile. Therefore, in our setting, in order to accommodate a non-flat volatility smile, we need to consider that some variables that define the information set $J(t)$ are not observed at time t by the representative investor:

$$J(t) = I(t) \vee \tilde{I}(t)$$

where $I(t)$ (resp $\tilde{I}(t)$) stands for the information that is available (resp. latent) for investor at time t. Note that in popular option pricing models, $\tilde{I}(t)$ becomes eventually observed at the latest at the maturity date of the option. For instance, in the classical stochastic volatility option pricing models (see e.g. Heston (1993)), the volatility path is eventually observed by investors. However, investors may not be able, given the present value of the spot volatility, to make a perfect forecast of the integrated volatility path until the maturity of the option. With short maturities in mind (maturity date at time $t+1$), the volatility path on the time interval $[t, t+1]$ will be our leading example of the information set $\tilde{I}(t)$.

In order to figure out the resulting shape of the volatility smile, it is then worth defining a latent stock price $\tilde{S}_t$ that would be the actual price if the information set $J(t)$ would have been observed at time t. We would have

$$\begin{aligned}
\tilde{r}_{t+1} &= \log\left(S_{t+1}/\tilde{S}_t\right) - r_{f,t} \\
&\Longrightarrow E^*[\exp(\tilde{r}_{t+1})\,|J(t)] = 1
\end{aligned}$$

meaning that:

$$\begin{aligned}
\tilde{S}_t &= e^{-r_{f,t}} E^*[S_{t+1}\,|J(t)] \\
&\Longrightarrow \tilde{S}_t = S_t \exp\left(\mu\,[J(t)] + \frac{\sigma^2\,[J(t)]}{2}\right)
\end{aligned}$$

Note that:

$$\begin{aligned}
S_t &= e^{-r_{f,t}} E^*[S_{t+1}\,|I(t)] = E^*[\tilde{S}_t\,|I(t)] \\
&\Longrightarrow E[\exp\left(\mu\,[J(t)] + \frac{\sigma^2\,[J(t)]}{2}\right)|I(t)] = 1
\end{aligned} \tag{2.2}$$

We must also acknowledge that, with a general equilibrium perspective, the interest rate process $r_{f,t}$ itself should be impacted by the broadening of the available information set from $I(t)$ to $J(t)$. However, following the dominant tradition for option pricing on equity, we overlook the interest rate risk and do not match the change of stock price (from S_t to $\tilde{S}_t$) by a corresponding change of the short term interest rate (see Garcia, Luger and Renault (2003) for a more comprehensive approach). Given information $J(t)$, option prices a time t would be conformable to the Black and Scholes option pricing formula (2.1) but with the value $\tilde{S}_t$ of the underlying stock price. Therefore, by the law of iterated expectations, we see that the actual option price when only information $I(t)$ is available is

$$C_t(K) = E^*[BS_{(t)}\left(K, \tilde{S}_t, \sigma^2\left[J(t)\right]\right)|I(t)] \tag{2.3}$$

It is worth noting that the conditional expectation in (2.3) is computed with respect to two sources of randomness, namely the joint distribution of $\tilde{S}_t$ and $\sigma^2\left[J(t)\right]$ given $I(t)$, that is a function of the conditional distribution of $J(t)$ given $I(t)$. Since this distribution does not depend on past and current returns (our maintained assumption), the option price $C_t(K)$ is, like the BS price $BS_{(t)}\left(K, \tilde{S}_t, \sigma^2\left[J(t)\right]\right)$, a function of the pair (S_t, K) that is homogeneous of degree one. As a result, the associated BS implied volatility $\sigma_{imp,t}(K)$, defined by:

$$BS_{(t)}\left(K, S_t, \sigma_{imp,t}(K)\right) = E^*[BS_{(t)}\left(K, S_t, \sigma^2\left[J(t)\right]\right)|I(t)]$$

depends on (S_t, K) only through the moneyness (K/S_t), or equivalently through the net log-moneyness:

$$x_t(K) = \log\left(K/S_t\right) - r_{f,t}$$

Note that (see Garcia and Renault (1998)) this homogeneity property would not hold in the case of GARCH option pricing. In any case, the non-linearity of the Black-Scholes pricing formula will in general imply that $\sigma_{imp,t}(K)$ does depend on the strike price K (or on the moneyness $x_t(K)$), leading to a non-flat volatility smile. The following Proposition 2.1 is an immediate corollary of a general result proved in Renault (1997):

Proposition 2.1:

We have $\tilde{S}_t \equiv S_t$ (almost surely) if and only if $\mu\left[J(t)\right] + \left(\sigma^2\left[J(t)\right]/2\right)$ belongs to the information set $I(t)$ and in this case the volatility smile, depicting implied volatilities $\sigma_{imp,t}(K)$ as functions of the log-moneyness $\left[\log\left(K/S_t\right) - r_{f,t}\right]$:

$$BS_{(t)}\left(K, S_t, \sigma_{imp,t}(K)\right) = E^*[BS_{(t)}\left(K, S_t, \sigma^2\left[J(t)\right]\right)|I(t)]$$

is an even function, minimum at zero log-moneyness (at the money option).

It is worth realizing that the condition $\tilde{S}_t \equiv S_t$ is not only sufficient but also almost necessary for a symmetric volatility smile. To see that, let us imagine, without loss of generality that:

$$\mu\left[J(t)\right]+\left(\sigma^{2}\left[J(t)\right]/2\right)=\mu_{1}\left[I(t)\right]+\mu_{2}\left[I(t),\tilde{Z}_{t}\right]$$

for some random vector $\tilde{Z}_t$. By (2.2):

$$\exp\left(\mu_{1}\left[I(t)\right]\right)E^{*}[\exp\left(\mu_{2}\left[I(t),\tilde{Z}_{t}\right]\right)|I(t)]=1$$

and thus:

$$\frac{\tilde{S}_{t}}{S_{t}}=\frac{\exp\left\{\mu_{2}\left[I(t),\tilde{Z}_{t}\right]\right\}}{E^{*}[\exp\left(\mu_{2}\left[I(t),\tilde{Z}_{t}\right]\right)|I(t)]}$$

For instance, in case of a linear function:

$$\mu_{2}\left[I(t),\tilde{Z}_{t}\right]=\lambda'\tilde{Z}_{t}$$

we have (see appendix for a proof)

Proposition 2.2:

$$\frac{\partial C_{t}(K)}{\partial\lambda}\left(\lambda=0\right)=S_{t}Cov^{*}[\tilde{Z}_{t},\Phi\left(d_{1,t}\left(K,\sigma^{2}\left[J(t)\right]\right)\right)|I(t)]$$

It is worth noting that the derivative computed in proposion 2.2 overlooks that $\sigma^2\left[J(t)\right]$ may also depend on the parameter λ. We will show in Section 3.4 below in a more specific model that $\sigma^2\left[J(t)\right]$ will actually depend on λ but that its derivative at $\lambda = 0$ is indeed zero. It is then pretty clear that the presence of a latent state variable $\lambda'\tilde{Z}_t$ will in general make the smile asymmetric around the moneyness, since locally around the symmetric smile, that we get for $\lambda = 0$, the distortion due to the term $\lambda'\tilde{Z}_t$ should be different for in or out the money options. To see that, we may imagine (almost without loss generality) that the first component $\tilde{Z}_{1,t}$ of the latent variable $\tilde{Z}_t$ is an increasing function the volatility $\sigma\left[J(t)\right]$ and prove the following:

Proposition 2.3:

If:

$$\tilde{Z}_{1,t}=h\left[\sigma^{2}\left[J(t)\right],I(t)\right]$$

with an increasing deterministic function $\sigma^2 \longmapsto h\left[\sigma^2, I(t)\right]$, then, with the log-moneyness $x_t(K) = \log\left(K/S_t\right) - r_{f,t}$:

$$x_{t}(K)>0\Longrightarrow\frac{\partial C_{t}(K)}{\partial\lambda_{1}}\left(\lambda=0\right)>0$$

In other words, for all options that are out of the money, the option price is an increasing function of λ_1. Moreover, the proof of Proposition 2.3 shows that the more out of the money the option is, the steeper is the slope of the option price as a function of λ_1. Thus, we can state that a non-zero λ_1 will distort the benchmark U-shape symmetric volatility smile that we get for $\lambda = 0$. With obvious notations, assuming to simplify that $\tilde{Z}_t$ (and λ) is unidimensional:

$$\sigma_{imp,t}(K) = BS^{-1}\left[C_t(K)_{(\lambda=0)} + \lambda \frac{\partial C_t(K)}{\partial \lambda}(\lambda = 0) + o(\lambda)\right]$$

Increasing the moneyness (in the direction of out-the-money options) will amplify the impact of a non-zero λ, producing a skewed volatility smile. It is worth noting that this skewed smile is an illustration of a general phenomenon; as explained by Renault and Touzi (1996) and Renault (1997), a non-zero leverage effect is the cause of a skewed smile, while a symmetric smile is obtained in the case of no-leverage. The latent variable $\tilde{Z}_t$ that enters the first two conditional moments and is not observed yet at time t accommodates a discrete time version of the instantaneous correlation between return and volatility that characterizes leverage effect. In this respect, it may be said that the occurence of leverage effect is identified by the occurence of a skewed volatility smile. With a negative correlation (as well documented for leverage effect), and thus a negative factor loading λ, one may expect that the volatility smile will be less steeply increasing (or even eventually decreasing) on the out-the-money side. This is in accordance with some well-documented stylized facts (see also our empirical section below).

3 Statistical identification through an exponential affine pricing model

3.1 Risk-Neutral return distribution

While the volatility smile may allow to identify the presence of leverage effect, its actual shape is a cross-sectional phenomenon that is fully determined by risk neutral parameters. For the purpose of statistical inference taking advantage of time series data, we need to bridge the gap between historical and risk neutral distributions through the specification of a pricing kernel. This pricing kernel should be stochastically time varying through the relevant state variables, namely not only the underlying asset returns but also, in the context of Proposition 2.3, the latent state variable $\sigma_{t+1}^2 = \lambda' \tilde{Z}_t$. The rationale for this notation is twofold:

First, since following the leverage effect interpretation of the result of Proposition 2.3, the latent state variable should be tightly related to volatility, we use the notation σ for a volatility factor whose exact relationship with return volatility will be made explicit only in Section 4.

Second, since the role of this state variable is to be eventually known by investors at the maturity date $t+1$ of the short term option, but unknown at the time t of option pricing, it is natural to use the time notation $t+1$. To make the information setting even more explicit, we will denote by $I^\sigma(t)$ the specific version of the information set $J(t)$:

$$I^\sigma(t) = I(t) \vee \left\{\sigma^2_{t+1}\right\} \subset I(t+1)$$

Note that we define the extended information set through only one latent factor σ^2_{t+1}. Of course, we could accommodate more generally a multi-factor volatility model where different factors, components of $\tilde{Z}_t$ and such that $\sigma^2_{t+1} = \lambda' \tilde{Z}_t$, may capture for instance slowly and fast varying components of volatility. The advantages of such a factorial hidden Markov volatility model have been recently enhanced by Augustyniak et al. (2018), albeit in a multiplicative (instead of additive) setup. It may accommodate a more flexible description of the term structure of volatility smiles, for options of different maturities. This is beyond the scope of this paper.

Therefore, we are conformable to the model described in Section 2 by assuming that the risk-neutral conditional moment generating function of asset returns is defined as follows:

$$E^*[\exp\left(-v r_{t+1}\right)|I^\sigma(t)] = \exp\left[-\alpha^*(v)\sigma^2_{t+1} - \beta^*(v)\sigma^2_t - \gamma^*(v)\right] \tag{3.1}$$

for some functions $\alpha^*(.), \beta^*(.), \gamma^*(.)$ such that:

$$\alpha^*(0) = \beta^*(0) = \gamma^*(0) = 0$$

Note that we allow in this section more generality than in Section 2 since we allow a general exponential affine model (Darolles et al. (2006)), without maintaining the assumption of conditional normality. Conditional normality for all possible value of σ^2_t and σ^2_{t+1} is actually equivalent to assuming that the functions $\alpha^*(.), \beta^*(.), \gamma^*(.)$ are quadratic with negative coefficients of v^2 in all three polynomials. The important property maintained in both sections is that returns are serially independent given state variables: the Laplace transform (3.1) does not depend on current and past returns.

3.2 An historical non-linear state-space model

The goal for statistical identification from return data is to write down a dynamic model that specifies not only the historical conditional distribution of return r_{t+1} given past and (latent) current values of the state variables process $\sigma_\tau^2, \tau \leq t+1$ (mean equation) but also a transition equation that describes the dynamics of the state variable process σ_t^2. We will do it as follows.

(i) Measurement equation:

$$E[\exp\left(-vr_{t+1}\right)|I^{\sigma}(t)] = \exp\left[-\alpha(v)\sigma_{t+1}^2 - \beta(v)\sigma_t^2 - \gamma(v)\right] \tag{3.2}$$

for some functions $\alpha(.), \beta(.), \gamma(.)$ such that:

$$\alpha(0) = \beta(0) = \gamma(0) = 0$$

(ii) Transition equation:

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

for some functions $a(.)$ and $b(.)$ such that:

$$a(0) = b(0) = 0$$

Several remarks are in order.

First, we have chosen to specify a measurement equation whose shape is identical to the one proposed before for the risk neutral distribution. While Section 2 has shown that this shape was convenient for the study of the volatility smile based on the risk neutral distribution, it is not really necessary to impose the same shape for the historical distribution. However there is some appeal to proceed like that and we will show below that an exponential affine pricing kernel does deliver the requested shape-preserving property when switching from the risk-neutral distribution to the historical one. It is worth mentioning that while Feunou and Tedongap (2012) study a discrete time model of leverage effect germane to ours, their affine specification with leverage cannot work simultaneously for the historical and the risk neutral distribution. Due to their use of the Inverse Gaussian distribution, a general exponentially affine pricing kernel is not structure preserving in their context. They can use their model either for risk neutral distribution or for the historical one, but not both.

Second, the transition equation maintains that the state variable process σ_t^2 is Markov of order one. It will of course be the case if we see it as the discrete time sampling of an underlying diffusion process. However, by using a multifactor model with a Markov of order one multivariate process $\tilde{Z}_t$ of volatility factors, we would have more generally an ARMA structure for $\sigma_{t+1}^2 = \lambda'\tilde{Z}_t$. Another possibility would be to allow more generally for H lags $\sigma_{t+1-h}^2, h = 1, ..., H$ in both risk neutral and historical conditional moment generating functions, while the same number of lags could be introduced in the exponential affine pricing kernel defined below.

As already announced, we will get the shape-preserving property by assuming that the Stochastic Discount Factor (SDF) is exponentially affine and does not depend on current and past returns. We specify it as follows:

$$M_{t+1}(\varsigma) = \exp(-r_{f,t} + m_0 + m_1\sigma_t^2)\exp\left(-\varsigma_1\sigma_{t+1}^2 - \varsigma_2 r_{t+1}\right) \tag{3.3}$$

This SDF introduces two parameters for the compensation of risk: ς_1 for compensation of the risk of the volatility factor, ς_2 for compensation of the underlying asset risk. As shown below, the two parameters m_0 and m_1 are imposed by the need to match the time variations of the short term interest rate $r_{f,t}$. The key issue is to show that such a SDF is able to bridge the gap between the historical and the risk neutral distributions defined above, meaning that we have the consistency identity:

$$\exp(-r_{f,t})\, E^*[H(\sigma_{t+1}^2, r_{t+1}, I(t))\,|I(t)] = E[M_{t+1}(\varsigma)\, H(\sigma_{t+1}^2, r_{t+1}, I(t))\,|I(t)] \tag{3.4}$$

for any function $H(\cdot,\cdot,\cdot)$. This is the purpose of Proposition 3.1 below.

Proposition 3.1:

A SDF specified by (3.3), makes consistent (in the sense of (3.4)) the risk neutral conditional distribution of returns defined by (3.1) and the historical one (3.2) when the two following conditions are fulfilled:

First, the functions $\alpha(.), \beta(.), \gamma(.)$ and $\alpha^*(.), \beta^*(.), \gamma^*(.)$ are related by the following identities

$$\begin{aligned}
\alpha^*(v) &= \alpha(\varsigma_2 + v) - \alpha(\varsigma_2) \\
\beta^*(v) &= \beta(\varsigma_2 + v) - \beta(\varsigma_2) \\
\gamma^*(v) &= \gamma(\varsigma_2 + v) - \gamma(\varsigma_2)
\end{aligned} \tag{3.5}$$

Second, the state variable process is Markov of order one for the risk neutral distribution with a moment generating function:

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

with functions $a^*(.)$ and $b^*(.)$ defined as follows:

$$\begin{aligned}
a^*(u) &= a(u + \varsigma_1 + \alpha(\varsigma_2)) - a(\varsigma_1 + \alpha(\varsigma_2)) \\
b^*(u) &= b(u + \varsigma_1 + \alpha(\varsigma_2)) - b(\varsigma_1 + \alpha(\varsigma_2))
\end{aligned} \tag{3.6}$$

In particular, the risk neutral conditional distribution of asset returns r_{t+1} is Gaussian if and only if the historical conditional distribution r_{t+1} is Gaussian.

Not surprisingly, Proposition 3.1 shows that it is a non-zero price ς_2 of the return risk that introduces a difference between the historical and the risk neutral return distribution. However, it is worth reminding that these distributions are conditional not only to the (possibly) observed state variable σ_t^2 but also to the unobserved one σ_{t+1}^2. For the purpose of statistical identification, we also need to chararacterize the conditional historical distribution of the asset return r_{t+1} given the available information $I(t)$ at time t:

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

This explains why the risk neutral distribution of volatility differs from the historical one not only through a non-zero price ς_1 for the volatility risk, but also through the function $\alpha\left(\varsigma_2\right)$ of the price ς_2 of the risk on return.

3.3 Identification of prices of risk

Risk compensation for the underlying asset return is encapsulated in the following identity:

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

that is equivalent to:

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

Note (see the proof of Proposition 3.1) that the joint conditional distribution of $\left(\sigma_{t+1}^2, r_{t+1}\right)$ given $I(t)$ can be characterized by the moment generating function:

$$\begin{aligned} E[\exp(-u\sigma_{t+1}^2-vr_{t+1})\,|I(t)] &= \exp\left(-l(u,v)\sigma_t^2-g(u,v)\right) \\ l\left(u,v\right) &= a\left(u+\alpha(v)\right)+\beta(v) \\ g(u,v) &= b\left(u+\alpha(v)\right)+\gamma(v) \end{aligned}$$

In particular, the proof of Proposition 3.1. shows that:

$$m_0\left(\varsigma\right) = g\left(\varsigma\right), m_1\left(\varsigma\right) = l\left(\varsigma\right)$$

Hence, the risk compensation for the underlying asset return can be characterized by the following two equations:

$$\begin{aligned} g(\varsigma) &= g(\varsigma_1, \varsigma_2 - 1) \\ l(\varsigma) &= l(\varsigma_1, \varsigma_2 - 1) \end{aligned}$$

meaning that:

$$\begin{aligned} a\left[\varsigma_1 + \alpha(\varsigma_2 - 1)\right] + \beta(\varsigma_2 - 1) &= a\left[\varsigma_1 + \alpha(\varsigma_2)\right] + \beta(\varsigma_2) \\ b\left[\varsigma_1 + \alpha(\varsigma_2 - 1)\right] + \gamma(\varsigma_2 - 1) &= b\left[\varsigma_1 + \alpha(\varsigma_2)\right] + \gamma(\varsigma_2). \end{aligned} \tag{3.7}$$

It is common to say that the price ς_1 of volatility risk cannot by identified from time series data on the underlying asset return only but that it takes some option price data. It is actually obvious that the system (3.7) of two equations with two unknown ς_1 and ς_2 may not be able to identify them. Clearly, if one can find a value of the parameter ς_2 solution of the three equations:

$$\begin{aligned} \alpha(\varsigma_2) &= \alpha(\varsigma_2 - 1) \\ \beta(\varsigma_2) &= \beta(\varsigma_2 - 1) \\ \gamma(\varsigma_2) &= \gamma(\varsigma_2 - 1) \end{aligned}$$

then the two equations (3.7) are automatically fulfilled independently of the value of ς_1. However, the main result of this subsection is to show that this identification issue cannot kick in when leverage effect in the sense of Section 2 is present. We can actually show that:

Proposition 3.2:

If the conditional distribution of r_{t+1} given $I^{\sigma}(t)$ is Gaussian, then:

$$E^*[r_{t+1}\,|I^{\sigma}(t)] + \frac{1}{2}Var^*[r_{t+1}\,|I^{\sigma}(t)] - [\alpha(\varsigma_2) - \alpha(\varsigma_2 - 1)]\,\sigma^2_{t+1}$$

belongs to the information set $I(t)$.

In other words, what we have dubbed leverage effect in Section 2, namely the fact that $E^*[r_{t+1}\,|I^{\sigma}(t)] + \frac{1}{2}Var^*[r_{t+1}\,|I^{\sigma}(t)]$ depends on the latent variable σ^2_{t+1} and is responsible for producing a skewed volatility smile cannot happen if:

$$\alpha(\varsigma_2) = \alpha(\varsigma_2 - 1)$$

When observing a skewed volatility smile, we can be sure that there is some leverage effect and that due to this fact, the price ς_1 of volatility risk can be identified from underlying asset return data (without need of option price data) by solving in $(\varsigma_1, \varsigma_2)$ the system of two equations (3.7). Of course this later statement assumes that when $\alpha(\varsigma_2) \neq \alpha(\varsigma_2 - 1)$, the two equations in (3.7) are not redundant. This should be obviously generically true and will be confirmed by the parameterization studied in the next sections. It is indeed not so surprising (see also Bandi and Reno (2016)) that instantaneous correlation between volatility and return may allow us to identify the price of volatility risk with only return data.

It is also worth noting that the result of Proposition 3.2 takes conditional normality only in order to ensure that the function α is quadratic. If one considers more generally a Taylor expansion of the function $\alpha(v)$ in a neighborhood of zero, it is always true that the occurences of σ_{t+1}^2 in the terms of degree 1 and 2 in v are cancelled out by the subtraction of $[\alpha(\varsigma_2) - \alpha(\varsigma_2 - 1)]\,\sigma_{t+1}^2$. The deep reason for that is that the computation of the first two risk neutral conditional moments from the conditional moment generating function only takes the first two derivatives at $v = 0$ of the functions $\alpha^*(v)$, $\beta^*(v)$ and $\gamma^*(v)$.

3.4 Identification of leverage effect

The lesson of the previous subsection is that leverage effect goes through the quantity:

$$[\alpha(\varsigma_2) - \alpha(\varsigma_2 - 1)]\,\sigma_{t+1}^2.$$

Since conditional normality of return r_{t+1} given "extended" information $I^\sigma(t)$ is a maintained assumption throughout the rest of the paper, we can rewrite this quantity:

$$\alpha'(0)\sigma_{t+1}^2 + (2\varsigma_2 - 1)\frac{\alpha''(0)}{2}\sigma_{t+1}^2.$$

This shows that leverage effect goes through two key parameters, namely $\alpha'(0)$ and $\alpha''(0)$. We can make that more explicit by looking at the first two conditional moments deduced from the moment generating function.

First, we have:

$$E[r_{t+1}\,|I^\sigma(t)] = \alpha'(0)\sigma_{t+1}^2 + \beta'(0)\sigma_t^2 + \gamma'(0)$$

This suggests the introduction of a first leverage effect parameter ψ defined by:

$$\begin{aligned} \psi \quad &= \quad \alpha'(0) \\ \iff \quad & E[r_{t+1}\,|I^\sigma(t)] - E[r_{t+1}\,|I(t)] = \psi\left\{\sigma_{t+1}^2 - E[\sigma_{t+1}^2\,|I(t)]\right\} \end{aligned}$$

The parameter ψ encapsulates instantaneous causality between the volatility factor σ_{t+1}^2 and the return process r_{t+1}. One unit of innovation in σ_{t+1}^2 leads to update the return forecast by ψ units. The concept of leverage effect as introduced by Black (1976) leads us to expect a negative value of the parameter ψ: bad news in the return process are more often than not associated to high volatility levels.

Second, we have also:

$$Var[r_{t+1}\,|I^{\sigma}(t)] = -\alpha''(0)\sigma_{t+1}^2 - \beta''(0)\sigma_t^2 - \gamma''(0)$$

This formula suggests to introduce another volatility factor as:

$$\tilde{\sigma}_{t+1}^2 = \sigma_{t+1}^2 + \frac{\beta''(0)}{\alpha''(0)}\sigma_t^2 + \frac{\gamma''(0)}{\alpha''(0)}$$

so that we get the two following occurences of leverage effect:

$$\begin{aligned} E[r_{t+1}\,|I^{\sigma}(t)] - E[r_{t+1}\,|I(t)] &= \psi\left\{\tilde{\sigma}_{t+1}^2 - E[\tilde{\sigma}_{t+1}^2\,|I(t)]\right\} \\ Var[r_{t+1}\,|I^{\sigma}(t)] &= -\alpha''(0)\tilde{\sigma}_{t+1}^2 \end{aligned}$$

When interpreting $\tilde{\sigma}_{t+1}^2$ as a volatility factor, we expect a high level of leverage effect to manifest itself in two ways:

First, a parameter ψ large in absolute value means a high level of instantaneous causality between return and volatility,

Second, a positive parameter $[-\alpha''(0)]$ significantly smaller than one means that conditioning by the volatility path reduces the return volatility. For this reason, we will adopt the following parameterization:

$$-\alpha''(0) = 1 - \phi^2$$

and keep in mind that large absolute values of both parameters ψ and ϕ mean high leverage effect. This duality of parameters for capturing the same effect will lead us to relate them below. We first summarize the above discussion by the decomposition of return variance:

$$Var[r_{t+1}\,|I(t)] = \psi^2 Var[\tilde{\sigma}_{t+1}^2\,|I(t)] + \left[1-\phi^2\right] E[\tilde{\sigma}_{t+1}^2\,|I(t)] \tag{3.8}$$

This decomposition allows us to characterize as follows the conditional linear correlation between return and volatility, which is commonly seen as an empirical assessment of leverage effect:

Proposition 3.3:

$$\begin{aligned} Corr[r_{t+1}, \tilde{\sigma}_{t+1}^2\,|I(t)] &= Corr[r_{t+1}, \sigma_{t+1}^2\,|I(t)] \\ &= \psi\left\{\psi^2 + \left[1-\phi^2\right]\frac{E[\tilde{\sigma}_{t+1}^2\,|I(t)]}{Var[\tilde{\sigma}_{t+1}^2\,|I(t)]}\right\}^{-1/2} \end{aligned}$$

As a straightforward corollary we get:

Corollary 3.4:

(i) $Corr[r_{t+1}, \sigma_{t+1}^2 \,|I(t)]$ is a time invariant deterministic number if and only if there is a positive real number k such that:

$$\frac{E[\tilde{\sigma}_{t+1}^2 \,|I(t)]}{Var[\tilde{\sigma}_{t+1}^2 \,|I(t)]} \equiv k^2$$

(ii) In this case:

$$Corr[r_{t+1}, \sigma_{t+1}^2 \,|I(t)] = \phi \Longleftrightarrow \psi = k\phi$$

Several comments are in order.

First, it has been common in the option pricing literature, at least since Heston (1993), to accommodate leverage effect through a time invariant (negative) conditional correlation coefficient between returns and volatility. It will then be our duty in the next sections to design a volatility factor that makes this restriction consistent with empirical evidence.

Second, it is important to note that this restriction is consistent, up to a constraint between parameters, with the transition equation for the state variable σ_t^2 introduced in subsection 3.2. According to this transition equation:

$$\begin{aligned} E[\sigma_{t+1}^2 \,|I(t)] &= \omega + \rho\sigma_t^2 \\ \rho &= a'(0) \in (0,1),\ \omega = b'(0) \\ Var[\sigma_{t+1}^2 \,|I(t)] &= \bar{\omega} + \bar{\rho}\sigma_t^2 \\ \bar{\rho} &= -a''(0),\ \bar{\omega} = -b''(0). \end{aligned}$$

Then:

$$\begin{aligned} E[\tilde{\sigma}_{t+1}^2 \,|I(t)] &= \omega + \rho\sigma_t^2 - \rho e \sigma_t^2 - f\omega \\ e &= \frac{\beta''(0)}{\rho\,(1-\phi^2)},\ f = \frac{\gamma''(0)}{\omega\,(1-\phi^2)}. \end{aligned}$$

Hence, condition (i) of corollary 3.4 is equivalent to the following constraint:

$$1 - f = (1-e)\,\frac{\rho}{\bar{\rho}}\frac{\bar{\omega}}{\omega}$$

and then:

$$k^2 = (1-e)\,\frac{\rho}{\bar{\rho}}.$$

Third, as already announced, it takes a restriction relating coefficients ψ and ϕ in order to capture constant leverage effect through one single parameter. The restriction $\psi = k\phi$ may sound natural since we deduce from the variance decomposition (3.8) that:

$$\psi = k\phi \Longrightarrow Var[r_{t+1} \,|I(t)] = E[\tilde{\sigma}^2_{t+1} \,|I(t)]$$

This formula is reminiscent of continuous time stochastic volatility models in which the volatility factor would be the quadratic variation of the price process over the time interval $[t, t+1]$. However, this formula must be adjusted in case of a risk premium proportional to variance that makes the drift also random and impacting the conditional variance of returns. This is the reason why we will instead impose the following identification constraint of leverage effect.

Leverage effect identification constraint:

Under the maintained assumption

$$k > 0, \frac{E[\tilde{\sigma}^2_{t+1} \,|I(t)]}{Var[\tilde{\sigma}^2_{t+1} \,|I(t)]} \equiv k^2$$

we assume in addition:

$$\psi = k\phi + \left(1 - \phi^2\right)\left(\varsigma_2 - \frac{1}{2}\right) \tag{3.9}$$

Note that this constraint introduces not only the risk premium effect (that is the price of risk ς_2) but also a Jensen effect (ς_2 replaced by $\varsigma_2 - \frac{1}{2}$). The correction term is well suited in order to identify ϕ as the driving parameter for the leverage effect in option pricing since then:

$$[\alpha(\varsigma_2) - \alpha(\varsigma_2 - 1)]\,\sigma^2_{t+1} = k\phi\sigma^2_{t+1}$$

Of course, beyond the above identification constraint, statistical identification of ϕ will take a device for connecting the volatility factor σ^2_{t+1} with observed returns. This device will be described in Section 4 below.

Since with the notations of Section 2, we have:

$$\begin{aligned} Var[r_{t+1} \,|I^{\sigma}(t)] &= \left(1 - \phi^2\right)\tilde{\sigma}^2_{t+1} \\ \frac{\tilde{S}_t}{S_t} &= \frac{\exp\left(k\phi\sigma^2_{t+1}\right)}{E[\exp\left(k\phi\sigma^2_{t+1}\right) |I(t)]} \end{aligned}$$

we can then rewrite the option pricing formula (2.3) as:

$$\begin{aligned} C_t(K) &= E^*[BS_{(t)}\left(K, \tilde{S}_t(\phi), \left(1 - \phi^2\right)\tilde{\sigma}^2_{t+1}\right) |I(t)] \\ \tilde{S}_t(\phi) &= S_t \frac{\exp\left(k\phi\sigma^2_{t+1}\right)}{E[\exp\left(k\phi\sigma^2_{t+1}\right) |I(t)]} \end{aligned} \tag{3.10}$$

Formula (3.10) revisits in discrete time a formula first shown by Romano and Touzi (1997) (see also **?**). The leverage effect parameter ϕ plays a double role in the option pricing formula.

On the one hand, the underlying asset price is distorted (from S_t to $\tilde{S}_t(\phi)$) by the factor $\exp\left(k\phi\sigma_{t+1}^2\right)$ divided by its conditional mean.

On the other hand, only the share $\left(1-\phi^2\right)\tilde{\sigma}_{t+1}^2$ of volatility $\tilde{\sigma}_{t+1}^2$ matters for option pricing. The rationale is that conditioning by the volatility path reduces the variance of return by the factor ϕ^2.

We can then state in this specific setting a formula similar to Proposition 2.2:

$$\begin{aligned}\frac{\partial C_t(K)}{\partial \phi}(\phi=0) &= kS_t Cov^*[\sigma_{t+1}^2, \Phi\left(d_{1,t}\left(K, \tilde{\sigma}_{t+1}^2\right)\right)|I(t)] \\ &= kS_t Cov^*[\tilde{\sigma}_{t+1}^2, \Phi\left(d_{1,t}\left(K, \tilde{\sigma}_{t+1}^2\right)\right)|I(t)]\end{aligned}$$

It is worth noting that since the leverage effect parameter ϕ has two occurences, the derivative with respect to ϕ will entail two terms. However, since one of the occurences goes through ϕ^2, it gives a zero derivative at $\phi = 0$. It confirms the assumption previously maintained for the proof of Proposition 2.2. The computation of the derivative of the option price and the argument for its positivity for out of the money options is then similar to Proposition 2.2.:

$$x_t(K) > 0 \Longrightarrow \frac{\partial C_t(K)}{\partial \phi}(\phi=0) > 0$$

Again, we can claim that with a negative leverage effect coefficient ϕ, one may expect that the volatility smile will be less steeply increasing on the out-the-money side.

4 A discrete time version of Heston option pricing model

4.1 Identification of the volatility factor

We develop in this section a stochastic volatility (SV) extension of the HEAVY-GARCH model previously proposed by Shephard and Sheppard (2010). While the sequence of returns $t = 1, 2, ..., T$, can be seen as daily, we want to take also advantage of observations of daily realized variances $RV_t, t = 1, 2, ..., T$. Strictly speaking, the availability of these observations means that the information sets contain intraday return data through observation of say n underlying asset prices $S_{t+i/n}, i = 1, 2, ..., n$ per day. For convenience, we will not change the notations for the information sets $I(t)$ and $I^\sigma(t)$, assuming that the availability of additional intraday data does not modify the conditional distributions we have described in the previous sections. Inspired by the GARCH(1,1) model, Shephard and Sheppard (2010) have proposed the following model:

$$\begin{aligned} \mu_t &= E[RV_{t+1} \,| I(t)] \\ \mu_t &= \omega_R + \alpha_R RV_t + \beta_R \mu_{t-1} \end{aligned}$$

Similarly to the analysis led in Meddahi and Renault (2004), we note that this GARCH-type model is a particular case of a SV-type model defined by the AR(1) dynamics of $\mu_t = E[RV_{t+1} \,| I(t)]$:

$$\mu_t = \omega_R + (\alpha_R + \beta_R)\mu_{t-1} + \nu_t, \;\; E[\nu_t \,| I(t-1)] = 0$$

In all cases the process RV_t is ARMA(1,1):

$$\begin{aligned} &\eta_{t+1} = RV_{t+1} - \mu_t \\ &\quad\Longrightarrow (RV_{t+1} - \eta_{t+1}) = \omega_R + (\alpha_R + \beta_R)\,(RV_t - \eta_t) + \nu_t \\ &\quad\Longrightarrow RV_{t+1} = \omega_R + (\alpha_R + \beta_R) RV_t - (\alpha_R + \beta_R)\eta_t + \eta_{t+1} + \nu_t \end{aligned}$$

$$E[-(\alpha_R + \beta_R)\eta_t + \eta_{t+1} + \nu_t \,| I(t-1)] = 0$$

However, in the general case the innovation process of this ARMA(1,1) is spanned by two not perfectly correlated processes η and ν, while in the GARCH-type model ν_t and thus also μ_t are deterministic functions of past and present values of $RV_\tau, \tau \leq t$, or equivalently of $\eta_\tau, \tau \leq t$. This dimension two means that μ_t is a genuinely latent AR(1) process that may be well suited for the identification of the space spanned by our stochastic volatility factor σ_t^2. We know from Section 2 that this genuine latency of σ_t^2 is needed for our purpose. Then, we end up with two latent AR(1) processes μ_t and σ_t^2 for which we may expect that they are related by an exact affine relationship. This will be implied by our model specification.

More precisely, we follow the logic of the introduction of the ARMA(1,1) process $\tilde{\sigma}_t^2$ in the former section to assume that a similar relationship ties the ARMA(1,1) process RV_t with the state variable process sigma:

$$RV_{t+1} = \sigma_{t+1}^2 - B\sigma_t^2 - D \tag{4.1}$$

Several remarks are in order.

First, we choose a unit coefficient for σ_{t+1}^2 in formula (4.1). This can be assumed without loss of generality since the latent volatility factor is obviously defined up to an arbitrary scaling factor, and a unit coefficient is consistent with the previous definition of $\tilde{\sigma}_t^2$. Note that since the latent factor σ_t^2 is by definition conformable to the AR(1) dynamics:

$$E[\sigma_t^2 \,| I(t)] = \omega + \rho\sigma_{t-1}^2$$

we get as already announced an affine relationship between μ_t and σ_t^2:

$$\mu_t = (\rho - B)\,\sigma_t^2 + \omega - D$$

Second, we allow the ARMA(1,1) process RV_t to differ from the ARMA process $\tilde{\sigma}_t^2$ introduced in Section 3 because we know that the optimal forecast of $\tilde{\sigma}_{t+1}^2$ does not exactly coincide with the conditional variance of return r_{t+1} while we will impose the restriction:

$$\mu_t = E[RV_{t+1}\,|I(t)] = Var[r_{t+1}\,|I(t)] \tag{4.2}$$

Note that this restriction has been extensively discussed in the HEAVY-GARCH literature. Shephard and Sheppard (2010) note that the conditional variance $Var[r_{t+1}\,|I(t)]$ is a"close-to-close" measure while $\mu_t = E[RV_{t+1}\,|I(t)]$ can be interpreted as an "open-to-close" conditional variance of returns. For this reason, Brownlees and Gallo (2010) have proposed the additional degree of freedom that μ_t and $Var[r_{t+1}\,|I(t)]$ would be only related by an exact affine relationship. However, they did not find compelling empirical evidence against the identity (4.2) that will be a maintained assumption throughout this paper.

Third, the maintained assumption (4.2) amounts to a couple of constraints between parameters since:

$$\begin{aligned} Var[r_{t+1}\,|I(t)] &= \psi^2 Var[\tilde{\sigma}_{t+1}^2\,|I(t)] + \left[1-\phi^2\right] E[\tilde{\sigma}_{t+1}^2\,|I(t)] \\ &= \psi^2\left[\bar{\omega} + \bar{\rho}\sigma_t^2\right] + \left[1-\phi^2\right]\left[\rho\sigma_t^2\,(1-e) + \omega\,(1-f)\right]. \end{aligned}$$

(4.2) is fulfilled if and only if:

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

By writing:

$$\psi^2 = k^2\phi^2 + A = (1-e)\,\frac{\rho}{\bar{\rho}}\phi^2 + A$$

and we get the two constraints:

$$\begin{aligned} \rho - B &= \rho\,(1-e) + A\bar{\rho} \\ \omega - D &= \bar{\omega}\left[A + (1-e)\,\frac{\rho}{\bar{\rho}}\right] \end{aligned}$$

To summarize, our model specification is:

$$
\begin{aligned}
RV_{t+1} &= \sigma^2_{t+1} - B\sigma^2_t - D \\
B &= \rho e - A\bar{\rho} \\
D &= \omega - \bar{\omega}\left[A + (1-e)\frac{\rho}{\bar{\rho}}\right] \\
A &= \psi^2 - k^2\phi^2 \\
&= \left[k\phi + \left(1-\phi^2\right)\left(\varsigma_2 - \frac{1}{2}\right)\right]^2 - k^2\phi^2 \\
k^2 &= (1-e)\frac{\rho}{\bar{\rho}}
\end{aligned}
$$

It is worth noting that:

$$
\begin{aligned}
RV_{t+1} &= \tilde{\sigma}^2_{t+1} + A\left[\bar{\rho}\sigma^2_t + \bar{\omega}\right] \\
&= \tilde{\sigma}^2_{t+1} + AVar[\sigma^2_{t+1}\,|I(t)]
\end{aligned} \tag{4.3}
$$

In particular:

$$
RV_{t+1} = \tilde{\sigma}^2_{t+1} \Longleftrightarrow A = 0 \Longleftrightarrow \psi^2 = k^2\phi^2 \Longleftrightarrow Var[r_{t+1}\,|I(t)] = E[\tilde{\sigma}^2_{t+1}\,|I(t)]
$$

Fourth, the relationship (4.3) between RV_{t+1} and $\tilde{\sigma}^2_{t+1}$ gives an easy way to check the empirical validity of the leverage effect identification assumption (condition (i) of corollary 3.4) since:

$$
\frac{E[\tilde{\sigma}^2_{t+1}\,|I(t)]}{Var[\tilde{\sigma}^2_{t+1}\,|I(t)]} = \frac{E[RV_{t+1}\,|I(t)] - AVar[\sigma^2_{t+1}\,|I(t)]}{Var[\sigma^2_{t+1}\,|I(t)]} = \frac{E[RV_{t+1}\,|I(t)]}{Var[RV_{t+1}\,|I(t)]} - A.
$$

Therefore, the condition (i) of corollary 3.4 is fulfilled if and only if $E[RV_{t+1}\,|I(t)]$ is proportional to $Var[RV_{t+1}\,|I(t)]$. In Appendix A, we propose an empirical assessment of this condition. In order to get a model-free assessment, we compute fitted values of the time series $E[RV_{t+1}\,|I(t)]$ and $Var[RV_{t+1}\,|I(t)]$ that are based on the estimation of an AR(1) model for the process RV_t with ARCH(1) innovations:

$$
\begin{aligned}
RV_{t+1} &= \omega_R + \alpha_R RV_t + \nu_t \\
\nu_{t+1} &= h_t^{1/2}u_{t+1}, E[u_{t+1}\,|I(t)] = 0, E[u^2_{t+1}\,|I(t)] = 1 \\
Var[RV_{t+1}\,|I(t)] &= h_t = \omega_h + \alpha_h\nu^2_t
\end{aligned}
$$

It is important to keep in mind that this specification for the dynamics of realized variance is not maintained throughout this paper. It is only used as a filter for computing fitted values of $E[RV_{t+1}\,|I(t)]$ and $Var[RV_{t+1}\,|I(t)]$. Figure 5 shows that over 16 years of daily data (realized variance of SNP500 from January 2000 to June 2016) it is a sensible approximation to see the time series $\{E[RV_{t+1}\,|I(t)]\,/Var[RV_{t+1}\,|I(t)]\}$ as a constant close to unity. The coefficient of variation (CV) of the ratio confirms this visual assessment: CV is only 0.21 and even drops to 0.15 when we eliminate the 5% most extreme observations.

4.2 Heston model as a continuous time limit

We want to check that the affine specification introduced above for the joint dynamics of $\left(r_{t+1}, \sigma_{t+1}^2\right)$ is a discrete time version of Heston (1993) option pricing model. As already well-known in the GARCH/SV literature, there is no such thing as a unique way to embed a discrete time model in a continuous model. However, our specification of joint affine dynamics of the process σ_t^2 for its first two conditional moments:

$$\begin{aligned} E[\sigma_{t+1}^2\,|I(t)] &= \omega + \rho\sigma_t^2 \\ Var[\sigma_{t+1}^2\,|I(t)] &= \bar{\omega} + \bar{\rho}\sigma_t^2 \end{aligned} \tag{4.4}$$

obviously amounts to a vector auto-regressive VAR(1) specification for the bivariate process $\left(\sigma_t^2, \sigma_t^4\right)$ for which temporal aggregation formulas are well-known. (see Meddahi and Renault (2004) for an extensive discussion of this approach). These temporal aggregation formulas give us an unambiguous guidance about how to address the continuous time limit issue. For this purpose we define a volatility factor:

$$\sigma_{t,H}^2(N) = \frac{1}{HN}\sum_{n=1}^{HN}\sigma_{t+\frac{n}{N}}^2$$

where N is the number of subintervals in a unit interval. Our normalization by the factor HN allows us to keep the interpretation of $\sigma_{t,H}^2(N)$ as a volatility factor on a given (the smallest possible) unit of time.

For the sake of getting the instantaneous analog of $\sigma_{t,H}^2(N)$, we will consider that the horizon H may go to zero, while always assuming $HN \geq 1$ and (for sake of notational simplicity) maintaining the assumption that HN is an integer.

The following lemma is then easy to guess (proof available upon request) and useful to get the continuous time limit of our model:

Lemma 4.1:

$$\lim_{H\to 0} E\left[\sigma^2_{t,H}(N)|I(t)\right] = \sigma^2_t$$
$$\lim_{H\to 0} \frac{1}{H} Var\left(\sigma^2_{t,H}(N)|I(t)\right) = \frac{1}{2}\lim_{H\to 0} \frac{1}{H} Var\left(\sigma^2_{t+H,H}(N)|I_t\right).$$

We can then prove the following proposition:

Proposition 4.2: (Continuous-time limit) If the equations in Lemma 4.1 hold, then we have for all integer N and $H \in [1/N, \infty)$,

$$\lim_{H\to 0} \frac{1}{H} E\left[\sigma^2_{t+H,H}(N) - \sigma^2_{t,H}(N)|\tilde{I}(t)\right] = -log(\rho)\left(\frac{\omega}{1-\rho} - \sigma^2_t\right),$$
$$\lim_{H\to 0} \frac{1}{H} Var\left(\sigma^2_{t,H}(N)|\tilde{I}(t)\right) = -\frac{\bar{\rho}}{\rho}\frac{log(\rho)}{1-\rho}\left(\sigma^2_t + \frac{\omega - 2\bar{\omega}(\rho/\bar{\rho})}{1+\rho}\right),$$

where a information set $\tilde{I}(t) = \left\{\sigma^2_{t-kH,H}(N), k \geq 1\right\}$.

In other words, the continuous time limit of this model is the affine model:

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

for some Wiener process W_t and

$$\kappa = -log(\rho) > 0,$$
$$\bar{\sigma}^2 = \frac{\omega}{1-\rho} = E[\sigma^2_t] > 0,$$
$$\eta = \frac{\kappa}{\omega}\frac{\bar{\rho}}{\rho}\bar{\sigma}^2,$$
$$\nu = \eta\frac{\omega - 2\bar{\omega}(\rho/\bar{\rho})}{1+\rho} \geq 0, \text{ if } \omega \geq 2\bar{\omega}(\rho/\bar{\rho}).$$

In particular, if $\omega = 2\bar{\omega}(\rho/\bar{\rho})$, we get for σ^2_t a square root process of Feller (1951), as used for interest rate by Cox et al. (1985) and for volatility by Heston (1993). The three parameters $(\kappa, \bar{\sigma}^2, \eta)$ are unconstrained (up to standard inequality constraints) one-to-one functions of the three initial parameters[1], $\rho, \bar{\omega}$, and $\bar{\rho}$. Therefore, as fas as the first two moments are concerned, any square root process can be seen as a continuous time limit of our volatility factor model. More generally, any affine process in continuous time (Duffie et al. (2000)) can be seen as the continuous time limit of our discrete time model thanks to the degree of freedom $\omega \neq 2\bar{\omega}(\rho/\bar{\rho})$.

The advantage of the discrete time specification is that, by contrast with Brownian diffusions, the specification of the first two conditional moments does not constrain us regarding higher order

[1] Note that $\omega = 2\bar{\omega}(\rho/\bar{\rho})$ is a function of $\rho, \bar{\omega}$, and $\bar{\rho}$ in this case.

moments. This may allow us in particular to accommodate stylized facts that take jumps both in returns and in volatility (see e.g. Bandi and Reno (2016)) to be captured by a continuous time model.

4.3 A fully parametric model

As already announced, the conditional distribution of r_{t+1} given $I^{\sigma}(t)$ is assumed to be Gaussian, and thus the functions $\alpha(\cdot), \beta(\cdot)$ and $\gamma(\cdot)$ that define this conditional distribution are quadratic functions, nil at $u = 0$, thus characterized by the values of $\alpha'(0), \alpha''(0), \beta'(0), \beta''(0), \gamma'(0)$ and $\gamma''(0)$. These six numbers are characterized by parameters that, as previously explained, must also be related to the parameters of volatility dynamics.

As far as volatility dynamics are concerned, we specify a discrete time model inspired by Heston (1993)'s continuous time model. Following Gourieroux and Jasiak (2006), we consider the simplest version where transition dynamics are driven by gamma distributions as in Heston (1993) model and its precursor Feller (1951)'s square root process. Extensions with mixture components to capture the tail effects of continuous time jumps are beyond the scope of this paper. We use more precisely the ARG(1) model defined by Gourieroux and Jasiak (2006) as follows:

(i) The conditional distribution of σ^2_{t+1} given some mixing variable U_t is gamma with a shape parameter $(\delta + U_t)$ and a scale parameter c,

(ii) The conditional distribution of U_t given σ^2_t is Poisson with parameter $\varrho\sigma^2_t/c$.

We easily verify that this parametric model is nested in the general affine model defined with functions $a(.)$ and $b(.)$ as in Section 3.2, with the specification:

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

and thus:

$$\begin{aligned} a'(0) &= \rho, \; a''(0) = -2\rho c \\ b'(0) &= \delta c, \; b''(0) = -\delta c^2. \end{aligned}$$

In other words:

$$\begin{aligned} E[\sigma^2_{t+1} \,| I(t)] &= \omega + \rho\sigma^2_t = \delta c + \rho\sigma^2_t \\ Var[\sigma^2_{t+1} \,| I(t)] &= \bar{\omega} + \bar{\rho}\sigma^2_t = \delta c^2 + 2\rho c\sigma^2_t. \end{aligned}$$

In summary, volatility dynamics is defined by a 3-dimensional vector:

$$\theta_\sigma = (\rho, \delta, c)'$$

We can then characterize the return parameters as follows:

(i) ϕ defined by:

$$1 - \phi^2 = -\alpha''(0), \ \phi < 0$$

(ii) e defined by:

$$e = \frac{\beta''(0)}{\rho\,(1-\phi^2)}.$$

Note that this allows us to also define:

$$k^2 = (1-e)\,\frac{\rho}{\bar{\rho}} = \frac{1-e}{2c}$$

and to get:

$$\gamma''(0) = f\omega\left(1-\phi^2\right) = f\delta c\left(1-\phi^2\right),$$

with:

$$f = 1 - (1-e)\,\frac{\rho}{\bar{\rho}}\frac{\bar{\omega}}{\omega} = 1 - \frac{1-e}{2}.$$

(iii) The risk premium parameter ς_2 is identified by:

$$\alpha'(0) = \psi = k\phi + \left(1-\phi^2\right)\left(\varsigma_2 - \frac{1}{2}\right).$$

This allows us to characterize the relationship between the observed process RV_t and the latent one σ_t^2:

$$\begin{aligned}
RV_{t+1} &= \sigma_{t+1}^2 - B\sigma_t^2 - D \\
B &= \rho e - 2A\rho c = \rho\,[e - 2Ac] \\
D &= \delta c - \delta c^2\left[A + \frac{1-e}{2c}\right] \\
&= \delta c\left[1 - Ac - \frac{1-e}{2}\right] \\
A &= \left[k\phi + \left(1-\phi^2\right)\left(\varsigma_2 - \frac{1}{2}\right)\right]^2 - k^2\phi^2.
\end{aligned}$$

(iv) Given $\beta''(0)$ and $\gamma''(0)$ defined by (ii) above, the complete parameterization of the quadratic functions $\beta(\cdot)$ and $\gamma(\cdot)$ is tantamount to the parameterization of the two numbers $\beta'(0)$ and $\gamma'(0)$. These two quantities are parameterized by the volatility risk premium parameter ς_1 through the identification conditions (3.7):

$$\begin{aligned}
\beta(\varsigma_2) - \beta(\varsigma_2 - 1) &= \beta'(0) + \beta''(0)\left(\varsigma_2 - \frac{1}{2}\right) \\
&= a\left[\varsigma_1 + \alpha(\varsigma_2 - 1)\right] - a\left[\varsigma_1 + \alpha(\varsigma_2)\right] \\
\gamma(\varsigma_2) - \gamma(\varsigma_2 - 1) &= \gamma'(0) + \gamma''(0)\left(\varsigma_2 - \frac{1}{2}\right) \\
&= b\left[\varsigma_1 + \alpha(\varsigma_2 - 1)\right] - b\left[\varsigma_1 + \alpha(\varsigma_2)\right].
\end{aligned}$$

Even though the volatility risk premium parameter ς_1 is theoretically overidentified by these two equations, our experience is that it is actually hard to estimate with data only on the underlying asset return and its realized volatility. Option data are definitely much more informative about this volatility risk parameter. However, we will call return parameters the parameters θ_r identified by return data, given the three volatility parameters θ_σ :

$$\begin{aligned}
\theta_r &= (\phi, e, \varsigma_1, \varsigma_2)' \\
\theta_\sigma &= (\rho, \delta, c)'.
\end{aligned}$$

5 GMM estimation of an exponential affine model

5.1 General framework

As explained in Section 4, our discrete time version of Heston's model is, as far as return and volatility data are concerned, a fully parametric model characterized by two exponential affine conditional distributions:

(i) The conditional distribution of σ_{t+1}^2 given σ_t^2, characterized by two parameterized functions $a(.)$ and $b(.)$, indexed by unknown parameters $\theta_\sigma = (\rho, \delta, c)'$.

(ii) The conditional distribution of r_{t+1} given σ_{t+1}^2 and σ_t^2, characterized by three parameterized functions $\alpha(.), \beta(.)$ and $\gamma(.)$ indexed by unknown parameters $\theta_r = (\phi, e, \varsigma_1, \varsigma_2)'$. However, these functions are defined only for a given value of $\theta_\sigma = (\rho, \delta, c)'$.

Even though maximum likelihood, whenever possible, would deliver efficient estimation, it is of course convenient to use the conditional moment restrictions directly provided by the exponential affine conditional moment generating function for a GMM strategy. Since the parameterized functions $\alpha(.), \beta(.)$ and $\gamma(.)$ are defined only for a given value of $\theta_\sigma = (\rho, \delta, c)'$, we will contemplate a two-step GMM:

(i) θ_σ is estimated in a first step, from the conditional moment restrictions provided by the ARG(1) specification of the conditional distribution of σ_{t+1}^2 given σ_t^2.

(ii) θ_r is estimated in a second step, from the conditional moment restrictions provided by the

Gaussian specification of the conditional distribution of r_{t+1} given σ_{t+1}^2 and σ_t^2, when θ_σ is replaced by a first step estimator $\hat{\theta}_\sigma$.

While it is well-known that such a two-step GMM procedure is not efficient, we leave for future work the application of an efficient approach as devised in Frazier and Renault (2017). Since, up to the two step issue, the estimation of θ_σ and θ_r raise similar issues, we will set the focus in this methodological section only on the estimation of θ_σ.

5.2 Choice of instruments and identification

While GMM based on realized variance data will be discussed in the next subsection, we first sketch what would be a GMM strategy based on the observation of the volatility factor σ_t^2.

Estimation of θ_σ is then based on the J conditional moment restrictions:

$$\begin{aligned}
E[\exp\left(-u_j\sigma_{t+1}^2\right) - \Psi_{t,\theta_\sigma}(u_j)\left|\sigma_t^2\right] &= 0, j = 1, 2, ..., J \\
\Psi_{t,\theta_\sigma}(u) &= \exp\left(-a(u)\sigma_t^2 - b(u)\right), \theta_\sigma = (\rho, \delta, c)' \\
a(u) &= \frac{\rho u}{1+cu}, b(u) = \delta \log(1+cu)
\end{aligned}$$

where $\{u_j; j = 1, ..., J\}$ is a grid of J values of the real (or complex) number u. It would be theoretically asymptotically optimal to elicit the largest possible grid. However, there are obviously some finite sample bias-variance trade off that will be described in Section 5.4 below.

For a given grid, the exponential affine structure is very convenient for an explicit computation of optimal instruments. They are given by the formula:

$$\left[\frac{\partial \Psi'_{t,\theta_\sigma}(u_1)}{\partial \theta}, ..., \frac{\partial \Psi'_{t,\theta_\sigma}(u_J)}{\partial \theta}\right] \Sigma_{t,\theta_\sigma}^{-1}(u_1, ..., u_J) \tag{5.1}$$

where $\Sigma_{t,\theta_\sigma}^{-1}(u_1, ..., u_J)$ is the $(J \times J)$ matrix whose (j, l) coefficient is:

$$\begin{aligned}
& Cov[\exp\left(-u_j\sigma_{t+1}^2\right), \exp\left(-u_l\sigma_{t+1}^2\right)\left|\sigma_t^2\right] \\
= \; & \Psi_{t,\theta_\sigma}(u_j + u_l) - \Psi_{t,\theta_\sigma}(u_j)\Psi_{t,\theta_\sigma}(u_l)
\end{aligned}$$

In other words, optimal instruments are obtained by combining functions of the type $\exp\left(-a(u)\sigma_t^2 - b(u)\right)$. To keep it simple, this may suggest to work with what Carrasco et al. (2007) have dubbed the Single Index (SI) moments:

$$E\left\{\exp\left(-u_j\sigma_t^2\right)\left[\exp\left(-u_j\sigma_{t+1}^2\right) - \Psi_{t,\theta_\sigma}(u_j)\right]\right\} = 0, j = 1, ..., J$$

However, it is worth reminding that an arbitrary choice of instruments may not deliver identification. As pointed out by Dominguez and Lobato (2004), even the supposedly optimal instruments may not deliver identification. It is then worth proving in the appendix that:

Proposition 5.1:

Assume that the observations $\left\{\sigma_t^2\right\}$ follow a stationary ARG(1) process. Then, for J sufficiently large, the J unconditional moment restrictions:

$$E\left\{\exp\left(-u_j\sigma_t^2\right)\left[\exp\left(-u_j\sigma_{t+1}^2\right)-\Psi_{t,\theta_\sigma}(u_j)\right]\right\}=0, j=1,...,J$$

identify the parameters θ_σ of the ARG(1) model.

In spite of the positive result of Proposition 5.1., we do not expect very accurate GMM estimators since the above unconditional moment conditions are only about the marginal distribution of σ_t^2 and $\left(\sigma_t^2+\sigma_{t+1}^2\right)$. While the marginal distribution of σ_t^2 is of course unable to identify the persistence parameter ρ (it is actually a gamma distribution with parameters δ and $c/\left(1-\rho\right)$)), its comparison with the marginal distribution of $\left(\sigma_t^2+\sigma_{t+1}^2\right)$ does the job but in a very noisy way. It will be of course much more efficient to consider the Double Index (DI) moments defined by:

$$E\left\{\exp\left(-u_k\sigma_t^2\right)\left[\exp\left(-u_j\sigma_{t+1}^2\right)-\Psi_{t,\theta_\sigma}(u_j)\right]\right\}=0, j=1,...,J, k=1,...,K$$

Figures 6 and 7 in the Appendix A report a compelling Monte Carlo evidence about the better accuracy of the DI method. Note that it will not provide in general the optimal instruments since we know from (5.1) that each optimal instrument should include all the indices $u_j, j=1,...,J$.

5.3 Estimation with realized variance

While we have learned from the previous subsections that a DI GMM strategy must be used, we will not in general be able to use it directly on the volatility factor σ_t^2 that we do not observe but only on realized variance that is tightly related:

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

For sake of simplicity, we will assume that $RV_{t+1}=\tilde{\sigma}_{t+1}^2$, meaning that we overlook the correction term $AVar[\sigma_{t+1}^2\left|I(t)\right]$ in (4.3). Our numerical experiments confirm that this approximation is warranted. The important advantage of this approximation is that it amounts to assume that:

$$B=\rho e, D=f\delta c=\delta c\frac{1+e}{2}$$

In particular, it allows us to remain true to our two-step GMM approach:

(i) DI-GMM estimation in a first step of $\theta_\sigma = (\rho, \delta, c)'$ augmented with e. This estimation will use moments as described in the former subsection (conditional moment generating function of σ^2_{t+1} given σ^2_t with DI choice of instruments) but from data on realized volatility RV_{t+1}.

(ii) DI-GMM estimation in a second step of $\theta_r = (\phi, e, \varsigma_1, \varsigma_2)'$ when θ_σ and e are replaced by first step estimators. This estimation will use moments given by the conditional moment generating function of r_{t+1} given $I^\sigma(t)$ (with DI choice of instruments) from data on asset returns. However, we acknowledge that ς_1 is not identified strongly enough from underlying asset return data and we calibrate it on option prices.

Note that to take into account the correction factor $AVar[\sigma^2_{t+1}\,|I(t)]$ we should have either performed a simultaneous estimation of all parameters θ_σ and θ_r or used a convoluted iterative approach (see e.g. Fan, Pastorello, and Renault (2015)).

The purpose of this subsection is to describe the first step (step (i) above). When B is nonzero, we can see that the realized variance is ARMA(1,1) which is not Markov and we cannot construct the conditional characteristic function (CF) in general. Since the joint CF is unknown in general, the simulated method of moments can be used (e.g., Carrasco, Chernov, Florens, and Ghysels (2007)). Instead of using simulated method, we employ GMM with moments based on approximated joint CF assuming the realized variance process is an invertible ARMA (i.e., $|B| < 1$). Then we can invert the ARMA(1,1) realized variance to the AR(∞) process using

$$\sigma^2_t = \sum_{k=0}^{\infty} B^k (RV_{t-k} + D).$$

By plugging this into the conditional CF of volatility given in Section 3.2, we get

$$E\left[\exp\left(-uRV_{t+1}\right)|F_t\right] = \exp\left\{-(a(u) - Bu)\left(\sum_{k=0}^{\infty} B^k RV_{t-k}\right) - b(u) - (a(u) - u)\frac{D}{1-B}\right\}.$$

The estimation based on the above conditional CF is not feasible since we do not observe the infinite number of data lags. However, since we assume that $|B| < 1$, $B^k \to 0$ as $k \to \infty$ and we have an approximation

$$E\left[\exp\left(-uRV_{t+1}\right)|F_t\right] \approx \exp\left\{-(a(u) - Bu)\left(\sum_{k=0}^{H-1} B^k RV_{t-k}\right) - b(u) - (a(u) - u)\frac{D}{1-B}\right\}$$

for some $H < \infty$ so that the realized variance process is approximately CAR(H). We construct the unconditional moment restrictions using the DI-instrument $A_t = \exp(-a_1 RV_t - a_2 RV_{t-1})$ for some $(a_1, a_2)' \in \mathbb{C}^2$. Note that we use the above instrument instead of using $A_t = \exp\left(-\sum_{k=0}^{H} a_k RV_{t-k}\right)$, since the reazlied variance is a restricted CAR(H) approximately and the (approximated) joint CF of $(RV_{t+1}, RV_t, RV_{t-1})'$ is enough to identify the parameters characterizing $a(u)$ and B.

For the estimation of the returns process, we first construct the estimation of the latent volatility factor as

$$\hat{\sigma}_{t+1}^2 = \sum_{k=0}^{H} \hat{B}^k RV_{t-k} + \frac{\hat{D}}{1-\hat{B}}$$

where $\hat{B}$ and $\hat{D}$ are the GMM estimators of B and D, respectively, computed with the realized variance data. Then we estimate the return parameters by GMM with moment conditions $g_t(\tau_2; \theta_r)$ where σ_{t+1}^2 and σ_t^2 are replaced by $\hat{\sigma}_{t+1}^2$ and $\hat{\sigma}_t^2$, respectively.

5.4 GMM estimation

We will use the DI moment conditions of volatility denoted as $E[h_t(\tau_1; \theta_{rv})]$ to estimate $\theta_{rv} = (\theta_\sigma', e)'$ and then in the next step, estimate θ_r with the DI moments of returns denoted as $E[g_t(\tau_2; \theta_r)]$ given the estimates of θ_σ and e. We will discuss the GMM estimation for θ_{rv} in this subsection and apply the same method to the estimation of θ_r. θ_{rv}^0 and θ_r^0 denote the true parameter values for the volatility and returns processes, respectively.

The GMM estimator is defined as:

$$\hat{\theta}_{\sigma,T} = \underset{\theta_{rv}}{Argmin}\ \bar{h}(\tau_1, \theta_{rv})' \hat{W}_T^{-1} \bar{h}(\tau_1, \theta_{rv}), \tag{5.2}$$

where $\bar{h}(\tau_1, \theta_{rv}) = (1/T) \sum_{t=1}^{T} h_t(\tau_1; \theta_{rv})$ and $\hat{W}_T$ is a sample analog of a positive definite matrix W s.t. $\hat{W}_T \xrightarrow{p} W$. It is already well-established that the optimal weighting matrix that leads to the smallest asymptotic variance among the class of a GMM estimator is

$$W = V = E\left[h_t(\tau_1; \theta_{rv}^0) h_t(\tau_1; \theta_{rv}^0)'\right],$$

and $\hat{W}_T = \hat{V}_T = (1/T) \sum_{t=1}^{T} h_t(\tau_1; \tilde{\theta}_{rv,T}) h_t(\tau_1; \tilde{\theta}_{rv,T})'$, where $\tilde{\theta}_{rv,T}$ is a preliminary consistent estimator of θ_{rv}^0 computed using, for example, an identity matrix as a weighting matrix.

However, even with a small dimensional τ_1, $\hat{W}_T$, the sample analog of the optimal weighting matrix, may not be invertible (or very close to be singular) and this can result in unstable estimation. We, in order to ensure consistent estimation, employ the Tikhonov method of regularization introduced by Carrasco and Florens (2000). That is, we replace V by a perturbed version of it using a tuning parameter $\alpha > 0$ such that:

$$W^{-1} = (V^2 + \alpha I)^{-1} V \tag{5.3}$$

where I is an identity matrix[2]. V is an unknown population moment and hence, we use the sample analog of W^{-1}:

$$\hat{W}_T^{-1} = (\hat{V}_T^2 + \alpha I)^{-1}\hat{V}_T.$$

Then the GMM estimator is defined as (5.2) with above $\hat{W}_T^{-1}$, the regularized optimal weighting matrix[3].

$\hat{\theta}_{rv,T}$ can then be shown to be consistent and asymptotically normal under some regularity conditions (e.g., conditions in theorem 2.6 and theorem 3.4 in Newey and McFadden (1994)). We have ensured that θ_{rv} is identified from the moment conditions and W^{-1} is nonsingular for some user-chosen $\alpha > 0$. The asymptotic distribution of $\hat{\theta}_{rv,T}$ is provided as following.

$$\begin{aligned}
\sqrt{T}\left(\hat{\theta}_{rv,T} - \theta_{rv}^0\right) &\xrightarrow{d} \mathcal{N}(0, A), \\
&A = E\left[G'W^{-1}G\right]^{-1} G'W^{-1}VWGE\left[G'W^{-1}G\right]^{-1}, \\
&G = E\left[\nabla_{\theta_{rv}} h_t(\tau_1; \theta_{rv}^0)\right].
\end{aligned}$$

Note that we do not use or propose a data dependent selection method of a turning parameter that leads to the efficient estimation since it is beyond the scope of this paper. Our focus is the consistent estimation and we will choose some positive α that is big enough to ensure a small enough bias.

6 Numerical illustrations

In this section we provide several Monte Carlo simulation results. The aim of these exercises is not to study the finite sample behavior of our GMM estimators but to illustrate the identifiability of leverage effect from discrete time affine models. We use the leading example of this paper, the ARG(1)-normal model for the SV model. We also present the results using a GARCH-type option pricing model that is another large class of discrete time models and make a comparison with the SV model. For a GARCH model, we consider the Heston and Nandi (2000)'s affine GARCH(1,1) model (HN hereafter).

[2](5.3) is computed from the solution to the Ridge regression problem

$$\underset{g}{Min}\, \|Vg - f\|^2 + \alpha\, \|g\|$$

for some finite dimensional vector f, where $\|\cdot\|$ denotes an l2-norm.

[3]We use an identity weighting matrix to compute $\tilde{\theta}_{rv}$ that $\hat{V}_T$ is computed with.

6.1 HN model

HN assume the following process for daily excess log returns

$$r_{t+1} = log(S_{t+1}/S_t) - r_{f,t} = \lambda h_t - \frac{1}{2}h_t + \sqrt{h_t}\epsilon_{t+1},$$

where $\epsilon_{t+1} \sim i.i.d.\mathcal{N}(0,1)$ and λ is the risk price of returns. The conditional variance h_t has the following process

$$h_{t+1} = \omega + \beta h_t + \alpha(\epsilon_{t+1} - \gamma\sqrt{h_t})^2,$$

where γ captures the asymmetric relationship between returns and volatility. The persistence of daily variance is captured by the form $(\beta + \alpha\gamma^2)$. The covariance between returns and volatility and the volatility of volatility can be derived as

$$Cov\left[r_{t+1}, h_{t+1}|I(t)\right] = -2\alpha\gamma h_t,\ Var(h_{t+1}|F_t) = 2\alpha^2(1 + 2\gamma^2 h_t).$$

The correlation coefficient between returns and volatility is then

$$Corr\left(r_{t+1}, h_{t+1}|I(t)\right) = \frac{-2\gamma\sqrt{h_t}}{\sqrt{2 + 4\gamma^2 h_t}}.$$

It can be seen that the negative correlation between returns and volatility increases (for positive γ) as γ gets larger (for a fixed h_t).

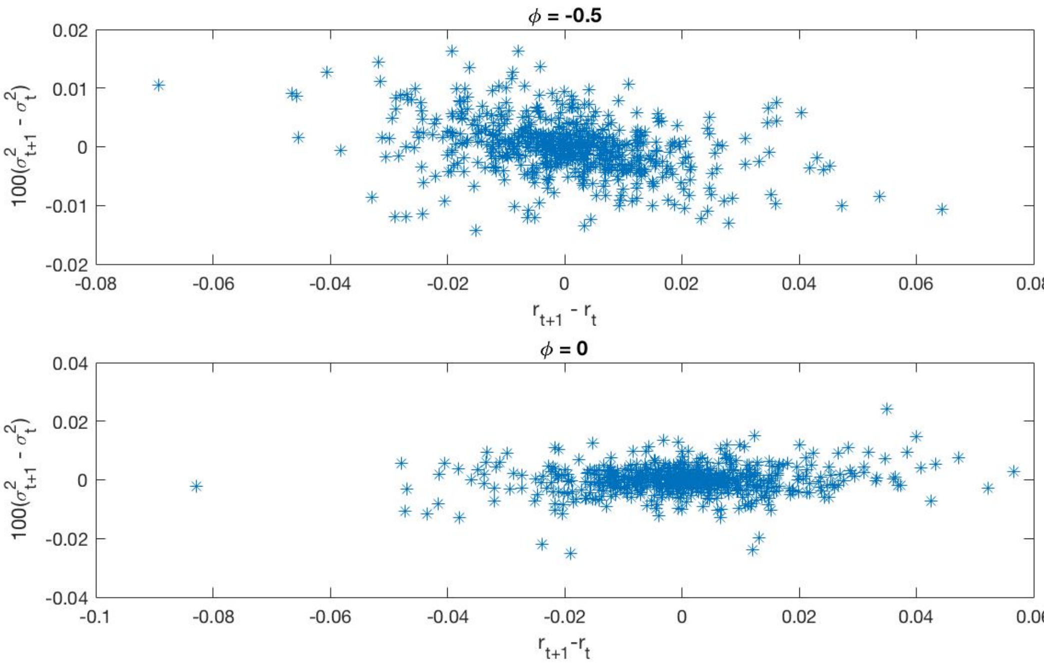
6.2 Monte Carlo evidence

6.2.1 Data generating process

The returns and volatility data are generated from the ARG(1)-Normal SV model derived in this paper with parameter values $\rho = 0.9, \delta = 1.1, E[\sigma^2_{t+1}] = 0.04/365$[4], and $e = 0$. The sampling frequency is one day with number of observations $T = 600$. We simulate the data separately for two cases with $\phi = -0.5$ (significant leverage effect) and with $\phi = 0$ (no leverage effect). Figure 1 shows the plots of the daily difference in volatility $(\sigma^2_{t+1} - \sigma^2_t)$ against the daily difference in returns $(r_{t+1} - r_t)$ for both cases.

[4]The scale parameter c is computed from the equation $E[\sigma^2_{t+1}] = \delta c/(1-\rho)$.

Figure 1: Daily correlation between changes in volatility and changes in returns



* The estimated correlations are -0.35 and 0.12 for the first and the second graph, respectively.

6.2.2 QML estimation of HN

In this section, we estimate the HN parameters $(\omega, \alpha, \beta, \gamma, \lambda)$ from this potentially misspecified SV model using the Quasi Maximum Likelihood (QML) estimation to see whether the leverage effect can be reproduced with reasonable estimates of other parameter values by the HN model. One such parameter is obviously γ. It captures the asymmetric correlation between returns and volatility that is shown through ϕ. γ is expected to be positive and to be increasing as the absolute value of ϕ increases. Another important parameter is the level of volatility persistence ($\rho = 0.9$) that is shown by $\beta + \alpha\gamma^2$ in the HN model. The Table 1 presents the estimation results.

As we can see from the Table 1, the HN model estimates γ to be positive for all $\phi = -0.1, -0.3, -0.5$ which means a negative correlation between returns and volatility. The γ estimate is also positive for $\phi = 0$ but it is not statistically significant. We plot the daily conditional correlation between returns and volatility implied by the HN model in Figure 2 and find that, although the model implied correlation becomes stronger for a larger (in absolute value) ϕ, it does not capture the true correlations correctly. The correlation seems to fluctuate at around -0.8 when $\phi = -0.3$, which shows that the model implied correlation is much bigger than the true correlation. Moreover, as the true leverage effect decreases (so that the estimated γ decreases), the estimated level of persistence of volatility departs from the true persistence ($\rho = 0.9$) significantly. For example, $\beta + \alpha\gamma^2$ is estimated to be

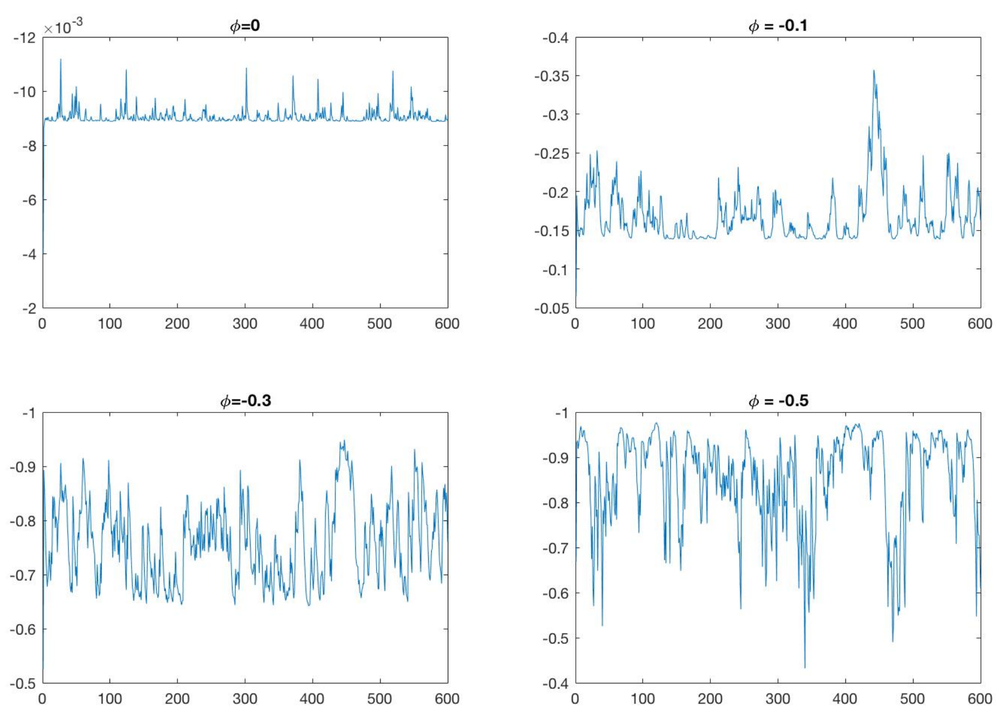
Table 1: HN parameter estimates for different leverage effects

Leverage effect (ϕ)	$\phi = 0$	$\phi = -0.1$	$\phi = -0.3$	$\phi = -0.5$
λ	1.7085	-7.5103	-26.1236	-35.9649
ω	4.13e-04	3.97e-04	9.53e-05	8.09e-09
α	8.21e-05	1.02e-04	1.09e-04	1.04e-04
β	0.2366	0.2289	0.4976	0.4560
γ	0.2699	4.3615	41.6121	60.6573
$\beta + \alpha\gamma^2$	0.2366	0.2309	0.6855	0.8393

* The parameter ω is computed from the equality $E[h_t] = (\omega + \alpha)/(1 - \beta - \alpha\gamma^2)$.

around 0.24 when there is no leverage effect.

Figure 2: Daily conditional correlation between returns and volatility



6.2.3 Patterns of implied volatility smiles

In this subsection, we simulate option prices for different strike prices and maturities to show how models identify a leverage effect from using options. The option prices are generated from each model with the spot price $S = 100$ and various strike prices $K = (91, 92, \cdots, 100, \cdots, 119, 110)'$ so that the log moneyness ($log(K/S)$) is between -10% and 10%.

Figure 3 and 4 present the shapes of volatility smiles from the SV model and the HN model, respectively, with various levels of a leverage effect (ϕ and γ). The time to maturity is $T = 30$ days for both graphs and the level of volatility is given as 0.04/365. As we can see, both the SV model and HN model produce a symmetric volatility smile when there exists no leverage effect[5] with the implied volatility minimized at the money, and symmetry starts to be distorted as the leverage effect increases.

This is consistent with what we have studied in Section 2 (also in Renault (1997)) that we get a volatility skewness in the presence of a leverage effect due to the stochastic nature of volatility and a latent stock price ($\tilde{S}_t$)). See how a small discrepancy of $\tilde{S}_t$ from S_t distorts the smiles of both models in Appendix A (Figure 8-11). In addition, even though volatility is not stochastic in GARCH models, volatility is presently (at time t) not known if we consider relatively long time to maturity and we get a desired volatility skewness with a leverage effect[6]. However, the volatility smile becomes linear with a negative slope when γ gets too big (simulations with various parameter values show that it looks linear when $\gamma > 100$) in the HN model and it is hard to identify the leverage effect in this case. These simulation exercises verify that the shape of volatility smile is closely related to the leverage effect, which provides a ground for identifying a leverage effect from discrete time models, but such relationship is stronger for the SV model than the GARCH model.

[5] The leverage effect is zero when $\phi = 0$ for the SV model and $\gamma = 0$ for the HN model.

[6] We get a flat volatility smile for $T = 1$ with the GARCH models.

Figure 3: Volatility smiles from the SV model

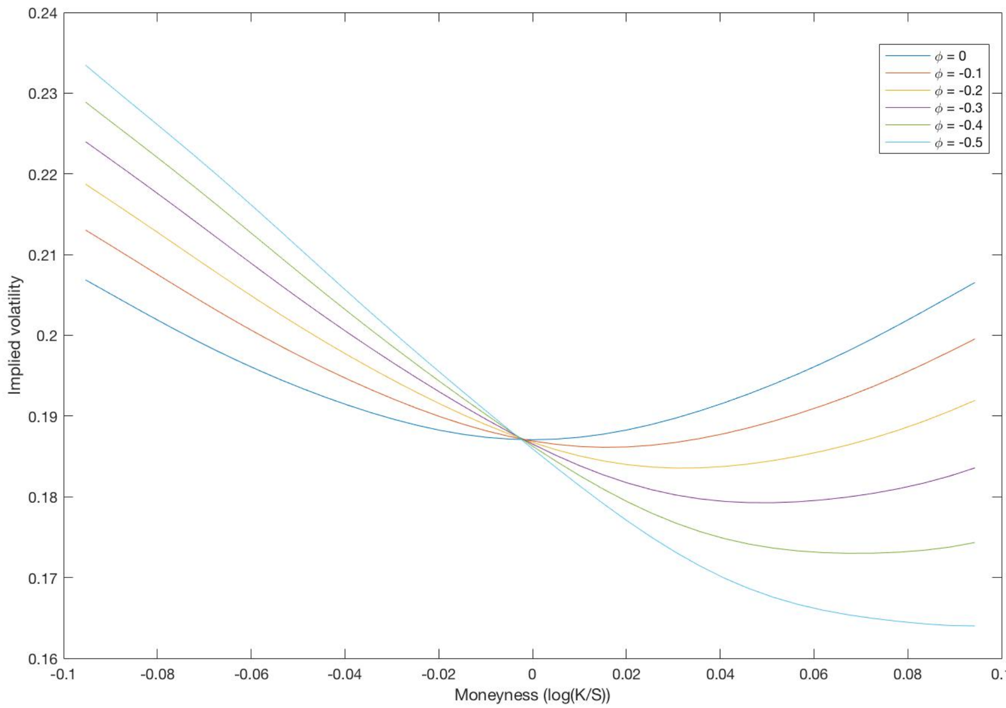
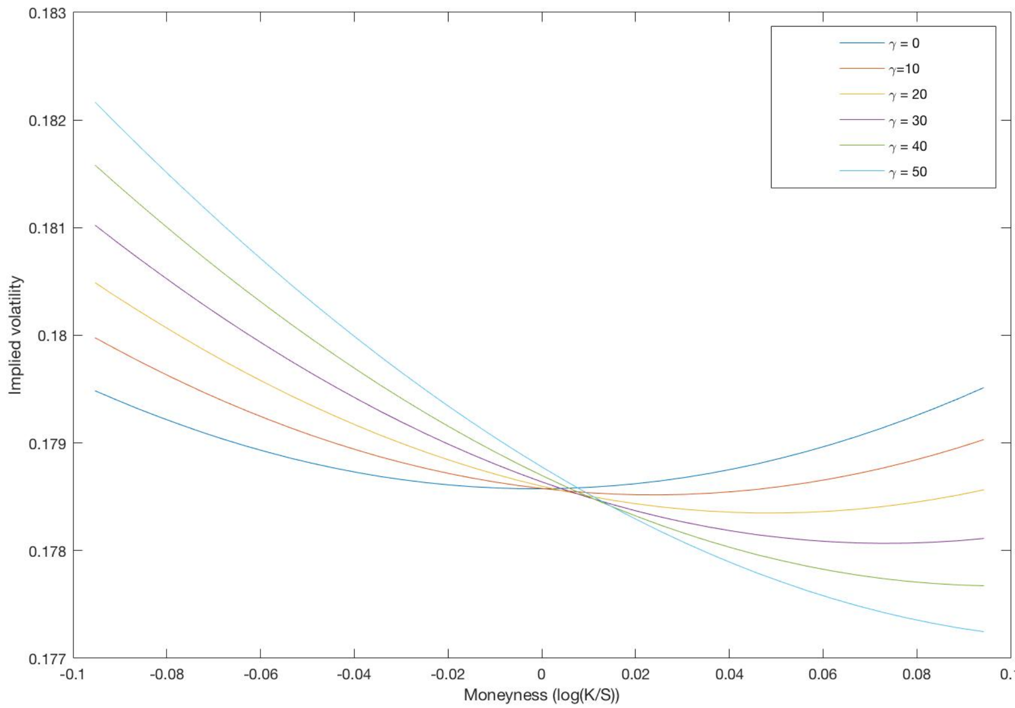


Figure 4: Volatility smiles from the HN model



* The implied volatility is generated with the risk neutral parameter values ($\alpha = 1.1e-06, \beta = 0.9, E[h_t] = 7.73e-05$) and the volatility 0.04/365 with different γ's.

6.3 Empirical results

6.3.1 Data

In this section, we present an empirical application of the our SV model applying the GMM estimation discussed in Section 5. The dataset was obtained from Oxford-Man Institute[7] 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[8]. The realized variance process $\{RV_t\}$ is computed from 5-minute intraday returns.

[7]Oxford-Man Institute's "realized library", http://realized.oxford-man.ox.ac.uk
[8]This rate is obtained from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data library.html

6.3.2 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 returns process r_{t+1} is Gaussian given $I^\sigma(t)$ with the functional forms of $a(\cdot)$, $b(\cdot)$, $\alpha(\cdot)$, $\beta(\cdot)$, and $\gamma(\cdot)$ given in Section 4.3. However, as discussed in Section 5.3, σ_t^2 is not observable and we use instead the daily observations of realized variance[9]:

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

We apply the two-step GMM procedure described (i) and (ii) in Section 5.3. We first construct the DI-moments described in the same section and estimate the parameters $\theta_{rv} = (\theta_\sigma', e)' = (\rho, \delta, c, e)'$. Let $h_t(\tau_1; \theta_{rv})$ denote such moments of realized variance with $\tau_1 = (a_1, a_2, u)'$, where a_1, a_2, and u are each vectors of 5 equally spaced complex numbers on an interval $[1, 10] \times 1i$. Then the moment conditions that we exploit are

$$E \begin{bmatrix} Re\left\{h_t(\tau_1; \theta_\sigma)\right\} \\ Im\left\{h_t(\tau_1; \theta_\sigma)\right\} \end{bmatrix}$$

where $Re\{a\}$ and $Im\{a\}$ are, respectively, the real and imaginary part of a complex vector a. This gives us $5 \times 5 \times 5 \times 2 = 250$ number of moment conditions in total. The approximation parameter $H = 10$ is chosen since our numerical experiments show that the estimation results do not depend on H much for $H \geq 10$. Lastly, we use the Tikhonov regularized weighting matrix as described in Section 5.4 with $\alpha = 0.05$.

Then we construct the volatility using $\hat{\theta}_{rv}$, the estimated θ_{rv} by approximating the form of σ_t^2 converted from realized variance:

$$\hat{\sigma}^2 = \sum_{k=0}^{H} (\hat{e}\hat{\rho})^k \left(RV_{t-k} + \hat{f}\hat{\delta}\hat{c}\right).$$

The DI-moments of returns are then formed with

$$E\left[\exp(-vr_{t+1})|I^\sigma(t)\right] = \exp\left\{-\alpha(v)\hat{\sigma}_{t+1}^2 - \beta(v)\hat{\sigma}_t^2 - \gamma(v)\right\}.$$

and the DI-instrument $A_t^\sigma = \exp\left(-b_1\hat{\sigma}_{t+1}^2 - b_2\hat{\sigma}_t^2\right)$ where v, b_1 and b_2 are each vectors of 4 equally spaced complex numbers on an interval $[-1, 1] \times 1i$. Let $\tau_2 = (b_1, b_2, v)'$ and $g_t(\tau_2; \theta_r)$ denote such DI-moment functions with e in θ_r replaced by $\hat{e}$. In the same manner with the estimation of θ_σ, we

[9] See the discussion given in Section 5.3 for this specification of realized variance.

use the following DI moment condition

$$E\begin{bmatrix} Re\left\{g_t(\tau_1;\theta_r)\right\} \\ Im\left\{g_t(\tau_1;\theta_r)\right\} \end{bmatrix},$$

and the regularization parameter $\alpha = 0.05$ is chosen.

As described in Section 5.3, the volatility risk parameter ς_1 is weakly identified from the returns data and we calibrate it with the options data. Specifically, we choose ς_1 that minimizes IVRMSE put forward by Renault (1997):

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

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

The estimation results are given in Table 2 with those of the HN model. Note that there is no volatility risk price (ς_1) in the HN model since it uses only the risk price of returns. The volatility persistence is estimated to be around 0.94 and 0.97 for the ARG(1)-Normal and the HN model, respectively, and both results are consistent with the empirical findings of high level of persistence in the literature. The estimated risk prices of returns (ς_2) of both models are also shown to be similar.

However, those two models produce a significance difference in reproducing the leverage effect. In the ARG(1)-Normal model, the leverage effect ϕ is estimated to be around -0.45. This high absolute value of ϕ provides empirical evidence of the existence of the significant leverage effect and shows that our model is able to capture such effect. On the other hand, the leverage effect, which is the conditional correlation coefficient between returns and volatility (formula given in Section 6.1), is almost a constant close to -1 (see Figure 12 in Appendix A). This result is in line with the restriction of the GARCH model that the innovations of returns and volatility share the same process.

6.4 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[10] and the observations range from January 3, 2000 to January 3, 2013. In order to ensure that we consider liquid options, we only maintain the ones with time to maturity[11] between 15 and 180 days and restrict our data to Wednesday options. Also, the observations with an implied volatility of more than 70% are discarded. Moreover, we only

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

Table 2: Estimates of the parameters

param.	ARG(1)-Normal	HN
ρ	0.9405	
	(0.0503)	
δ	0.6475	
	(0.2111)	
c	1.56e-5	
	(6.04e-6)	
e	0.3386	
	(0.0291)	
ω		6.88e-9
α		3.81e-6
		(3.83e-7)
β		0.7771
		(0.0047)
$E[h_t]$		1.11e-4
Persistence	0.9405	0.9654
ϕ	-0.4465	
	(0.0763)	
ς_2	1.7680	
	(2.0062)	
ς_1	-10	
λ		1.3986
		(1.3223)
γ		222.26
		(11.4435)

* The standard errors are given in parentheses.
* ω is computed as $E[h_t]\left(1-\beta-\alpha\gamma^2\right)-\alpha$ where the unconditional mean $E\left[h_t^{RV}\right]$ is the sample mean of realized variance.

Table 3: Option pricing performances

	$ARG(1) - Normal$	HN
$IVRMSE$	4.7443	6.0381

consider out of the money call options in order to maintain the data in a manageable size. The same analysis can be done for put options as well. The total number of observations is 39,158.

We categorize options according to their time to maturity and moneyness where moneyness is defined as (K/S_t) with K and S_t denoting a strike price and a price of the underlying asset at time t.

We 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 and the HN models. In order to analyze the option pricing performances of each model, we use the percentage implied volatility (IVRMSE) defined in the previous subsection. The results are presented in Table 3 which shows that our model outperforms the HN model in terms of the option pricing errors.

Table 4 presents some descriptions of the options data and IVRMSE for each maturity and moneyness category of the $ARG(1)$-Normal model and the HN model. In terms of IVRMSE, the option pricing performance of the ARG(1)-normal model 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 IVRMSE for both models.

7 Conclusion

We have addressed in this paper two identification issues that are known to be puzzling. They are both related to leverage effect.

First, as documented by Bollerslev et al. (2006), discrete time return data do not allow to disentangle leverage effect from volatility feedback. In the context of a conditional distribution of return that is a mixture of lognormal, we are able to pin down the parameter that properly characterizes the amount of leverage effect since it is the only responsible for skewness of the volatility smile. From this benchmark, we are able to write down an identification constraint that relates three parameters:

First, a parameter for the joint occurrence of leverage and volatility feedback in conditional mean of return given current volatility (our parameter ψ),

Second, the price of risk on asset return that is responsible for volatility feedback (our parameter ς_2),

Table 4: Option pricing performances by maturity and moneyness (K/S)

By Maturity	Less than 30	30 to 60	60 to 120	120 to 150	More than 150
No. of obs	7,653	16,498	11,033	1,888	2,068
Ave. IV(%)	18.93	18.56	18.49	18.39	18.34
$ARG(1)$-Normal					
IVRMSE	4.6444	4.7466	4.7538	4.8498	4.9241
HN					
IVRMSE	7.2305	6.1030	5.5282	4.9849	5.2378

By Moneyness	Less than 1.02	1.02 to 1.04	1.04 to 1.06	1.06 to 1.1	More than 1.1
No. of obs	7,004	6,428	6,069	9,431	10,226
Ave. IV(%)	18.14	16.94	16.49	17.35	22.34
$ARG(1)$-Normal					
IVRMSE	4.6403	4.6075	4.5516	4.4388	5.2569
HN					
IVRMSE	4.7497	4.8922	5.1410	6.2750	10.6456

* Moneyness is K/S.
* IV stands for Implied Volatility.

Third, the leverage effect parameter ϕ that matters for the conditional of variance of return given current volatility.

This constraint is devised in order that the parameter ϕ is sole responsible for the skewness of the volatility smile. The direction of its impact can be characterized in closed form, at least in the neighborhood of $\phi = 0$.

A second classical identification issue is about the parameter ς_1 of price of volatility risk. It is often said that only option price data allow to identify this parameter. Interestingly enough, we prove that the price of volatility risk is actually identified from return data only if and only if our leverage effect parameter is non-zero.

Even though the paper also provides some compelling empirical evidence that the model is validated by a reasonable goodness of fit (and sensible values of estimated parameters) on the S&P500 daily data, it is obviously for statistical fit and inference that the paper paves the way for future research.

First, even though theoretically ensured through leverage, identification of volatility risk price without option price data is not compelling empirically. Following an argument put forward by Bandi and Reno (2016), we suspect that our identification strategy based on leverage would be more reliable when reinforced by jumps in both returns and volatility. While capturing jumps with a discrete time model is a challenging task, a Factorial Hidden Markov a la Augustyniak et al. (2018) would do the job.

Second, admitting that strong identification can be ensured, a good deal of work remains to be done for efficient estimation. The fact that the volatility factor should be filtered from data on daily realized variance implies complicated nonlinear interactions between the different parts of the model. In this paper, we have simplified the estimation task by making approximations allowing a two-step procedure: first estimation of the volatility dynamics to filter the volatility factor and second estimation of the return dynamics is based on first stage estimators of both filtered values of volatility and coefficients of identification constraint for leverage. Besides the hopefully negligible bias implied by our approximation, the multi-step estimation procedure should be revisited in the spirit of Fan et al. (2015) to ensure asymptotic efficiency of estimators . Since, as documented by Ait-Sahalia et al. (2013), the leverage effect puzzle is also due to an estimation challenge, efficient estimation should be of foremost importance.

References

Ait-Sahalia, Y., J. Fan, and Y. Li (2013): "The leverage effect puzzle: Disentangling sources of bias at high frequency," *Journal of Financial Economics*, 109, 224–249.

Augustyniak, M., L. Bauwens, and A. Dufays (2018): "A new approach to volatility modeling: The Factorial Hidden Markov volatility model," *Journal of Business and Economic Statistics*, published online.

Bandi, F. and R. Reno (2016): "Price and volatility co-jumps," *Journal of Financial Economics*, 119, 107–146.

Barone-Adesi, G., R. Engle, and L. Mancini (2008): "GARCH option pricing model with filtered historical simulation," *Review of Financial Studies*, 21, 1223–1258.

Bertholon, H., A. Monfort, and F. Pegoraro (2008): "Econometric asset pricing modelling," *Journal of Financial Econometrics*, 6, 407–458.

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

Bollerslev, T., J. Litvinova, and G. Tauchen (2006): "Leverage and volatility feedback effects in high-frequency data," *Journal of Financial Econometrics*, 4, 353–384.

Brownlees, C. and G. Gallo (2010): "Comparison of volatility measures: A risk management perspective," *Journal of Financial Econometrics*, 8, 29–56.

Carrasco, M., M. Chernov, J. Florens, and E. Ghysels (2007): "Efficient estimation of general dynamic models with a continuum of moment conditions," *Journal of Econometrics*, 140, 529–573.

Carrasco, M. and J. Florens (2000): "Generalization of GMM to a continuum of moment conditions," *Econometric Theory*, 16, 797– 834.

Christoffersen, P., R. Elkamhi, B. Feunou, and K. Jacobs (2010): "Option valuation with conditional heteroskedasticity and nonnormality," *Review of Financial Studies*, 23, 2139–2183.

Christoffersen, P., K. Jacobs, and C. Ornthanalai (2013): "GARCH option valuation: Theory and evidence," *The Journal of Derivatives*, 21, 8–41.

Cox, J., J. Ingersoll, and S. Ross (1985): "A theory of the term structure of interest rates," *Econometrica*, 53, 385–407.

Darolles, S., C. Gourieroux, and J. Jasiak (2006): "Structural Laplace transform and compound autoregressive models," *Journal of Time Series Analysis*, 27, 477–503.

DOMINGUEZ, M. AND I. LOBATO (2004): "Consistent estimation of models defined by conditional moment restrictions," *Econometrica*, 72, 1601–1615.

DUAN, J.-C. (1995): "The GARCH option pricing model," *Mathematical Finance*, 5, 13–32.

DUFFIE, D., J. PAN, AND K. SINGLETON (2000): "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica*, 68, 1343–1376.

FAN, Y., S. PASTORELLO, AND E. RENAULT (2015): "Maximization by parts in extremum estimation," *The Econometrics Journal*, 18, 147–171.

FELLER, W. (1951): "Two singular diffusion problems," *Annals of Mathematics*, 54, 173–182.

FEUNOU, B. AND R. TEDONGAP (2012): "A stochastic volatility model with conditional skewness," *Journal of Business and Economic Statistics*, 30, 576–591.

FRAZIER, D. AND E. RENAULT (2017): "Efficient two-step estimation via targeting," *Journal of Econometrics*, 201, 212–227.

GAGLIARDINI, P., C. GOURIEROUX, AND E. RENAULT (2011): "Efficient derivative pricing by the extended method of moments," *Econometrica*, 79, 1181–1232.

GARCIA, R., R. LUGER, AND E. RENAULT (2003): "Empirical assessment of an intertemporal option pricing model with latent variables," *Journal of Econometrics*, 116, 49–83.

——— (2005): "Viewpoint: Option prices, preferences, and state variables," *Canadian Journal of Economics*, 38, 1–27.

GARCIA, R. AND E. RENAULT (1998): "A note on GARCH option pricing and hedging," *Mathematical Finance*, 8, 153–161.

GOURIEROUX, C. AND J. JASIAK (2006): "Autoregressive gamma processes," *Journal of Forecasting*, 25, 129–152.

HESTON, S. (1993): "A closed form solution for options with stochastic volatility with applications to bond and currency options," *The Review of Financial Studies*, 6, 327–343.

HESTON, S. AND S. NANDI (2000): "A closed-form GARCH option valuation model," *The Review of Financial Studies*, 13, 585–625.

MEDDAHI, N. AND E. RENAULT (2004): "Temporal aggregation of volatility models," *Journal of Econometrics*, 119, 355–379.

MYKLAND, P. AND L. ZHANG (2009): "Inference for continuous semimartingales observed at high frequency," *Econometrica*, 77, 1403–1445.

Newey, W. and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 4, 2111–2245.

Pan, J. (2002): "The jump-risk premia implicit in options: evidence from an integrated time-series study," *Journal of Financial Economics*, 63, 3–50.

Renault, E. (1997): "Econometric models of option pricing errors. In D. M. Kreps & K. F. Wallis (Eds.)," *Advances in Economics and Econometrics: Theory and Application*, 3, 223–278.

Renault, E., K. Sekkat, and A. Szafarz (1998): "Testing for spurious causality in exchange rates," *Journal of Empirical Finance*, 5, 47–66.

Renault, E. and N. Touzi (1996): "Option hedging and implied volatilities in a stochastic volatility model," *Mathematical Finance*, 6, 279–302.

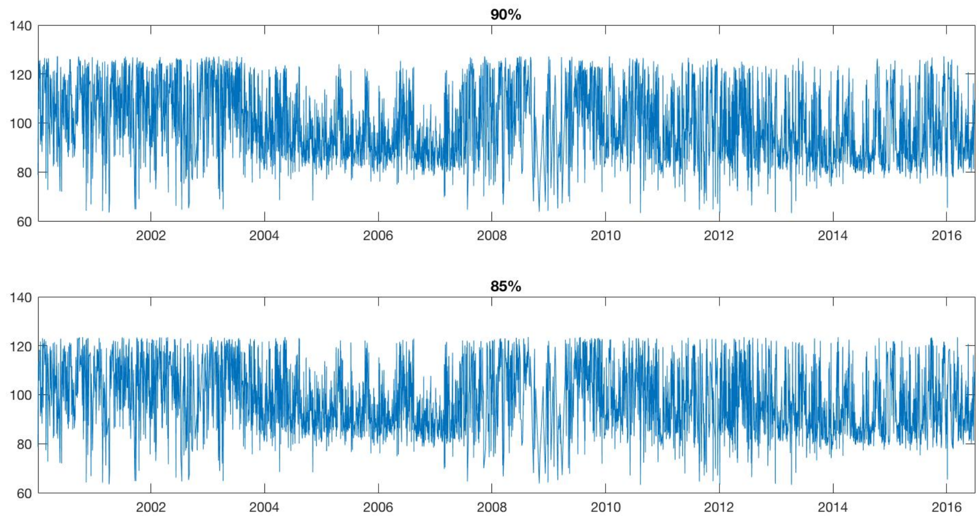
Romano, M. and N. Touzi (1997): "Contingent claims and market completeness in a stochastic volatility model," *Mathematical Finance*, 7, 399–410.

Shephard, N. and K. Sheppard (2010): "Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models," *Journal of Applied Econometrics*, 25, 197–231.

Appendix

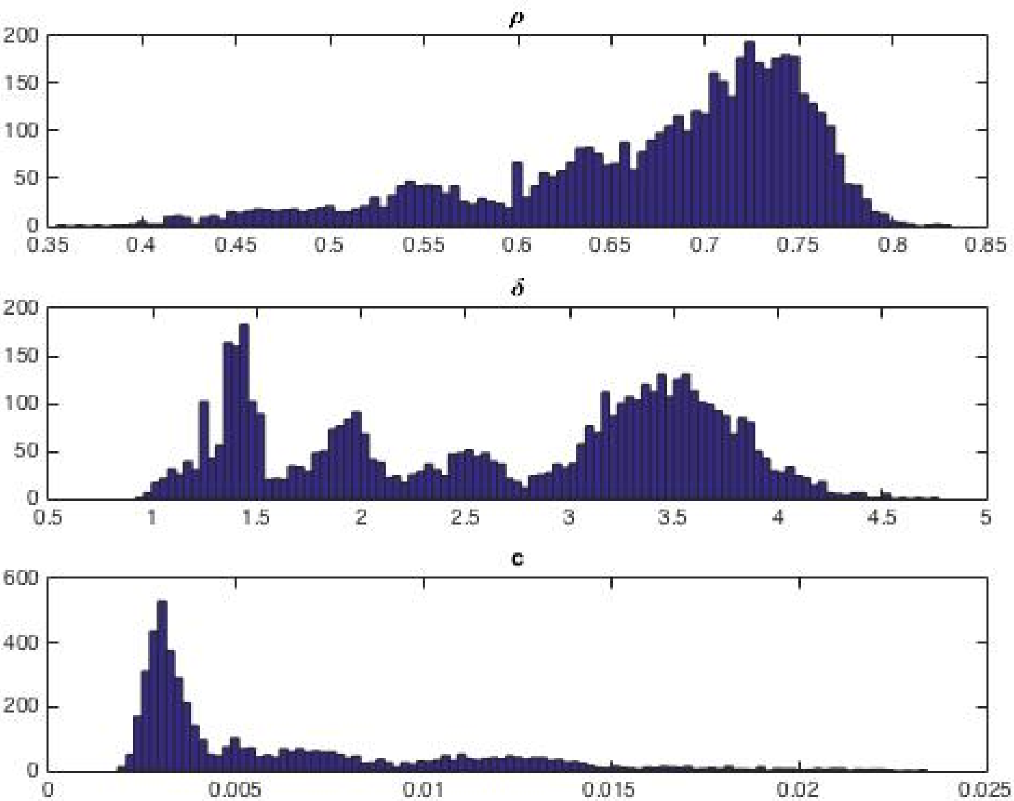
Appendix A

Figure 5: Time series of $\sqrt{E\left[RV_{t+1}|I(t)\right]/Var\left[RV_{t+1}|I(t)\right]}$



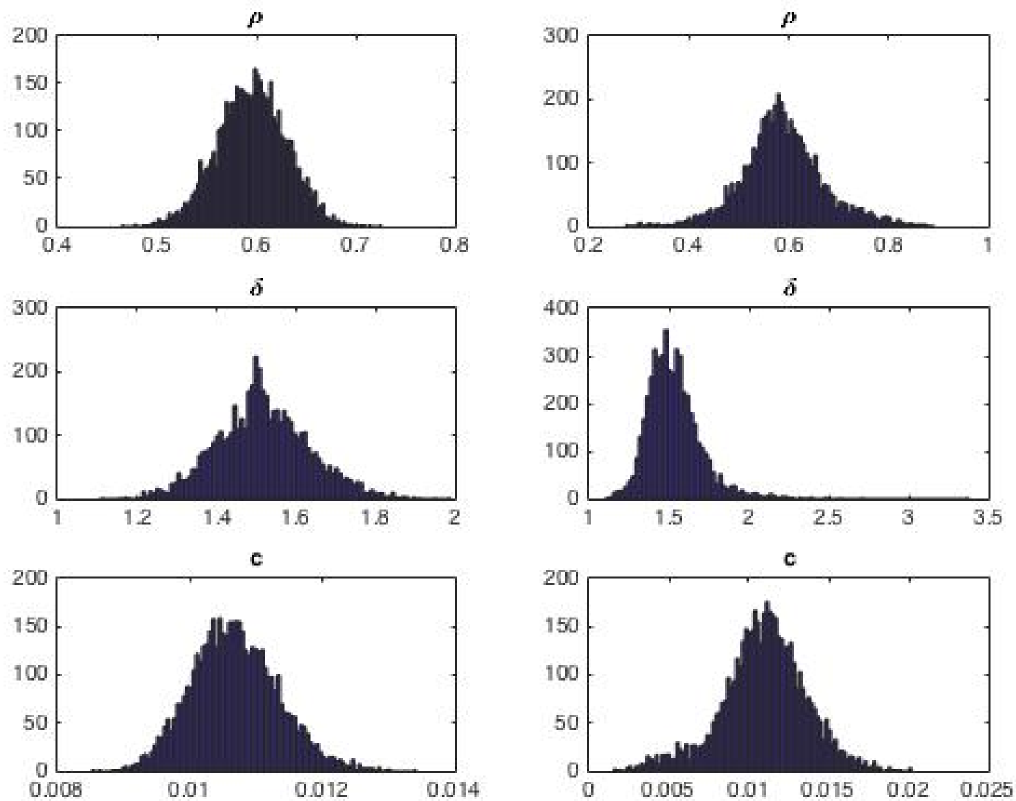
** $\sqrt{E\left[RV_{t+1}|I(t)\right]/Var\left[RV_{t+1}|I(t)\right]}$ is calculated by fitting realized variance with AR(1) with ARCH(1) disturbance.*
** The first graph excludes the 5% largest and 5% smallest values.*
** The second one excludes the 10% largest and 5% smallest values.*

Figure 6: Distribution of GMM estimates for ARG(1) volatility model with the SI-moments



* *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 7: Distribution of GMM estimates for ARG(1) volatility model with the DI-moments



** The true values are:* $(\rho^0 = 0.6, \delta^0 = 1.5, c^0 = 0.0106)$.

** We used 5 equally spaced* u*'s on* $[1i, 10i]$.

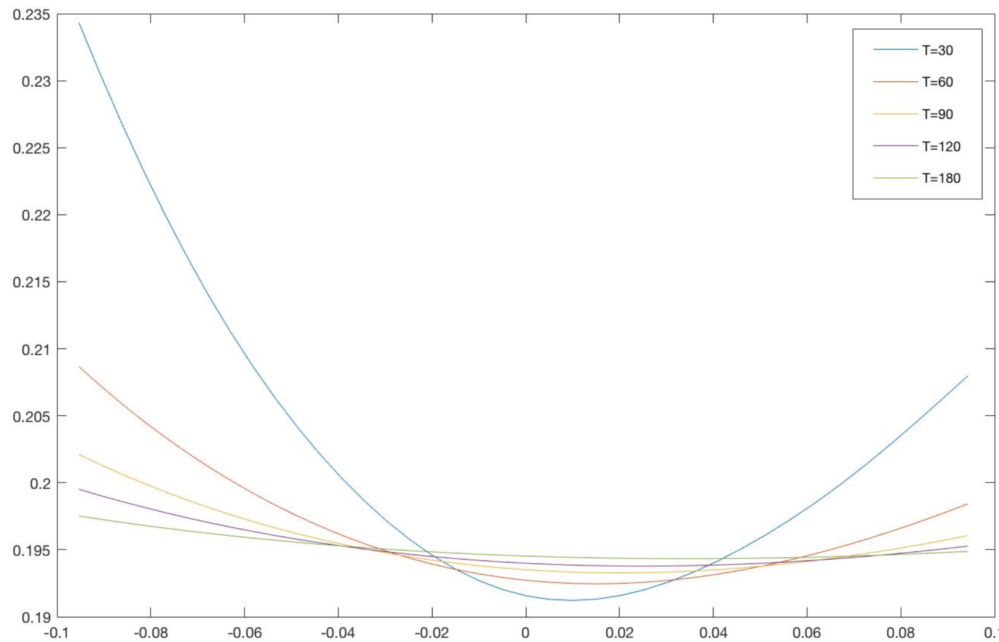
** The right hand side panel used* $v = 1i$ *and the left hand side one used* $v = 1i$ *and* $v = 10i$.

** An identity weighting matrix is used.*

** 10 randomly generated values were used as initial values for each* ρ, δ, *and* c.

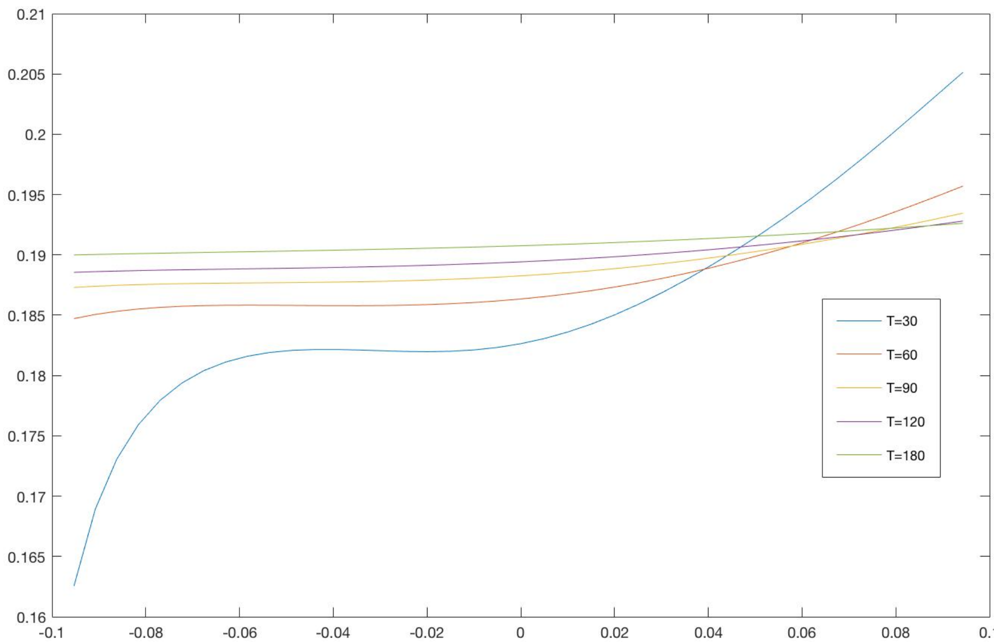
** 5000 replications*

Figure 8: Distortion of smile of the ARG(1)-Normal model



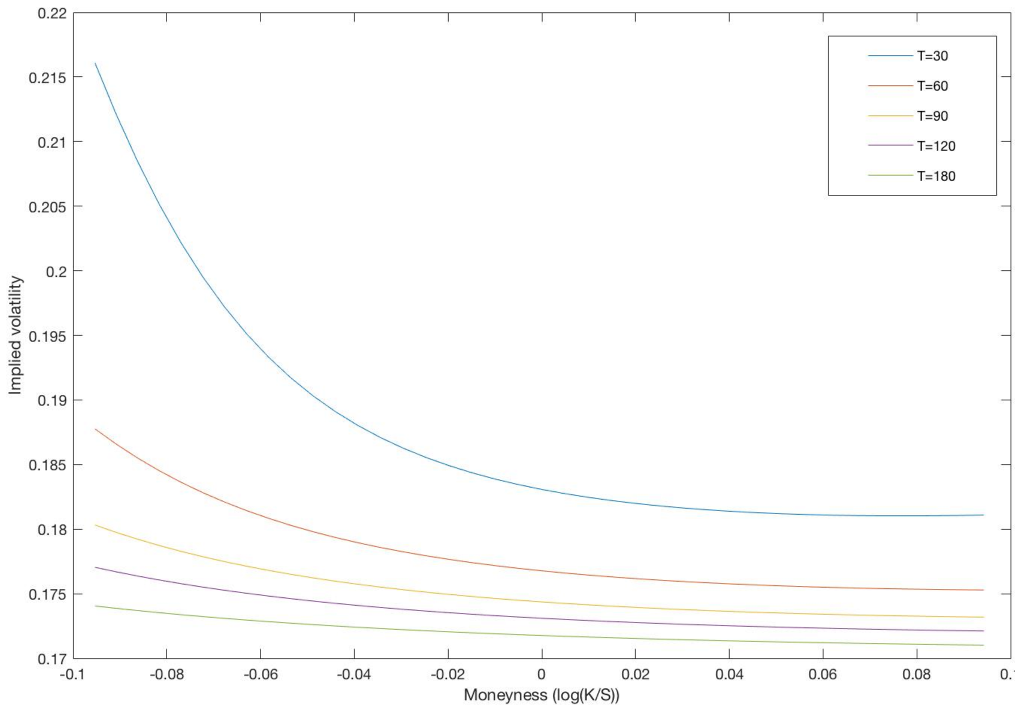
** The option prices are computing with the latent stock price* $\tilde{S}_t = (1.001) \times S_t = (1.001) \times (100)$.
** The Black-Scholes implied volatilities are computed using the observed stock price* $S_t = 100$.
** The moneyness is* $log(K/S_t)$ *where* $K = [91, 92, \cdots, 100, 101, \cdots, 110]$.
*** T *represents the time to maturity in days.*

Figure 9: Distortion of smile of the ARG(1)-Normal model



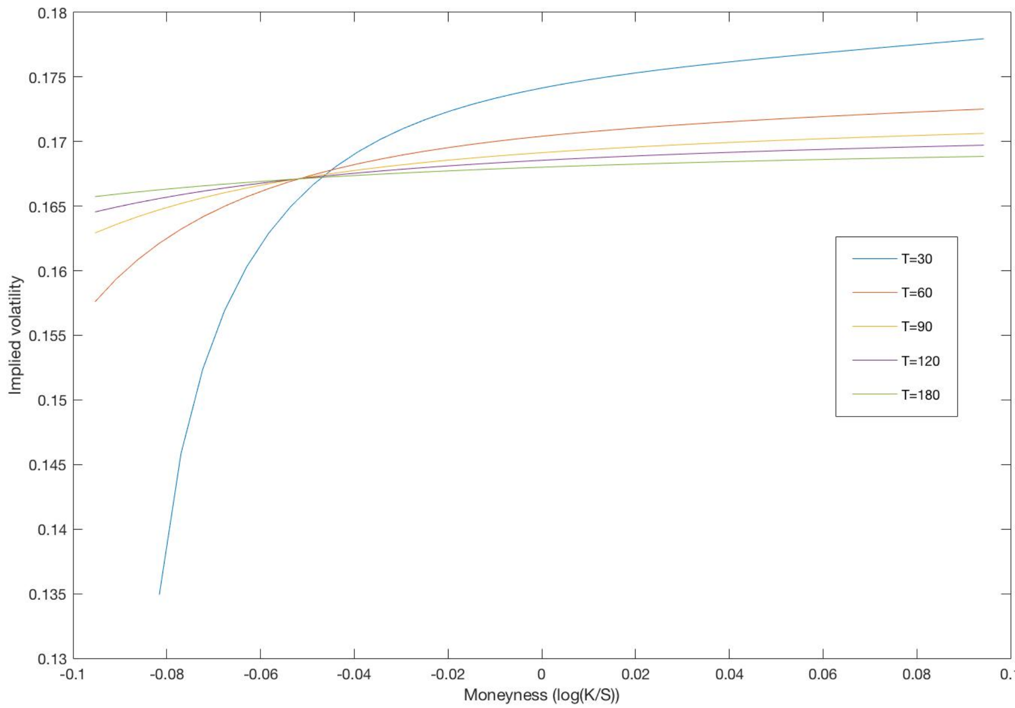
* *The option prices are computing with the latent stock price* $\tilde{S}_t = (1 - 0.001) \times S_t = (0.999) \times (100)$.
* *The Black-Scholes implied volatilities are computed using the observed stock price* $S_t = 100$.
* *The moneyness is* $log(K/S_t)$ *where* $K = [91, 92, \cdots, 100, 101, \cdots, 110]$.
* T *represents the time to maturity in days.*

Figure 10: Distortion of smile of the HN model



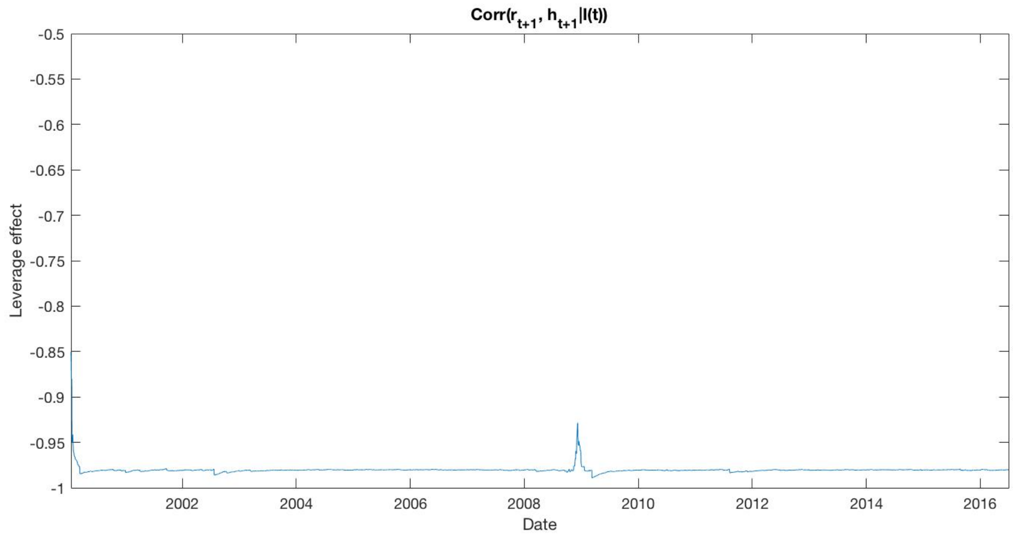
* *The option prices are computing with the latent stock price* $\tilde{S}_t = (1.001) \times S_t = (1.001) \times (100)$.
* *The Black-Scholes implied volatilities are computed using the observed stock price* $S_t = 100$.
* *The moneyness is* $log(K/S_t)$ *where* $K = [91, 92, \cdots, 100, 101, \cdots, 110]$.
* T *represents the time to maturity in days.*

Figure 11: Distortion of smile of the HN model



* *The option prices are computing with the latent stock price* $\tilde{S}_t = (1 - 0.001) \times S_t = (0.999) \times (100)$.
* *The Black-Scholes implied volatilities are computed using the observed stock price* $S_t = 100$.
* *The moneyness is* $log(K/S_t)$ *where* $K = [91, 92, \cdots, 100, 101, \cdots, 110]$.
* T *represents the time to maturity in days.*

Figure 12: HN model: Daily conditional correlation between returns and volatility



Appendix B

Proof of Proposition 2.2

$$
\begin{aligned}
C_t(K) \quad &= \quad E^*[BS_{(t)}\left(K, \tilde{S}_t, \sigma^2\left[J(t)\right]\right)|I(t)] \\
&\Longrightarrow \quad \frac{\partial C_t(K)}{\partial \lambda} = E^*\left[\Phi\left[d_{1,t}\left(K, \sigma^2[J(t)]\right)\right]\frac{\partial \tilde{S}_t}{\partial \lambda}|I(t)\right]
\end{aligned}
$$

with:

$$
\frac{\partial \tilde{S}_t}{\partial \lambda} = S_t \frac{\partial \xi_t}{\partial \lambda}
$$

and:

$$\begin{aligned}
\xi_t &= \frac{\exp\left\{\lambda' \tilde{Z}_t\right\}}{E^*[\exp\left\{\lambda' \tilde{Z}_t\right\}|I(t)]} \\
&\Longrightarrow \frac{\partial \xi_t}{\partial \lambda} = \xi_t \tilde{Z}_t - \exp\left\{\lambda' \tilde{Z}_t\right\} \frac{E^*[\tilde{Z}_t \exp\left\{\lambda' \tilde{Z}_t\right\}|I(t)]}{\left\{E^*[\exp\left\{\lambda' \tilde{Z}_t\right\}|I(t)]\right\}^2} \\
&= \xi_t \left[\tilde{Z}_t - \frac{E^*[\tilde{Z}_t \exp\left\{\lambda' \tilde{Z}_t\right\}|I(t)]}{\left\{E^*[\exp\left\{\lambda' \tilde{Z}_t\right\}|I(t)]\right\}}\right] \\
&\Longrightarrow \frac{\partial \xi_t}{\partial \lambda}_{|\lambda=0)} = \tilde{Z}_t - E^*[\tilde{Z}_t\,|I(t)] \\
&\Longrightarrow \frac{\partial C_t(K)}{\partial \lambda}_{|\lambda=0)} = S_t Cov^*[\tilde{Z}_t, \Phi\left(d_{1,t}\left(K, \sigma^2\left[J(t)\right]\right)\right)|I(t)]
\end{aligned}$$

Proof of Proposition 2.3

$$\begin{aligned}
d_{1,t}(K,V) &= \frac{1}{\sqrt{V}}\left[\log\left(S_t/K\right) + r_{f,t}\right] + \frac{\sqrt{V}}{2} \\
&\Longrightarrow \frac{\partial d_{1,t}(K,V)}{\partial V} = -\frac{1}{2V\sqrt{V}}\left[\log\left(S_t/K\right) + r_{f,t}\right] + \frac{1}{4\sqrt{V}}
\end{aligned}$$

Hence:

$$\frac{\partial d_{1,t}(K,V)}{\partial V} > 0 \Leftrightarrow x_t(K) = \log\left(K/S_t\right) - r_{f,t} > -\frac{V}{2}$$

Moreover:

$$\frac{\partial C_t(K)}{\partial \lambda_1}(\lambda = 0) = S_t Cov^*[\tilde{Z}_{1,t}, \Phi\left(d_{1,t}\left(K, \sigma^2\left[J(t)\right]\right)\right)|I(t)]$$

with, given $I(t)$, $\tilde{Z}_{1,t}$ increasing function of $\sigma^2\left[J(t)\right]$.Thus, at least for positive $x_t(K)$ we can conclude:

$$x_t(K) > 0 \Longrightarrow \frac{\partial C_t(K)}{\partial \lambda_1}(\lambda = 0) > 0$$

Proof of Proposition 3.1

The historical distribution defined by (3.2) is such that:

$$\begin{aligned}
& E[\exp(-u\sigma_{t+1}^2 - vr_{t+1})\,|I(t)] \\
= \;& E[\exp(-u\sigma_{t+1}^2)E[\exp(-vr_{t+1})\,|I^{\sigma}(t)]\,|I(t)] \\
= \;& E[\exp\left(-\left(u+\alpha(v)\right)\sigma_{t+1}^2\right)|I(t)]\exp\left(-\beta(v)\sigma_t^2 - \gamma(v)\right) \\
= \;& \exp\left(-a\left(u+\alpha(v)\right)\sigma_t^2 - b\left(u+\alpha(v)\right)\right)\exp\left(-\beta(v)\sigma_t^2 - \gamma(v)\right) \\
= \;& \exp\left(-l(u,v)\sigma_t^2 - g(u.v)\right)
\end{aligned}$$

with:

$$\begin{aligned}
l(u,v) &= a\left(u+\alpha(v)\right)+\beta(v) \\
g(u,v) &= b\left(u+\alpha(v)\right)+\gamma(v)
\end{aligned} \tag{7.1}$$

Note that it gives us the moment generating function of the conditional distribution of $\left(\sigma_{t+1}^2, r_{t+1}\right)$ given $I(t)$.

The SDF parameters m_0 and m_1 must then be computed in order to match the exogenously specified dynamics of interest rate, which is akin to the following restriction:

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

leading to pick m_0 and m_1 as functions $m_0\left(\varsigma\right)$ and $m_1\left(\varsigma\right)$ such that:

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

which is equivalent to:

$$m_0\left(\varsigma\right) = g\left(\varsigma\right), m_1\left(\varsigma\right) = l\left(\varsigma\right)$$

By uniqueness of a moment generating function, we can claim that the consistency identity (3.4) will be fulfilled if and only if it is fulfilled for all functions $H = H_{(u,v)}$ (for any pair (u,v) of real numbers) such that:

$$H_{(u,v)}\left(\sigma_{t+1}^2, r_{t+1}\right) = \exp\left(-u\sigma_{t+1}^2 - vr_{t+1}\right)$$

In other words, we want to show that for all (u,v) :

$$\begin{aligned}
& E^*[\exp\left(-u\sigma_{t+1}^2 - vr_{t+1}\right)|I(t)] \\
= \;& \exp(g\left(\varsigma\right)+l\left(\varsigma\right)\sigma_t^2)E[\exp\left(-(\varsigma_1+u)\sigma_{t+1}^2 - (\varsigma_2+v)r_{t+1}\right)|I(t)]
\end{aligned}$$

which is tantamount to:

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

with:

$$\begin{aligned} l^*(u,v) &= l(\varsigma_1 + u, \varsigma_2 + v) - l(\varsigma_1, \varsigma_2) \\ g^*(u,v) &= g(\varsigma_1 + u, \varsigma_2 + v) - g(\varsigma_1, \varsigma_2) \end{aligned} \tag{7.2}$$

By a computation similar to the one performed for proving (7.1), we should have:

$$\begin{aligned} l^*(u,v) &= a^*(u + \alpha^*(v)) + \beta^*(v) \\ g^*(u,v) &= b^*(u + \alpha^*(v)) + \gamma^*(v) \end{aligned} \tag{7.3}$$

Thus, it is sufficient to show that identities (7.2) and (7.3) are equivalent. We start from:

$$\begin{aligned} l^*(u,v) &= a^*(u + \alpha^*(v)) + \beta^*(v) \\ &= a(u + \alpha^*(v) + \varsigma_1 + \alpha(\varsigma_2)) - a(\varsigma_1 + \alpha(\varsigma_2)) \\ &\quad + \beta(\varsigma_2 + v) - \beta(\varsigma_2) \\ &= a(u + \varsigma_1 + \alpha(\varsigma_2 + v)) + \beta(\varsigma_2 + v) \\ &\quad - a(\varsigma_1 + \alpha(\varsigma_2)) - \beta(\varsigma_2) \\ &= l(\varsigma_1 + u, \varsigma_2 + v) - l(\varsigma_1, \varsigma_2) \end{aligned}$$

A similar computation would obviously give:

$$\begin{aligned} g^*(u,v) &= b^*(u + \alpha^*(v)) + \gamma^*(v) \\ &= g(\varsigma_1 + u, \varsigma_2 + v) - g(\varsigma_1, \varsigma_2) \end{aligned}$$

completing the proof of equivalence between (7.2) and (7.3).

Identities (3.5) show that the functions $\alpha^*(.), \beta^*(.)$ and $\gamma^*(.)$ are quadratic if and only if the functions $\alpha(.), \beta(.)$ and $\gamma(.)$ are quadratic. Therefore, risk neutral and historical conditional lognormality are equivalent.

Proof of Proposition 3.2

By definition of the moment generating function, we can compute the first two moments as follows:

$$\begin{aligned} E^*[r_{t+1}\,|I^\sigma(t)] &= \alpha^{*\prime}(0)\sigma_{t+1}^2 + \beta^{*\prime}(0)\sigma_t^2 + \gamma^{*\prime}(0) \\ Var^*[r_{t+1}\,|I^\sigma(t)] &= -\alpha^{*\prime\prime}(0)\sigma_{t+1}^2 - \beta^{*\prime\prime}(0)\sigma_t^2 - \gamma^{*\prime\prime}(0) \end{aligned}$$

We deduce from Proposition 3.1 that:

$$\begin{aligned} \alpha^{*\prime}(v) &= \alpha'(\varsigma_2 + v) \\ \alpha^{*\prime\prime}(v) &= \alpha''(\varsigma_2 + v) \end{aligned}$$

Hence, the coefficient of σ_{t+1}^2in $E^*[r_{t+1}\,|I^\sigma(t)] + \frac{1}{2}Var^*[r_{t+1}\,|I^\sigma(t)]$ is:

$$\alpha^{*\prime}(0) - \frac{\alpha^{*\prime\prime}(0)}{2} = \alpha'(\varsigma_2) - \frac{\alpha''(\varsigma_2)}{2}$$

However, since $\alpha(v)$ is a quadratic function with $\alpha(0) = 0$:

$$\alpha(v) = \alpha'(0)v + \frac{\alpha''(0)}{2}v^2$$

and thus:

$$\begin{aligned} \alpha(\varsigma_2) - \alpha(\varsigma_2 - 1) &= \alpha'(0) + \frac{\alpha''(0)}{2}\left[2\varsigma_2 - 1\right] \\ &= \alpha'(0) - \frac{\alpha''(0)}{2} + \alpha''(0)\varsigma_2 \end{aligned}$$

while:

$$\alpha'(\varsigma_2) - \frac{\alpha''(\varsigma_2)}{2} = [\alpha'(0) - \alpha''(0)\varsigma_2] - \frac{\alpha''(0)}{2}$$

which confirms that:

$$\alpha^{*\prime}(0) - \frac{\alpha^{*\prime\prime}(0)}{2} = \alpha(\varsigma_2) - \alpha(\varsigma_2 - 1)$$

Proof of Proposition 3.3

We have:

$$\begin{aligned} Cov[r_{t+1}, \tilde{\sigma}_{t+1}^2\,|I(t)] &= Cov\left\{E[r_{t+1}\,|I^\sigma(t)]\,, \tilde{\sigma}_{t+1}^2\,|I(t)]\right\} \\ &= Cov\left\{\psi\tilde{\sigma}_{t+1}^2, \tilde{\sigma}_{t+1}^2\,|I(t)]\right\} = \psi Var[\tilde{\sigma}_{t+1}^2\,|I(t)] \end{aligned}$$

Hence, the correlation coefficient:

$$Corr[r_{t+1}, \tilde{\sigma}^2_{t+1} \,|I(t)] = \psi \left\{ \frac{Var[\tilde{\sigma}^2_{t+1} \,|I(t)]}{Var[r_{t+1} \,|I(t)]} \right\}^{1/2}$$

We then deduce from the decomposition of variance (3.8) that:

$$Corr[r_{t+1}, \tilde{\sigma}^2_{t+1} \,|I(t)] = \psi \left\{ \psi^2 + \left[1 - \phi^2\right] \frac{E[\tilde{\sigma}^2_{t+1} \,|I(t)]}{Var[\tilde{\sigma}^2_{t+1} \,|I(t)]} \right\}^{-1/2}$$

Proof of Corollary 3.4

The first part is straightforward. Then:

$$\begin{aligned}
Corr[r_{t+1}, \tilde{\sigma}^2_{t+1} \,|I(t)] &= \phi \\
&\Longrightarrow \phi^2 \left\{\psi^2 + \left[1 - \phi^2\right] k^2\right\} = \psi^2 \\
&\Longleftrightarrow \left[1 - \phi^2\right] \psi^2 = \left[1 - \phi^2\right] k^2 \phi^2 \\
&\Longleftrightarrow |\psi| = k \,|\phi|
\end{aligned}$$

We clearly get the complete equivalence with:

$$\psi = k\phi$$

Proposition A.1:

The volatility factor $\sigma^2_{t,H}(N)$ given in Section 4.2 satisfies two $ARMA(1,1)$-type conditional moment restrictions:

$$E\left[\sigma^2_{t+H,H}(N) - \rho^H \sigma^2_{t,H}(N) - \omega(H)|\tilde{I}(t)\right] = 0,$$
$$E\left[\sigma^4_{t+H,H}(N) - \rho^{2H} \sigma^4_{t,H}(N) - a(H;N)\sigma^2_{t,H}(N) - b(H;N)|\tilde{I}(t)\right] = 0,$$

for deterministic coefficients $\omega(H)$, $a(H;N)$, and $b(H;N)$ are given in the proof below in (7.4), (7.15), and (7.16),and

$$\sigma^4_{t,H}(N) = \left[\frac{1}{HN} \sum_{n=1}^{HN} \sigma^2_{t+n/N}\right]^2,$$

for any $H, N = 1, 2, \cdots$, and information set $\tilde{I}(t) = \left\{\sigma^2_{t-kH,H}(N), k \geq 1\right\}$.

Proof of Proposition A.1

Everywhere below we use the following notation:

$$E_t[X] = E[X|I(t)],\ V_t[X] = Var[X|I(t)],$$

for any random variable X.

From the first moment of volatility (see Section 4.2) we have

$$E_t[\sigma_{t+2}^2] = \rho E_t[\sigma_{t+1}^2] + \omega,$$

which leads to:

$$E_t[\sigma_{t+2}^2] = \rho^2 E_t[\sigma_t^2] + \omega(1+\rho).$$

Then by iterating H times the same argument, we get

$$\begin{aligned} E_t[\sigma_{t+H}^2] &= \rho^H E_t[\sigma_t^2] + \omega(1+\rho+\cdots+\rho^{H-1}) \\ &= \rho^H E_t[\sigma_t^2] + \omega\frac{1-\rho^H}{1-\rho} \\ &= \rho^H E_t[\sigma_t^2] + \omega(H) \end{aligned} \tag{7.4}$$

where

$$\omega(H) = \omega\frac{1-\rho^H}{1-\rho}.$$

Then we see that for any real $h \geq 0$,

$$E_{t+h}[\sigma_{t+H+h}^2] = \rho^H E_{t+h}[\sigma_{t+h}^2] + \omega(H),$$

and, by the law of iterated expectations,

$$E_t[\sigma_{t+H+h}^2] = \rho^H E_t[\sigma_{t+h}^2] + \omega(H).$$

Adding all above equations for $h = \frac{1}{N}, \frac{2}{N}, \cdots, HN-1, HN$, and dividing by HN, we get

$$E_t\left[\sigma_{t+H,H}^2(N)\right] = \rho^H E_t\left[\sigma_{t,H}^2(N)\right] + \omega(H).$$

From the second moment of volatility (see Section 4.2) we have

$$E_t\left[\sigma_{t+1}^4\right] = \rho^2 E_t\left[\sigma_t^4\right] + aE_t\left[\sigma_t^2\right] + b,$$

where

$$
\begin{aligned}
a &= 2\rho\omega + \bar{\rho} \\
b &= \omega^2 + \bar{\omega},
\end{aligned}
\tag{7.5}
$$

and using the same argument,

$$E_{t+1}\left[\sigma_{t+2}^4\right] = \rho^2 E_{t+1}\left[\sigma_{t+1}^4\right] + aE_{t+1}\left[\sigma_{t+1}^2\right] + b.$$

Then, by the law of iterated expectations we get

$$E_t\left[\sigma_{t+2}^4\right] = \rho^4 E_t\left[\sigma_t^2\right] + a\left(E_t\left[\sigma_{t+1}^2\right] + \rho^2 E_t\left[\sigma_t^2\right]\right) + b(1+\rho^2)$$

and by iterating H times the same argument, we get

$$E_t\left[\sigma_{t+H}^4\right] = \rho^{2H} E_t\left[\sigma_t^4\right] + a\sum_{h=0}^{H-1}\rho^{2(H-1-h)}E_t\left[\sigma_{t+h}^2\right] + b\sum_{h=0}^{H-1}\rho^{2h}.$$

By applying (7.4) to the second term in the above equation, we get

$$
\begin{aligned}
\sum_{h=0}^{H-1}\rho^{2(H-1-h)}E_t\left[\sigma_{t+h}^2\right] &= \sum_{h=0}^{H-1}\rho^{2(H-1-h)}\left(\rho^h E_t[\sigma_t^2] + \omega(h)\right) \\
&= \rho^{H-1}\frac{1-\rho^H}{1-\rho}E_t[\sigma_t^2] + C_1,
\end{aligned}
$$

where

$$C_1 = \omega\sum_{h=0}^{H-1}\rho^{2(H-1-h)}\frac{1-\rho^H}{1-\rho} = \omega\frac{1-\rho^{H-1}(1-\rho^H)}{(1-\rho)(1-\rho)^2} = \omega(H)\frac{1-\rho^{H-1}}{1-\rho}. \tag{7.6}$$

Hence,

$$E_t\left[\sigma_{t+H}^4\right] = \rho^{2H}E_t\left[\sigma_t^4\right] + a\rho^{H-1}\frac{1-\rho^H}{1-\rho}E_t[\sigma_t^2] + C_2,$$

where

$$C_2 = aC_1 + b\sum_{h=0}^{H-1}\rho^{2h} = aC_1 + b\frac{1-\rho^{2H}}{1-\rho^2}. \tag{7.7}$$

Then now we see that for any $h \geq 0$,

$$E_t\left[\sigma_{t+H+h}^4\right] = \rho^{2H} E_t\left[\sigma_{t+h}^4\right] + a\rho^{H-1}\frac{1-\rho^H}{1-\rho}E_t[\sigma_{t+h}^2] + C_2,$$

which can be rewritten using lag operate L as

$$E_t\left[(1-\rho^{2H}L^H)\sigma_{t+H+h}^4\right] = a\rho^{H-1}\frac{1-\rho^H}{1-\rho}E_t[\sigma_{t+h}^2] + C_2. \tag{7.8}$$

Adding all of the above equations for $h = \frac{1}{N}, \frac{2}{N}, \cdots, HN-1, HN$, and dividing by HN, we get

$$E_t\left[(1-\rho^{2H}L^H)\frac{1}{HN}\sum_{n=0}^{HN}\sigma_{t+H+n/N}^4\right] = a\rho^{H-1}\frac{1-\rho^H}{1-\rho}E_t\left[\sigma_{t,H}^2(N)\right] + C_2. \tag{7.9}$$

Now we want to compute $E_t\left[\sigma_{t+H,H}^4(N)\right]$, where

$$\begin{aligned}
\sigma_{t,H}^4(N) &= \left[\frac{1}{HN}\sum_{n=1}^{HN}\sigma_{t+n/N}^2\right]^2 \\
&= \frac{1}{H^2N^2}\sum_{n=1}^{HN}\sigma_{t+n/N}^4 + \frac{2}{H^2N^2}\sum_{j=1}^{HN-1}\sum_{n=1}^{HN-j}\sigma_{t+n/N}^2\sigma_{t+(n+j)/N}^2.
\end{aligned}$$

This means, after multiplying by HN and shifting time by H, that

$$\frac{1}{HN}\sum_{n=1}^{HN}\sigma_{t+H+n/N}^4 = HN\sigma_{t+H,H}^4(N) - \frac{2}{HN}\sum_{j=1}^{HN-1}\sum_{n=1}^{HN-j}\sigma_{t+H+n/N}^2\sigma_{t+H+(n+j)/N}^2.$$

Making the corresponding substitution in (7.9) and dividing by HN, we can write

$$\begin{aligned}
E_t\left[\left(1-\rho^{2H}L^H\right)\sigma_{t+H,H}^4(N)\right] = E_t&\left[\frac{2}{H^2N^2}\sum_{j=1}^{HN-1}\sum_{n=1}^{HN-j}\left(1-\rho^{2H}L^H\right)\sigma_{t+H+n/N}^2\sigma_{t+H+(n+j)/N}^2\right] \\
&+ a_0(H;N)E_t\left[\sigma_{t,H}^2(N)\right] + \frac{1}{HN}C_2,
\end{aligned}$$

where

$$a_0(H;N) = \frac{1}{HN}a\rho^{H-1}\frac{1-\rho^H}{1-\rho}. \tag{7.10}$$

By the law of iterated expectations and (7.4), the expectation of cross-term is

$$\begin{aligned} E_t\left[\sigma^2_{t+H+n/N}\sigma^2_{t+H+(n+j)/N}\right] &= E_t\left[\sigma^2_{t+H+n/N}E_{t+H+n/N}\left[\sigma^2_{t+H+(n+j)/N}\right]\right] \\ &= \rho^{j/N}E_t\left[\sigma^4_{t+H+n/N}\right] + \omega(j/N)E_t\left[\sigma^2_{t+H+n/N}\right]. \end{aligned}$$

For $h = n/N$, the equation (7.8) is

$$E_t\left[\left(1-\rho^{2H}L^H\right)\sigma^4_{t+H+n/N}\right] = a]rho^{H-1}\frac{1-\rho^H}{1-\rho}E_t\left[\sigma^2_{t+n/N}\right] + C_2.$$

Applying (7.4) to $E_t\left[\sigma^2_{t+H+n/N}\right]$ gives us

$$\begin{aligned} E_t\left[\left(1-\rho^{2H}L^H\right)\sigma^2_{t+H+n/N}\right] &= E_t\left[\sigma^2_{t+H+n/N}\right] - \rho^{2H}E_t\left[\sigma^2_{t+n/N}\right] \\ &= \rho^H(1-\rho^H)E_t\left[\sigma^2_{t+n/N}\right] + \omega(H). \end{aligned}$$

Hence, the expectation of the cross-term multiplied by $\left(1-\rho^{2H}L^H\right)$ is

$$\begin{aligned} E_t\left[\left(1-\rho^{2H}L^H\right)\sigma^2_{t+H+n/N}\sigma^2_{t+H+(n+j)/N}\right] &= \rho^{j/N}E_t\left[\left(1-\rho^{2H}L^H\right)\sigma^4_{t+H+n/N}\right] \\ &+ \omega(j/N)E_t\left[\left(1-\rho^{2H}L^H\right)\sigma^2_{t+H+n/N}\right] \\ &= \left[a\rho^{j/N}\rho^{H-1}\frac{1-\rho^H}{1-\rho} + \omega(j/N)\rho^H(1-\rho^H)\right]E_t\left[\sigma^2_{t+n/N}\right] \\ &+ C_3(j), \end{aligned} \tag{7.11}$$

where

$$\begin{aligned} C_3(j) &= \rho^{j/N}C_2 + \omega(j/N)\omega(H) \\ &= \rho^{j/N}\left[a\omega(H)\frac{1-\rho^{H-1}}{1-\rho} + b\frac{1-\rho^{2H}}{1-\rho^2}\right] + \omega^2\frac{1-\rho^{j/N}}{1-\rho}\frac{1-\rho^H}{1-\rho}. \end{aligned} \tag{7.12}$$

Next, we need to express $E_t\left[\sigma^2_{t+n/N}\right]$ in terms of $E_t\left[\sigma^2_{t,H}(N)\right]$. For that purpose we apply (7.4) again:

$$E_t\left[\sigma^2_{t+n/N}\right] = \rho^{n/N}E_t\left[\sigma^2_t\right] + \omega(n/N).$$

We find that

$$\begin{aligned}
E_t\left[\sigma^2_{t,H}(N)\right] &= \frac{1}{HN}\sum_{n=1}^{HN} E_t\left[\sigma^2_{t+n/N}\right] \\
&= \frac{1}{HN}\sum_{n=1}^{HN}\left(\rho^{n/N}E_t\left[\sigma^2_t\right] + \omega(n/N)\right) \\
&= \frac{1}{HN}\sum_{n=1}^{HN}\rho^{n/N}E_t\left[\sigma^2_t\right] + \frac{1}{HN}\sum_{n=1}^{HN}\omega(n/N) \\
&= \frac{\rho^{1/N}}{HN}\frac{1-\rho^H}{1-\rho^{1/N}}E_t\left[\sigma^2_t\right] + C_4,
\end{aligned}$$

where

$$C_4 = \frac{1}{HN}\sum_{n=1}^{HN}\omega(n/N) = \frac{\omega}{HN}\frac{(1-\rho^{1/N}) - \rho^{1/N}(1-\rho^H)}{(1-\rho)(1-\rho^{1/N})}. \tag{7.13}$$

Solving for $E_t\left[\sigma^2_t\right]$ we have

$$E_t\left[\sigma^2_t\right] = \frac{HN}{\rho^{1/N}}\frac{1-\rho^{1/N}}{1-\rho^H}\left(E_t\left[\sigma^2_{t,H}(N)\right] - C_4\right).$$

and

$$\begin{aligned}
E_t\left[\sigma^2_{t+n/N}\right] &= \rho^{n/N}E_t\left[\sigma^2_t\right] + \omega(n/N) \\
&= HN\rho^{(n-1)/N}\frac{1-\rho^{1/N}}{1-\rho^H}\left(E_t\left[\sigma^2_{t,H}(N)\right] - C_4\right) + \omega(n/N).
\end{aligned}$$

Substituting this result to the expression (7.11) for the cross-terms, we obtain that

$$E_t\left[\left(1-\rho^{2H}L^H\right)\sigma^2_{t+H+n/N}\sigma^2_{t+H+(n+j)/N}\right]$$

is equal to

$$\begin{aligned}
&\left(a\rho^{j/N}\rho^{H-1}\frac{1-\rho^H}{1-\rho} + \omega(j/N)\rho^H(1-\rho^H)\right) \\
&\times\left(HN\rho^{(n-1)/N}\frac{1-\rho^{1/N}}{1-\rho^H}\left(E_t\left[\sigma^2_{t,H}(N)\right] - C_4\right) + \omega(n/N)\right) + C_3(j) \\
&= \left(a\rho^{j/N}\rho^{H-1}\frac{1-\rho^H}{1-\rho} + \omega(j/N)\rho^H(1-\rho^H)\right)\rho^{(n-1)/N}\frac{1-\rho^{1/N}}{1-\rho^H}HNE_t\left[\sigma^2_{t,H}(N)\right] + C_5(j,n),
\end{aligned}$$

where

$$C_5(j,n) = \left(a\rho^{j/N}\rho^{H-1}\frac{1-\rho^H}{1-\rho} + \omega(j/N)\rho^H(1-\rho^H)\right)$$
$$\times\left(\omega(n/N) - HN\rho^{(n-1)/N}\frac{1-\rho^{1/N}}{1-\rho^H}C_4\right) + C_3(j). \tag{7.14}$$

Collecting the terms, we find that the coefficients in the second part of Proposition A.1 are

$$a(H;N) = a_0(H;N)$$
$$+\frac{2}{HN}\frac{\rho^H}{1-\rho^H}(1-\rho^{1/N})\sum_{j=1}^{HN-1}\sum_{n=1}^{HN-j}\left(\rho^{j/N}\frac{a}{\rho}\frac{1-\rho^H}{1-\rho} + \omega(H)(1-\rho^{n/N})\right)\rho^{(n-1)/N}, \tag{7.15}$$

and

$$b(H;N) = \frac{2}{H^2N^2}\sum_{j=1}^{HN-1}\sum_{n=1}^{HN-j} C_5(j,n) \tag{7.16}$$

with $a_0(H;N)$ defined above in (7.10), $\omega(H)$ defined in (7.4), the coefficients C_1 through C_5 defined in (7.6), (7.7), (7.12), (7.13), and (7.14).

Proof of Proposition 4.2

We deduce from the first equations in Proposition A.1 and Lemma 4.1:

$$\lim_{H\to 0}\frac{1}{H}E_t\left[\sigma^2_{t+H,H}(N) - \sigma^2_{t,H}(N)\right] = \lim_{H\to 0}\frac{\rho^H-1}{H}E_t\left[\sigma^2_{t,H}(N)\right] + \lim_{H\to 0}\frac{1}{H}\omega(H)$$
$$= \log(\rho)\left(\sigma_t^2 - \frac{\omega}{1-\rho}\right)$$

since

$$\lim_{H\to 0}\frac{1}{H}\omega(H) = \lim_{H\to 0}\frac{\omega}{1-\rho}\frac{1-\rho^H}{H} = -\frac{\omega}{1-\rho}(1-\rho).$$

This proves the first part of Proposition 4.2.

For the second part, we see from the definition of variance and the equation (7.4) that the conditional moment restrictions given in Proposition A.1 are the same as

$$\begin{aligned}
V_t\left[\sigma^2_{t+H,H}(N)\right] &= -\left(E_t\left[\sigma^2_{t+H,H}(N)\right]\right)^2 + \rho^{2H}E_t\left[\sigma^4_{t,H}(N)\right] + a(H;N)E_t\left[\sigma^2_{t,H}(N)\right] + b(H;N)\\
&= -\left(\rho^H E_t\left[\sigma^2_{t,H}(N)\right] + \omega(H)\right)^2 + \rho^{2H}E_t\left[\sigma^4_{t,H}(N)\right]\\
&+ a(H;N)E_t\left[\sigma^2_{t,H}(N)\right] + b(H;N)\\
&= \rho^{2H}V_t\left[\sigma^2_{t,H}(N)\right] + \left(a(H;N) - 2\rho^H\omega(H)\right)E_t\left[\sigma^2_{t,H}(N)\right] + \left(b(H;N) - (\omega(H))^2\right).
\end{aligned}$$

Next, divide this expression on both sides by H and take the limit ($N \to \infty$ implicitly since $\sigma^2_{t,H}(N)$ is only defined for $H \geq 1/N$):

$$\begin{aligned}
\lim_{H\to\infty}\frac{1}{H}V_t\left[\sigma^2_{t+H,H}(N)\right] = \lim_{H\to\infty}\frac{\rho^{2H}}{H}V_t\left[\sigma^2_{t,H}(N)\right] + \lim_{H\to\infty}\frac{a(H;N)-2\rho^H\omega(H)}{H}E_t\left[\sigma^2_{t,H}(N)\right]\\
+ \lim_{H\to\infty}\frac{b(H;N)-(\omega(H))^2}{H}.
\end{aligned}$$

Using the second part in Lemma 4.1, we get

$$\begin{aligned}
\lim_{H\to\infty}\frac{1}{H}V_t\left[\sigma^2_{t+H,H}(N)\right] - \lim_{H\to\infty}\frac{\rho^{2H}}{H}V_t\left[\sigma^2_{t,H}(N)\right] &= \lim_{H\to\infty}\frac{2}{H}V_t\left[\sigma^2_{t,H}(N)\right] - \lim_{H\to\infty}\frac{\rho^{2H}}{H}V_t\left[\sigma^2_{t,H}(N)\right]\\
&= \lim_{H\to\infty}\frac{1}{H}V_t\left[\sigma^2_{t,H}(N)\right]\left(2-\rho^{2H}\right)\\
&= \lim_{H\to\infty}\frac{1}{H}V_t\left[\sigma^2_{t,H}(N)\right],
\end{aligned}$$

and deduce that the above limit expression can be rewritten as

$$\lim_{H\to\infty}\frac{1}{H}V_t\left[\sigma^2_{t,H}(N)\right] = \lim_{H\to\infty,N\to\infty}\frac{a(H;N)-2\rho^H\omega(H)}{H}E_t\left[\sigma^2_{t,H}(N)\right] + \lim_{H\to\infty,N\to\infty}\frac{b(H;N)-(\omega(H))^2}{H}.$$

Simplifying $a(H;N)$: Before taking the limit with respect to $N \to \infty$, we need to simplify $a(H;N)$ by getting rid of summations in

$$\begin{aligned}
a(H;N) - a_0(H;N) = &\frac{2}{HN}\frac{\rho^H}{1-\rho^H}\left(1-\rho^{1/N}\right)\\
&\times\sum_{j=1}^{HN-1}\left[\left(\rho^{j/N}\frac{a}{\rho}\frac{1-\rho^H}{1-\rho} + \omega(H)\right)\sum_{n=1}^{HN-j}\rho^{(n-1)N} - \rho^{1/N}\omega(H)\sum_{n=1}^{HN-j}\rho^{2(n-1)/N}\right],
\end{aligned}$$

with $a_0(H;N)$ and a defined in (7.10) and (7.5) . Here the inner summations are reduced to

$$\sum_{n=1}^{HN-j}\rho^{(n-1)/N} = \sum_{n=0}^{HN-j-1}\rho^{n/N} = \frac{1-\rho^{H-j/N}}{1-\rho^{1/N}},$$

and

$$\sum_{n=1}^{HN-j} \rho^{2(n-1)/N} = \frac{1-\rho^{2H-2j/N}}{1-\rho^{2/N}}.$$

So the coefficient becomes

$$\begin{aligned} a(H;N) - a_0(H;N) = &\frac{2}{HN}\frac{\rho^H}{1-\rho^H}\left(1-\rho^{1/N}\right) \\ &\times \sum_{j=1}^{HN-1}\left[\left(\rho^{j/N}\frac{a}{\rho}\frac{1-\rho^H}{1-\rho}+\omega(H)\right)\frac{1-\rho^{H-j/N}}{1-\rho^{1/N}} - \rho^{1/N}\omega(H)\frac{1-\rho^{2H-2j/N}}{1-\rho^{2/N}}\right], \end{aligned}$$

or

$$\begin{aligned} a(H;N) - a_0(H;N) = &\frac{2}{HN}\frac{\rho^H}{1-\rho^H}\left(\frac{a}{\rho}\frac{1-\rho^H}{1-\rho}\sum_{j=1}^{HN-1}(\rho^{j/N}-\rho^H) + \omega(H)\sum_{j=1}^{HN-1}(1-\rho^{H-j/N})\right) \\ &- \frac{2}{HN}\frac{\rho^H}{1-\rho^H}\frac{\rho^{1/N}}{1+\rho^{1/N}}\omega(H)\sum_{j=1}^{HN-1}\left(1-\rho^{2H-2j/N}\right). \end{aligned}$$

In this expression, we have three summations over j:

$$\sum_{j=1}^{HN-1}\rho^{H-j/N} = \sum_{j=1}^{HN-1}\rho^{j/N} = \frac{\rho^{1/N}-\rho^H}{1-\rho^{1/N}}, \text{ and } \sum_{j=1}^{HN-1}\rho^{2H-2j/N} = \frac{\rho^{2/N}-\rho^{2H}}{1-\rho^{2/N}}$$

Substituting these we have

$$\begin{aligned} a(H;N) - a_0(H;N) = &-\frac{2}{HN}\frac{a}{\rho}\frac{\rho^H}{1-\rho}\left(\rho^H(HN-1) - \frac{\rho^{1/N}-\rho^H}{1-\rho^{1/N}}\right) \\ &+ \frac{2}{HN}\frac{\rho^H}{1-\rho^H}\omega(H)\left((HN-1) - \frac{\rho^{1/N}-\rho^H}{1-\rho^{1/N}}\right) \\ &- \frac{2}{HN}\frac{\rho^H}{1-\rho^H}\frac{\rho^{1/N}}{1+\rho^{1/N}}\omega(H)\left((HN-1) - \frac{\rho^{2/N}-\rho^{2H}}{1-\rho^{2/N}}\right). \end{aligned}$$

Taking the limit with $N \to \infty$: Taking the limit with respect to $N \to \infty$, the coefficient becomes

$$\begin{aligned} \lim_{N\to\infty} a(H;N) = &-\frac{2}{H}\frac{a}{\rho}\frac{\rho^H}{1-\rho}\left(\frac{1-\rho^H}{log(\rho)} + \rho^H H\right) \\ &+ 2\frac{\rho^H}{1-\rho^H}\frac{\omega(H)}{H}\left(H + \frac{1-\rho^H}{log(\rho)}\right) \\ &- \frac{\rho^H}{1-\rho^H}\frac{\omega(H)}{H}\left(H + \frac{1-\rho^{2H}}{log(\rho^2)}\right). \end{aligned} \tag{7.17}$$

while

$$\lim_{N\to\infty} a_0(H;N) = \lim_{N\to\infty} \frac{1}{HN} a\rho^{H-1} \frac{1-\rho^H}{1-\rho} = 0.$$

Now divide (7.17) by H:

$$\begin{aligned}\frac{1}{H} \lim_{N\to\infty} a(H;N) = & -2\frac{a}{\rho}\frac{\rho^H}{1-\rho}\frac{1}{H}\left(\frac{1-\rho^H}{Hlog(\rho)}+\rho^H\right)\\ & +2\frac{\rho^H}{1-\rho^H}\frac{\omega(H)}{H}\left(1+\frac{1-\rho^H}{Hlog(\rho)}\right)\\ & -\frac{\rho^H}{1-\rho^H}\frac{\omega(H)}{H}\left(1+\frac{1-\rho^{2H}}{Hlog(\rho^2)}\right).\end{aligned}$$

Series expansion of this expression around $H = 0$ gives the following result:

$$\frac{1}{H} \lim_{N\to\infty} a(H;N) = -\frac{a}{\rho}\frac{log(\rho)}{1-\rho} + O(H).$$

Hence,

$$\lim_{H\to 0, N\to\infty} \frac{a(H;N)}{H} = -\frac{a}{\rho}\frac{log(\rho)}{1-\rho}.$$

Taking the limit of the constant we obtain[12]

$$\lim_{H\to 0, N\to\infty} \frac{b(H;N)}{H} = \left(\frac{a}{\rho}-\frac{2b}{\omega}\right)\frac{\omega log(\rho)}{1-\rho^2}.$$

with b defined in (7.5).

Finally,

$$\lim_{H\to 0} \frac{(\omega(H))^2}{H} = \lim_{H\to 0}\left(\frac{\omega(H)}{H}\right)^2 H = 0.$$

This result concludes the proof and shows explicitly that

$$\begin{aligned}\lim_{H\to 0, N\to\infty} \frac{1}{H} V_t\left[\sigma^2_{t,H}(N)\right] = & \left(-\frac{a}{\rho}\frac{log(\rho)}{1-\rho}+2\omega\frac{log(\rho)}{1-\rho}\right)\sigma_t^2 + \left(\frac{a}{\rho}-\frac{2b}{\omega}\right)\frac{\omega log(\rho)}{1-\rho^2}\\ & -\frac{log(\rho)}{1-\rho}\left[\left(\frac{a}{\rho}-2\omega\right)\sigma_t^2+\left(\frac{a}{\rho}-\frac{2b}{\omega}\right)\frac{\omega}{1+\rho}\right].\end{aligned}$$

[12]The analytical expression for $b(H;N)$ after taking all summations is several pages long. Taking the limit of this expression by hand does not seem feasible. These operations were performed in Mathematica software and available upon request.

In case of affine first two moments as in (4.4), we have

$$a = 2\rho\omega + \bar{\rho}, \ b = \omega^2 + \bar{\omega}.$$

Hence, the limit becomes

$$\lim_{H\to 0, N\to\infty} \frac{1}{H} V_t\left[\sigma^2_{t,H}(N)\right] = -\frac{\bar{\rho}}{\rho}\frac{log(\rho)}{1-\rho}\left(\sigma_t^2 + \frac{\omega - 2\bar{\omega}(\rho/\bar{\rho})}{1+\rho}\right).$$

For the ARG(1) case, where $\omega = 2\bar{\omega}(\rho/\bar{\rho})$, the same limit becomes

$$\lim_{H\to 0, N\to\infty} \frac{1}{H} V_t\left[\sigma^2_{t,H}(N)\right] = -\frac{\bar{\rho}}{\rho}\frac{log(\rho)}{1-\rho}\sigma_t^2,$$

as expected from a particular case of Gourieroux and Jasiak (2006, p.137).

Proof of Proposition 5.1

We look for values of parameters $\theta_\sigma = (\rho, \delta, c)'$ solution of

$$E\left\{\exp\left[-u\left(\sigma_t^2 + \sigma_{t+1}^2\right)\right]\right\} = E\left\{\exp\left(-u\sigma_t^2\right)\Psi_{t,\theta_\sigma}(u)\right\}$$

for several possible values of the complex number u. From the moment generating function of $\left(\sigma_t^2 + \sigma_{t+1}^2\right)$ and the marginal distribution of σ_t^2 which is , this equation can be rewritten:

$$\begin{aligned} & \left(1+c^0u\right)^{-\delta^0}\left(1+\frac{c^0}{1-\rho^0}\left[\frac{\rho^0 u}{1+c^0u}+u\right]\right)^{-\delta^0} \\ = \ & \left\{(1+cu)^{\delta/\delta^0}\right\}^{-\delta^0}\left(1+\frac{c^0}{1-\rho^0}\left[\frac{\rho u}{1+cu}+u\right]\right)^{-\delta^0} \end{aligned}$$

where $\left(\rho^0, \delta^0, c^0\right)$ stands for the true unknown value of (ρ, δ, c). In ohher words, we must have $A(u) = B(u)$ with:

$$\begin{aligned} A(u) &= (1+cu)^{\delta/\delta^0}\left(1+\frac{c^0}{1-\rho^0}\left[\frac{\rho u}{1+cu}+u\right]\right) \\ B(u) &= \left(1+c^0u\right)\left(1+\frac{c^0}{1-\rho^0}\left[\frac{\rho^0 u}{1+c^0u}+u\right]\right) \\ &= 1+c^0u+\frac{c^0}{1-\rho^0}\left[\rho^0u+u+c^0u^2\right] \\ &= 1+\frac{1}{1-\rho^0}\left[2c^0u+\left(c^0\right)^2u^2\right]. \end{aligned}$$

In particular:

$$\begin{aligned}
A'(0) &= B'(0) \\
&\iff \frac{2c^0}{1-\rho^0} = c\frac{\delta}{\delta^0} + \frac{c^0}{1-\rho^0}(1+\rho) \\
&\iff \frac{c^0}{1-\rho^0} = \frac{c}{1-\rho}.\frac{\delta}{\delta^0}.
\end{aligned}$$

By plugging in, we can rewrite:

$$\begin{aligned}
A(u) &= (1+cu)^{\delta/\delta^0}\left(1+\frac{c}{1-\rho}.\frac{\delta}{\delta^0}\left[\frac{\rho u}{1+cu}+u\right]\right) \\
&= (1+cu)^{x-1}\left(1+cu+\frac{cx}{1-\rho}\left[\rho u+u+cu^2\right]\right) \\
&= (1+cu)^{x-1}\tilde{A}(u)
\end{aligned}$$

where $x = \delta/\delta^0$.

Note that $\tilde{A}(u)$ and $B(u)$ are polynomial of degree two with:

$$\begin{aligned}
\tilde{A}(-1/c) &= -\frac{\rho x}{1-\rho} \neq 0 \\
B(-1/c) &= -\frac{\rho^0}{1-\rho^0} \neq 0.
\end{aligned}$$

Therefore:

$$\begin{aligned}
A(u) &= (1+cu)^{x-1}\tilde{A}(u) = B(u), \forall u \\
&\implies x = 1 \implies \tilde{A}(u) = B(u), \forall u \\
&\implies \rho = \rho^0 \implies c = c^0 \\
&\implies (\rho,\delta,c) = \left(\rho^0,\delta^0,c^0\right).
\end{aligned}$$