The leverage effect puzzle revisited: identification in discrete time

Hyojin Han* Stanislav Khrapov† Eric Renault‡

July 29, 2019

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. The discrete time approach allows to make more transparent the role of various parameters: leverage versus volatility feedback effect, connection with daily realized volatility, impact of leverage on the volatility smile, etc. Even more importantly it sheds some new light on the identification of leverage effect and of the various risk premium parameters through link functions in closed form. 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. However, the link functions are almost flat if the leverage effect is close to zero, making estimation of the volatility risk price difficult and paving the way for identification robust inference.

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
†DB Schenker, khrapovs@gmail.com
‡Department of Economics, University of Warwick, Eric.Renault@warwick.ac.uk

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." Moreover, 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 the presence of 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. Although we only consider one latent volatility factor, we may extend our model to accommodate different factors. In that sense, 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. 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.

The contribution of the paper is threefold with two general results completed by an empirical illustration with a more restrictive econometric specification.

(i) The first general result completes a former analysis by Renault and Touzi (1996) (see also Renault (1997) for a more transparent and general proof) that had established that the absence of leverage was a sufficient condition for a symmetric volatility smile. We now prove that conversely the presence of leverage will necessarily manifests itself by a smirk. In addition, we precisely describe the shape of this smirk by showing that the volatility smile will be less steeply increasing (or even possibly decreasing) on the out-of-the-money side. For this first result, the only object of interest is the risk neutral distribution.

(ii) The second general result states that the price of volatility risk can be identified by using only observations on the underlying asset return if and only if there is a significant leverage effect. In other words, by contrast with a common belief, option price data may not be necessary to identify the price of volatility risk.

(iii) Besides the exogeneity of the volatility factor and the affine structure of the conditional distributions (with conditional normality of returns given the path of the latent volatility factor), these general results do not take any specific parametric model. Our main focus of interest on leverage effect and the need to identify it from return data eventually lead us to introduce some constraints between the parameters of the volatility dynamics. We end up with an ARG(1)-Normal (Auto-Regressive-Gamma) model that is an extension with leverage of the one

of Gourieroux and Jasiak (2006). The characterization of this model through the conditional moment generating function provides conditional moment restrictions for statistical inference as a discrete time version of the work of Pan (2002). Even though it is beyond the scope of this paper, it is worth noting that this also paves the way for an extension of the semi-parametric approach of Gagliardini, Gourieroux, and Renault (2011).

The rest of the paper is organized as follows. Section 2 proposes a general characterization of the shape of volatility smiles through a risk neutral distribution seen as a 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 shifts the interest towards an historical distribution coherent with the risk neutral one of Section 2 through an exponentially affine pricing kernel. It then proves that the price of volatility risk can be identified from the historical statistical characteristics of the underlying asset return if an only if there is a significant leverage effect. Section 4 proposes a fully parametric model of returns and volatility. This ARG(1)-Normal model is shown (in Appendix C) to be a discrete time version of Heston (1993) model. For the purpose of statistical identification of leverage effect, this parametric model is a highly constrained version of the model of Section 2 and 3. These constraints could be relaxed at the cost of a less straightforward interpretation of the parameterization of leverage effect. In Section 5, we show how intraday data on realized variance can be used for user-friendly GMM inference and devise a general two-step GMM estimation procedure based on the conditional moment generating function. We provide an empirical illustration on daily log returns and realized volatilities of the S&P500 over 16 years starting in January 2000. 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 6 concludes. The mathematical proofs of theoretical results and figures are relegated to Appendices A and B. We also show in Appendix C that our model is a discrete time version of Heston (1993) model and discuss the choice of instruments for GMM in Appendix D.

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 patterns 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 $\pounds^*(r_{t+1}\,|J(t))$ at stake for the purpose of option pricing. Therefore, with obvious notations, we maintain throughout the following assumption:

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

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 setup). We actually show that the standard continuous time option pricing model with stochastic volatility Heston (1993) can be seen as a continuous time limit of our setup (see Section 4.3 and appendix C).

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 so that:

$$I(t) \subsetneqq J(t)$$

where $I(t)$ stands for the information that is available for investor at time t. Note that in popular option pricing models, the complete information set $J(t)$ 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 current 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 part of the information set $J(t)$ that is not available to investor at time t.

In line with the stochastic volatility example, a maintained assumption will be:

Assumption: (Exogeneity of volatility) The risk-neutral probability distribution of $J(t)$ given $I(t)$ does not depend on past and current returns.

This exogeneity of the volatility factor is fulfilled in standard stochastic volatility models while it is violated in GARCH-type option pricing models. 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 $J(t)$ was 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:

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

Note that then:

$$\begin{aligned} S_t &= e^{-r_{f,t}} E^*[S_{t+1}\,|I(t)] = E^*[\tilde{S}_t\,|I(t)] \\ &\Longrightarrow E[\exp\left(\mu^*\left[J(t)\right] + \frac{\sigma^{*2}\left[J(t)\right]}{2}\right)|I(t)] = 1. \end{aligned} \tag{2.1}$$

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 at time t would be conformable to the Black and Scholes option pricing formula 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

$$C_t(K) = E^*[BS_{(t)}\left(K, \tilde{S}_t, \sigma^{*2}\left[J(t)\right]\right)|I(t)], \tag{2.2}$$

where $BS_{(t)}\left(K, \tilde{S}_t, \sigma^{*2}\left[J(t)\right]\right)$ is the Black and Scholes option pricing formula for an European call with strike price K.

It is worth noting that the conditional expectation in (2.2) 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 is deduced from the above assumption of exogeneity of volatility and 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 for a symmetric volatility smile but also necessary in very general circumstances. To see that, it is worth contemplating a ratio $\left(\tilde{S}_t/S_t\right)$ that is log-linear w.r.t. some strictly increasing function of $\sigma^{*2}\left[J(t)\right]$, so that we can prove the following result:

Proposition 2.2.:

Assume that we have a parametric model

$$\mu^* [J(t)] + \frac{1}{2}\sigma^{*2} [J(t)] = A [\lambda, J(t)], \lambda \in \mathbb{R}^p$$

with:

$$\begin{aligned} A [\lambda, J(t)] &= \lambda_1 \tilde{Z}(t) + \tilde{A} \left[\lambda_{)1(}, J(t)\right] \\ \lambda &= \left(\lambda_1, \lambda'_{)1(}\right)', \tilde{A} [0, J(t)] = 0 \end{aligned}$$

and:

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

where the deterministic function $\sigma^2 \longrightarrow h \left[\sigma^2, I(t)\right]$ is strictly increasing and:

$$Var[\sigma^{*2} [J(t)] | I(t)] > 0.$$

Then for any out-of-the money option:

$$\log \left[\frac{K}{S(t)}\right] > r_{f,t} \Longrightarrow \frac{\partial C_t(K)}{\partial \lambda_1} (\lambda = 0) > 0.$$

In other words, for all out-of-the money options, the option price is an increasing function of the slope coefficient λ_1. Moreover, the proof of Proposition 2.2. 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-of-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 occurrence of leverage effect is identified by the occurrence 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-of-the-money side. This is in accordance with some well-documented stylized facts (see also our empirical section below).

3 Statistical identification of risk premium parameters

We have proved in Section 2 that a non-flat volatility smile takes a latent state variable and that the skew of this smile identifies the presence of leverage effect in a general conditionally log-normal risk-neutral model. The purpose of this Section 3 is to further characterize the information content of a skewed volatility smile. We will argue that by contrast with a common belief, risk premium parameters on latent volatility risk may be identified without resorting to option prices data, that is using only times series of the underlying asset returns. It is precisely when leverage effect is present, that is when the volatility smile is skewed, this identification is possible.

3.1 An exponentially affine pricing kernel

Identification of risk-neutral parameters based only on historical data on underlying asset returns takes a bridge between risk-neutral and historical parameters, that is a pricing kernel. This pricing kernel will define the compensation for the different sources of risk. To keep it simple, we assume that the information $I(t)$ available to all investors at time t consists of only the past and present values $r(\tau), \tau \leq t$, of underlying asset return, while only one latent process characterizes the difference between the investor information and the complete, partly latent, information set $J(t)$. Following the discussion in Section 2, it is natural to denote this additional stochastic process σ_{t+1}^2 and dub it a volatility factor, even though its exact connection with the stochastic volatility of asset return will be characterized later. For sake of clarity, we will use the following notations:

$$J(t) = I^{\sigma}(t) = I(t) \vee \left\{\sigma_{t+1}^2\right\}.$$

The rationale for these notations is twofold: First, since the volatility factor is still latent at time t for investors, but eventually observed by them at time $(t+1)$, it is natural to index by $(t+1)$ the random element σ_{t+1}^2 that makes the difference between $J(t)$ and $I(t)$. Second, to stress that this difference is encapsulated in the volatility factor, we replace the notation $J(t)$ for the complete information set by $I^{\sigma}(t)$.

We are then led to define a stochastic discount factor $M_{t+1}(\varsigma)$ that depends on two risk premium parameters ς_1 and ς_2 that define respectively the risk compensation for the random elements σ_{t+1}^2 and

r_{t+1}. As commonly done in the literature, we will then consider an exponentially affine SDF:

$$\begin{aligned} M_{t+1}(\varsigma) &= \exp(-r_{f,t}) M_{0,t}(\varsigma) \exp\left[-\varsigma_1 \sigma_{t+1}^2 - \varsigma_2 r_{t+1}\right] \\ \varsigma &= (\varsigma_1, \varsigma_2), M_{0,t}(\varsigma) \in I(t). \end{aligned}$$

The pricing of a riskless payoff imposes the no-arbitrage restriction:

$$E[\exp\left(-\varsigma_1 \sigma_{t+1}^2 - \varsigma_2 r_{t+1}\right) | I(t)] = [M_{0,t}(\varsigma)]^{-1}. \tag{3.1}$$

This restriction shows that the specification of such a pricing kernel for any possible value of the risk premium parameters $(\varsigma_1, \varsigma_2)$ amounts to specify the joint historical conditional probability distribution (through the conditional Laplace transform) of $\left(\sigma_{t+1}^2, r_{t+1}\right)$ given $I(t)$. To complete this specification, we must specify this function by taking into account the following remarks:

First, if we maintain the assumption of exogeneity of volatility, not only for the risk-neutral distribution but also for the historical distribution, we see that $[M_{0,t}(\varsigma)]$ should depend on $I(t)$ only through the past and current values $\sigma_\tau^2, \tau \leq t$, of the volatility factor.

Second, we remain true to the common practice to assume that the state variable process is Markov of order one. Otherwise, one would consider a higher dimensional state variable process. Then, $[M_{0,t}(\varsigma)]$ should depend on $I(t)$ only through the value of σ_t^2.

Third, if we remain true to the exponential point of view, we will specify $[M_{0,t}(\varsigma)]$ as:

$$M_{0,t}(\varsigma) = \exp\left[l(\varsigma)\sigma_t^2 + g(\varsigma)\right].$$

We end up with the following exponentially affine SDF:

$$M_{t+1}(\varsigma) = \exp(-r_{f,t}) \exp\left[l(\varsigma)\sigma_t^2 + g(\varsigma)\right] \exp\left[-\varsigma_1 \sigma_{t+1}^2 - \varsigma_2 r_{t+1}\right], \tag{3.2}$$

and the CAR historical model:

$$E[\exp\left(-u\sigma_{t+1}^2 - v r_{t+1}\right) | I(t)] = \exp\left[-l(u,v)\sigma_t^2 - g(u,v)\right]. \tag{3.3}$$

3.2 The CAR risk-neutral model

It is worth stressing that the simple fact that we consider an exponential affine pricing kernel with one state variable (in addition to the asset return), jointly with the assumption of exogeneity and Markovianity of the state variable process, has led us naturally to a CAR historical model. We are going to show in this subsection that it also necessarily leads to a CAR risk-neutral model that must stay in the same model class. When going from the physical to the risk neutral measure

(and vice versa) one is constrained to remain in the same parametric family, irrespective of our preferred parametric model. The general message is that our structural point of view limits the allowed flexibility for an empirical specification. For instance, even though 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, so that their empirical specification is at odds with our structural specification.

The necessary coherency condition between the risk-neutral and the historical model is described by the following result.

Proposition 3.1.:

(i) The risk-neutral joint distribution of $\left(\sigma_{t+1}^2, r_{t+1}\right)$ is given by the CAR model:

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

where:

$$l^*\left(u,v\right) = l\left(u+\varsigma_1, v+\varsigma_2\right) - l\left(\varsigma_1,\varsigma_2\right).$$

(ii) If the conditional mean $\mu^*\left[I^\sigma(t)\right]$ and the conditional variance $\sigma^{*2}\left[I^\sigma(t)\right]$ are both affine functions of σ_{t+1}^2 and σ_t^2, there exist quadratic functions $\alpha^*(\cdot), \beta^*(\cdot)$ and $\gamma^*(\cdot)$ such that:

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

with:

$$a^*(u) = l^*\left(u,0\right); b^*(u) = g^*\left(u,0\right).$$

Then the joint distribution of $\left(\sigma_{t+1}^2, r_{t+1}\right)$ can be factorized along the two following univariate CAR models:

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

Several remarks are in order:

First, the assumption that both conditional mean $\mu^*\left[I^\sigma(t)\right]$ and conditional variance $\sigma^{*2}\left[I^\sigma(t)\right]$ are affine functions of σ_{t+1}^2 and σ_t^2 is just a reinforcement of the assumption already maintained in Proposition 2.2. about the combined quantity $\left[\mu^*\left[I^\sigma(t)\right] + \frac{1}{2}\sigma^{*2}\left[I^\sigma(t)\right]\right]$,

Second, the marginal CAR model for r_{t+1} given $I^{\sigma}(t)$ implies:

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

Third, since $\alpha^{*}(\cdot),\beta^{*}(\cdot)$, and $\gamma^{*}(\cdot)$ are quadratic functions, they all fulfil the following identity:

$$\varpi(-1)=\frac{\varpi''(0)}{2}-\varpi'(0),\ \varpi(.)\in\left\{\alpha^{*}(\cdot),\beta^{*}(\cdot),\gamma^{*}(\cdot)\right\}.$$

The following corollary is then straightforward.

Corollary 3.2.:

The condition of symmetry of the volatility smile (dubbed "absence of leverage effect") can be written equivalently as:

(i) $\mu^{*}\left[I^{\sigma}(t)\right]+\frac{\sigma^{*2}[I^{\sigma}(t)]}{2}$ belongs to the information set $I(t)$

(ii) $E^{*}[\exp\left(r_{t+1}\right)|I^{\sigma}(t)]$ does not depend on σ_{t+1}^{2}

(iii) We have:

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

Note that Corollary 3.2. allows us to reconcile two common views about what characterizes the absence of leverage effect:

On the one hand, it is tantamount to symmetry of the volatility smile,

On the other hand, it means that the risk-neutral optimal forecast of the next asset return:

$$\exp\left(r_{t+1}\right)=\exp(-r_{f,t})\frac{S_{t+1}}{S_{t}}$$

cannot be improved by the knowledge of the contemporaneous volatility factor σ_{t+1}. By contrast, leverage effect would be encapsulated in a positive coefficient $\alpha^{*}\left(-1\right)$, meaning that the asset return r_{t+1} and the volatility factor σ_{t+1}^{2} are negatively correlated given $I(t)$. Note, however, that the risk-neutral distribution must, by definition fulfil the following price identity (risk-neutral pricing of the underlying asset):

$$E^{*}[\exp\left(r_{t+1}\right)|I(t)]=1.$$

From the above formulas we deduce the following.

Corollary 3.3.:

The risk-neutral pricing of the underlying asset is tantamount to the following constraints on the

marginal risk-neutral distributions:

$$\begin{aligned} a^* \left[\alpha^*(-1)\right] + \beta^*(-1) &= 0, \\ b^* \left[\alpha^*(-1)\right] + \gamma^*(-1) &= 0. \end{aligned}$$

It is worth noticing that these constraints involve the volatility dynamics, that are the functions $a^*(\cdot)$ and $b^*(\cdot)$ only if there is leverage effect ($\alpha^*(-1) \neq 0$). Otherwise, these constraints are tantamount to the following equalities:

$$\alpha^*(-1) = \beta^*(-1) = \gamma^*(-1) = 0.$$

This remark will have important implications for statistical identification of risk premium parameters from data on the underlying asset price.

3.3 Identification of risk premium parameters

Identification of risk premium parameters must be based on the CAR historical model (3.3). It is worth noting that this model is endowed with a structure similar to the structure of the risk-neutral model emphasized in Proposition 3.1.(ii). More precisely, we can show:

Proposition 3.2.:

If the conditional mean $\mu^* \left[I^\sigma(t)\right]$ and the conditional variance $\sigma^{*2} \left[I^\sigma(t)\right]$ are both affine functions of σ_{t+1}^2 and σ_t^2, there exist quadratic functions $\alpha(\cdot), \beta(\cdot)$ and $\gamma(\cdot)$ such that:

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

with:

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

while:

$$\begin{aligned} a(u) &= l\left(u, 0\right) \Longrightarrow a^*(u) = a\left(u + \varsigma_1 + \alpha\left(\varsigma_2\right)\right) - a\left(\varsigma_1 + \alpha\left(\varsigma_2\right)\right) \\ b(u) &= g\left(u, 0\right) \Longrightarrow b^*(u) = b\left(u + \varsigma_1 + \alpha\left(\varsigma_2\right)\right) - b\left(\varsigma_1 + \alpha\left(\varsigma_2\right)\right). \end{aligned}$$

Then the joint historical distribution of $\left(\sigma_{t+1}^2, r_{t+1}\right)$ can be factorized along the two following univariate CAR models:

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

Then, time series data on the underlying asset return r_{t+1} should allow us to consistently estimate the functions $\alpha(\cdot), \beta(\cdot), \gamma(\cdot), a(\cdot)$ and $b(\cdot)$. This may in some cases allow to identify the risk premium parameters ς_1 and ς_2 insofar as arbitrage pricing relationships draw a connection with the time series characteristics $\alpha(\cdot), \beta(\cdot), \gamma(\cdot), a(\cdot)$ and $b(\cdot)$. We set the focus in this section on the case where the underlying stock price is the only relevant observation, so that the only available arbitrage pricing relationships correspond to Corollary 3.3.:

$$\begin{aligned} a^*\left[\alpha^*(-1)\right] + \beta^*(-1) &= 0 \\ b^*\left[\alpha^*(-1)\right] + \gamma^*(-1) &= 0 \end{aligned} \tag{3.4}$$

We are then led to consider two cases:

1st case: No leverage effect: $\alpha^*(-1) = 0$

In this case, the two arbitrage pricing relationships (3.4) are tantamount to:

$$\beta^*(-1) = \gamma^*(-1) = 0.$$

In terms of the historical characteristics that can be estimated from the underlying asset price we end up with the equations:

$$\begin{aligned} \alpha^*(-1) &= 0 \Longleftrightarrow \alpha\left(\varsigma_2\right) = \alpha\left(\varsigma_2 - 1\right) \\ \beta^*(-1) &= 0 \Longleftrightarrow \beta\left(\varsigma_2\right) = \beta\left(\varsigma_2 - 1\right) \\ \gamma^*(-1) &= 0 \Longleftrightarrow \gamma\left(\varsigma_2\right) = \gamma\left(\varsigma_2 - 1\right). \end{aligned} \tag{3.5}$$

Two remarks are in order:

First, when there is no leverage effect, only the risk premium parameter ς_2 is possibly identified from historical data on the underlying asset price. It would take option price data to identify the other risk premium parameter ς_1. Not surprisingly, the risk premium attached to the volatility risk σ_{t+1}^2 through the definition of the SDF $M_{t+1}(\varsigma_1, \varsigma_2)$ is not identified from observations on the underlying asset price.

Second, the set of constraints (3.5) implies that the quadratic functions $\alpha(\cdot), \beta(\cdot), \gamma(\cdot)$ must be

either of degree 2, or constant, but not linear.

2nd case: Presence of leverage effect $\alpha^*(-1) \neq 0$

Then, statistical identification of the two risk premium parameters may be achieved from the two arbitrage pricing relationships (3.4) that can be rewritten:

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

Two remarks are in order.

First, contrary to a common belief, both risk premium parameters ς_1 and ς_2 may be identified through these equations. In other words, historical data on the underlying asset that would allow to consistently estimate the functions $\alpha(\cdot), \beta(\cdot), \gamma(\cdot), a(\cdot)$ and $b(\cdot)$ would in turn deliver identification of both parameters ς_1 and ς_2.

Second, not surprisingly, the strength of identification of the volatility risk premium parameter ς_1 is tightly related to the strength of leverage effect as characterized by the amplitude of the difference:

$$\alpha\left(\varsigma_2 - 1\right) - \alpha\left(\varsigma_2\right) = \alpha^*(-1).$$

However, it is worth noting that this difference would be more or less enhanced, depending on the slope of the functions $a(\cdot)$ and $b(\cdot)$ that define the volatility factors dynamics. Admitting that these slopes are properly assessed by their values at 0, we can interpret them from the moment functions:

$$\begin{aligned} E[\sigma_{t+1}^2 \,|I(t)] &= a'(0)\sigma_t^2 + b'(0) \\ Var[\sigma_{t+1}^2 \,|I(t)] &= -a''(0)\sigma_t^2 - b''(0) \\ E\left[\sigma_{t+1}^2\right] &= \frac{b'(0)}{1 - a'(o)}. \end{aligned} \tag{3.6}$$

Not surprisingly, the identification power of leverage effect about the volatility risk parameter ς_1 is enhanced by a large volatility persistence $a'(0)$ $(0 \leq a'(0) < 1)$ and a large unconditional level of the volatility factor through large values of both $a'(0)$ and $b'(0)$.

4 Statistical Interpretation of Leverage Effect

Besides the exogeneity of the volatility factor and the affine structure of the conditional distributions (with conditional normality of returns given the path of the latent volatility factor), we have maintained no restrictive assumptions yet. Our main focus of interest on leverage effect and the need

to identify it from return data will lead us in this section to introduce some constraints between the parameters of the functions $\alpha(\cdot), \beta(\cdot), \gamma(\cdot), a(\cdot)$ and $b(\cdot)$ that define the joint distribution of return and volatility factor. These constraints could be relaxed at the cost of a less straightforward interpretation of the parameterization of leverage effect. Before discussing the well suited constraints, it is worth disentangling genuine leverage effect from volatility feedback.

4.1 Leverage effect vs volatility feedback

As already mentioned, leverage effect is generally understood as an instantaneous correlation (given $I(t)$) between asset return and its volatility factor. Our historical model actually says that the knowledge of σ_{t+1}^2 would possibly allow to improve the forecast of the contemporaneous asset return through the following conditional expectation formula:

$$E[\exp(r_{t+1})\,|I^{\sigma}(t)] = \exp\left[-\alpha(-1)\,\sigma_{t+1}^2 - \beta(-1)\,\sigma_t^2 - \gamma(-1)\right].$$

In other words, the impact of σ_{t+1}^2 on the optimal forecast of $\exp(r_{t+1})$ is encapsulated in the number:

$$\alpha(-1) = \frac{\alpha''(0)}{2} - \alpha'(0).$$

It is worth realizing that this quantity does not exactly correspond to the leverage effect as quantified in the previous section by:

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

with:

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

so that, by taking into account that the functions α and α^* are quadratic:

$$Lev = -\frac{\alpha^{*\prime\prime}(0)}{2} + \alpha^{*\prime}(0) = -\frac{\alpha''(0)}{2} + \alpha'(\varsigma_2) = -\frac{\alpha''(0)}{2} + \alpha'(0) + \alpha''(0)\varsigma_2. \tag{4.1}$$

Note that, according to Corollary 3.2., we define the measure Lev of leverage effect as a negative quantity corresponding to the idea of negative risk-neutral correlation between return and volatility. Then, (4.1) shows that this negative quantity actually encapsulates two effects:

$$Lev = -\alpha(-1) + \alpha''(0)\varsigma_2. \tag{4.2}$$

On the one hand, the negative coefficient $[-\alpha(-1)]$ of σ_{t+1}^2 in the optimal forecast of the return $\exp(r_{t+1})$,

On the other hand, since leverage effect also reduces the variance of return r_{t+1} by adding the negative quantity $\alpha''(0)\sigma_{t+1}^2$, the corresponding risk compensation $[-\alpha''(0)\varsigma_2]$ must be subtracted to capture the total leverage effect.

This correction term confirms a well-known difficulty with leverage effect assessment in discrete time. As argued by Bollerslev et al. (2006), it is hardly possible in discrete time to disentangle leverage effect and volatility feedback. This difficulty is confirmed by identity (4.2) that shows that the negative number $[-\alpha(-1)]$ encapsulates the sum of two effects. If we multiply both sides of equality (4.2) by the current volatility factor σ_{t+1}^2, we actually see that the total value includes:

Not only the negative number $(Lev)\sigma_{t+1}^2$ whose absolute value measures the amplitude of leverage effect,

But also the positive number $[-\alpha''(0)\varsigma_2]\,\sigma_{t+1}^2$ that measures the amplitude of volatility feedback effect, as defined by the risk premium parameter ς_2 for risk on return multiplied by the additional variance of return produced by σ_{t+1}^2 :

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

4.2 Variance, correlation and leverage

The above formula suggests to introduce an alternative volatility factor more directly related to the conditional variance of asset returns. By defining:

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

we have:

$$Var[r_{t+1}\,|I^{\sigma}(t)] = -\alpha''(0)\tilde{\sigma}_{t+1}^2.$$

While the initial volatility factor is by definition a Markov process of order 1, the conditional variance of r_{t+1} is proportional to $\tilde{\sigma}_{t+1}^2$ which will be an ARMA(1,1) process. In this respect, we expect to be able to connect the volatility factor with the realized variance of returns between t and $t+1$. The observable conditional variance of return is:

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

with:

$$k_t^2 = \frac{E\{\tilde{\sigma}_{t+1}^2 \,|I(t)\}}{Var\{\sigma_{t+1}^2 \,|I(t)\}} = \frac{E\{\tilde{\sigma}_{t+1}^2 \,|I(t)\}}{Var\{\tilde{\sigma}_{t+1}^2 \,|I(t)\}}. \tag{4.3}$$

The conditional covariance of return with the volatility factor is:

$$\begin{aligned} Cov[r_{t+1}, \sigma_{t+1}^2 \,|I(t)] &= Cov\{E[r_{t+1} \,|I^{\sigma}(t)] \,, \sigma_{t+1}^2 \,|I(t)\} \\ &= Cov\{\alpha'(0)\sigma_{t+1}^2, \sigma_{t+1}^2 \,|I(t)\} = \alpha'(0) Var\{\sigma_{t+1}^2 \,|I(t)\} \,. \end{aligned}$$

Leverage effect must be related to the conditional correlation between return and volatility factor:

$$Corr[r_{t+1}, \sigma_{t+1}^2 \,|I(t)] = \frac{\alpha'(0)}{\left\{[\alpha'(0)]^2 - \alpha''(0) k_t^2\right\}^{1/2}}. \tag{4.4}$$

This formula nicely confirms the intuition relating leverage effect and conditional correlation between return and volatility factor. As seen in (4.1), leverage effect involves a combination of two effects:

On the one hand, the parameter $\alpha'(0)$ encapsulates the impact of the knowledge of σ_{t+1}^2 on the optimal forecast of r_{t+1},

On the other hand, the parameter $\alpha''(0)$ displays how much return variance reduction is allowed by the knowledge of σ_{t+1}^2.

When these two parameters are combined, it is natural to use the weight:

$$k_t^2 = \frac{E\{\tilde{\sigma}_{t+1}^2 \,|I(t)\}}{Var\{\tilde{\sigma}_{t+1}^2 \,|I(t)\}}$$

to transform variance units into expectation units. However, this weight may introduce some random time variation in the conditional correlation between return and volatility factor. Following a common practice in the option pricing literature, we rather want to assume that this correlation is constant, which is tantamount to assuming that the volatility dynamics is such that:

Assumption (1st assumption about leverage): There exists a constant $k > 0$ such that:

$$E\{\tilde{\sigma}_{t+1}^2 \,|I(t)\} = k^2 Var\{\tilde{\sigma}_{t+1}^2 \,|I(t)\} \,.$$

Note that by definition:

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

Note that this first assumption about leverage is not overly restrictive in the context of an affine model, where both the conditional mean and the conditional variance are affine functions of the

volatility factor. By using (3.6), straightforward computations show that the above assumption is tantamount to the following constraint between the parameters of the historical distribution:

$$\frac{1}{a''(0)}\left[a'(0)+\frac{\beta''(0)}{\alpha''(0)}\right]=\frac{1}{b''(0)}\left[b'(0)+\frac{\gamma''(0)}{\alpha''(0)}\right].$$

We will be even more specific by assuming that the two leverage parameters $\alpha'(0)$ and $\alpha''(0)$ are related as follows:

Assumption (2nd assumption about leverage):

$$[\alpha'(0)]^2-\alpha''(0)k^2=k^2.$$

Thanks to these two assumptions, we end up with a nice parameterization of leverage effect by defining:

$$\phi=Corr[r_{t+1},\sigma_{t+1}^2\,|I(t)]=\frac{\alpha'(0)}{k}\in[0,1].$$

In other words, the two above assumptions about leverage allow us to characterize leverage through the parameter ϕ of conditional correlation between return and its volatility factor. In particular, for a given weight k, ϕ characterizes the two leverage parameters $\alpha'(0)$ and $\alpha''(0)$:

$$\begin{aligned}\alpha'(0) &= k\phi\\ -\alpha''(0) &= 1-\frac{[\alpha'(0)]^2}{k^2}=1-\phi^2\geq 0.\end{aligned}$$

The distorted assessment of leverage effect from historical data as documented by (4.1) and (4.2) can be revisited with this reparameterization:

$$Lev=k\phi-\left(1-\phi^2\right)\left(\varsigma_2-\frac{1}{2}\right).$$

While the conditional correlation parameter ϕ, when weighted by k, encapsulates the statistical notion of leverage, we must subtract the term $\left(1-\phi^2\right)\varsigma_2$ and add the term $\frac{1-\phi^2}{2}$ to get the exact measure of leverage. While the latter term can be interpreted as a correction for a Jensen effect, the subtraction of $(1-\phi^2)\varsigma_2$ makes leverage even more negative to compensate for perverse effect of volatility feedback.

With our parameterization, we can then rewrite the option pricing formula (2.2) as:

$$\begin{aligned}C_t(K) &= E^*[BS_{(t)}\left(K,\tilde{S}_t(\phi),\left(1-\phi^2\right)\tilde{\sigma}_{t+1}^2\right)|I(t)] \qquad (4.5)\\ \tilde{S}_t(\phi) &= S_t\frac{\exp\left(k\phi\sigma_{t+1}^2\right)}{E[\exp\left(k\phi\sigma_{t+1}^2\right)|I(t)]}.\end{aligned}$$

Formula (4.5) revisits in discrete time a formula first shown by Romano and Touzi (1997) (see also Garcia, Ghysels, and Renault (2010)). The leverage effect parameter ϕ plays a double role in the option pricing formula.

On 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.

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-of-the-money side. Figure 4 in Appendix presents the shapes of volatility smiles from our SV model for different values of ϕ (thus, different levels of leverage effect). As we can see, our SV model produces a symmetric volatility smile when there exists no leverage effect (i.e., $\phi = 0$) with the implied volatility minimized at the money, and symmetry starts to be distorted as the leverage effect increases.

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.

4.3.1 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_{t+1}^2 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_t^2 is Poisson with parameter $\varrho\sigma_t^2/c$.

We easily verify that this parametric model is nested in the general affine model defined with functions $a(\cdot)$ and $b(\cdot)$ as in Section 3.3., with the specification:

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

Then:

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

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

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

Note that while the sequence of returns $t = 1, 2, \cdots, T$, can be seen daily, the volatility factor σ is latent. We want to take advantage of observations of daily realized variances RV_t, $t = 1, 2, \cdots, T$ for the identification of this volatility factor. But, in order to do so, we need to completely characterize the conditional distribution of returns.

4.3.2 Return dynamics

The return process conditioning on latent volatility factor is Gaussian:

$$r_{t+1} | I^\sigma(t) \sim \mathcal{N}\left(E\left[r_{t+1} | I^\sigma(t)\right], Var\left[r_{t+1} | I^\sigma(t)\right]\right),$$

with the following two conditional moments:

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

So far, we have parameterized the function $\alpha(\cdot)$ only by characterizing the two leverage parameters $\alpha'(0)$ and $\alpha''(0)$ with ϕ given the weight k. In this section, we provide the parametric model for the return dynamics by completing the parameterization of k, $\beta'(0), \beta''(0), \gamma'(0)$ and $\gamma''(0)$. Through the previous sections, we developed the identification schemes for the volatility risk premium ς_1 and the leverage effect. As we have seen, they involve some constraints among the historical parameters of returns and volatility that we must take into account.

We first set our focus on the restriction for the constant leverage effect. It is imposed by the first

assumption about leverage, which is equivalent to imposing a constraint between $\beta''(0)$ and $\gamma''(0)$. See that, with our parametric model of the volatility, the constraint given in Section 4.2 can be written as:

$$\frac{1}{2\rho c}\left[\rho + \frac{\beta''(0)}{\alpha''(0)}\right] = \frac{1}{\delta c^2}\left[\delta c + \frac{\gamma''(0)}{\alpha''(0)}\right],$$

which shows that $\beta''(0)$ and $\gamma''(0)$ are determined by each other given the other parameters. Let us define new parameters e and f as:

$$e = \frac{\beta''(0)}{\rho(1-\phi^2)} \text{ and } f = \frac{\gamma''(0)}{\delta c(1-\phi^2)}.$$

Then, by rearranging the above constraint we see that f is written in terms of e such that:

$$\gamma''(0) = \frac{1}{2}\left[1 + \frac{\beta''(0)}{\rho(1-\phi^2)}\right]\delta c(1-\phi^2) \iff f = \frac{(1+e)}{2}.$$

With this, we can now provide the specification of the weight k in terms of the historical parameters:

$$k^2 = \frac{1-e}{2c},$$

and rewrite the alternative volatility factor $\tilde{\sigma}^2_{t+1}$ as:

$$\tilde{\sigma}^2_{t+1} = \sigma^2_{t+1} - e\rho\sigma^2_t - f\delta c.$$

Now it takes the parameterization of the two numbers $\beta'(0)$ and $\gamma'(0)$ for the complete parameterization of the return dynamics. These two quantities are parameterized through the conditions for the statistical identification of the volatility risk premium parameter ς_1 provided by Corollary 3.3. From them (see the rewritten conditions), we get:

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

since:

$$\bar{\omega}(\varsigma_2) - \bar{\omega}(\varsigma_2 - 1) = \bar{\omega}'(0) + \bar{\omega}''(0)\left(\varsigma_2 - \frac{1}{2}\right),\ \bar{\omega} \in \{\beta(\cdot), \gamma(\cdot)\}.$$

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 :

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

4.3.3 Comparison to continuous-time affine models

As we show in Appendix C, our parametric model has the Heston (1993) model as its continuous time limit. In addition, the general version of our affine model can be considered as the discrete time extension of the continuous time affine models that are affine diffusion models (AD). In this section, we briefly discuss the benefits and costs of modeling in discrete time rather than in continuous time.

Both our model and AD compute the derivative prices in semi-closed form. However, discrete time models have a more degree of freedom to reproduce the higher order moments of returns such as negative skewness than AD since AD assumes the conditional normality of returns, up to jumps, which are the only degrees of freedom for higher order moments. On the other hand, discrete time models are not restricted to assume conditional normality. In fact, the third conditional moment of the returns of our model is expressed as

$$\begin{aligned}s_t &= E\left[(r_{t+1} - \mu_t)^3 | I(t)\right] \\ &= 3(1-\phi^2)k\phi Var\{\sigma_{t+1}^2 \,|I(t)\} + 3(k\phi)^3 E\{(\sigma_{t+1}^2 - E[\sigma_{t+1}^2 | I(t)])^3 \,|I(t)\} \\ &= 3(1-\phi^2)k\phi(2\rho c\sigma_t^2 + \delta c^2) + 3(k\phi)^3(6\rho c^2\sigma_t^2 + 2\delta c^3) \\ &= k_1\sigma_t^2 + k_2,\end{aligned}$$

where $k_1 = 6k\phi\rho c((1-\phi^2) + 3(k\phi)^2 c)$ and $k_2 = 3k\phi\delta c^2((1-\phi^2) + 2(k\phi)^2 c)$. The time varying skewness is allowed unless $\phi = 0$, i.e. no leverage effect.

However, as already discussed, it is hard to distinguish the leverage effect from volatility feedback effect in discrete time and we leave room for a mixture of these two effects. While continuous time modeling characterizes the leverage effect as an instantaneous correlation between returns and volatility, the leverage effect is time-varying in nature and we have to impose a constraint on the parameters in order to have it as a time-invariant constant (see Section 4.2) to see its effect on the shape of volatility smile.

5 Estimation methodology

5.1 Estimation procedure

5.1.1 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):

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

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

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:

$$RV_{t+1} = \sigma_{t+1}^2 - B\sigma_t^2 - D. \tag{5.1}$$

Several remarks are in order.

First, we choose a unit coefficient for σ_{t+1}^2 in formula (5.1). 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 will impose the restriction:

$$\mu_t = E[RV_{t+1} \,|I(t)] = Var[r_{t+1} \,|I(t)]\,. \tag{5.2}$$

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 (5.2) that will be a maintained assumption throughout this paper.

By the assumptions about leverage provided in the subsection 4.2, the conditional variance of returns is:

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

so that the optimal forecast of $\tilde{\sigma}_{t+1}^2$ coincides with it. Then the maintained assumption (5.2) together with the specification of RV in (5.1) amounts to $B = e\rho$ and $D = f\delta c$ so that:

$$RV_{t+1} \;=\; \tilde{\sigma}_{t+1}^2 \;=\; \sigma_{t+1}^2 - e\rho\sigma_t^2 - f\delta c. \tag{5.3}$$

The equality (5.3) of RV_{t+1} and $\tilde{\sigma}^2_{t+1}$ gives an easy way to check the empirical validity of the first assumption about leverage. 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:

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

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 the S&P500 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.

5.1.2 Two step estimation with returns and realized variance

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 exponentially affine conditional distributions:

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

(ii) The conditional distribution of r_{t+1} given σ^2_{t+1} and σ^2_t, 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 θ_σ.

Even though we have a parametric model where the likelihood function exists and Maximum Likelihood Estimation (MLE) would deliver efficient estimation, the support of volatility ($\sigma^2_{t+1} > 0$) depends on the unknown parameters, which make it difficult to use MLE. Then it is convenient to use the conditional moment restrictions directly provided by the exponential affine conditional moment generating function for a GMM strategy.

In addition, assuming $|e\rho| < 1$ and inverting the ARMA(1,1) realized variance given in the

previous section, we have:

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

Then, since $(e\rho)^k \to 0$ as $k \to \infty$ by the assumption that $|e\rho| < 1$, we can use the approximation

$$\sigma_t^2 \approx \sum_{k=0}^{H} (e\rho)^k RV_{t-k} + \frac{f\delta c}{(1-e\rho)} \tag{5.4}$$

for some constant H.

Using this, we implement the following two-step estimation[1]. Seeing that the approximated process of realized variance is characterized by $\theta_{rv} = (\theta_\sigma', e)' = (\rho, \delta, c, e)'$, we fist estimate θ_{rv} using GMM and construct

$$\hat{\sigma}_t^2 = \sum_{k=0}^{H} (\hat{e}\hat{\rho})^k \left(RV_{t-k}\right) + \frac{\hat{f}\hat{\delta}\hat{c}}{1-\hat{e}\hat{\rho}}.$$

Then, with $\hat{\sigma}_{t+1}^2$ as an approximated observation of the latent volatility factor, we estimate the returns parameters ϕ, ς_1, and ς_2 by MLE for given estimates of θ_{rv} from the first step. However, the identification of ς_1 is weak from the observations of returns. We find the estimate of it too sensitive to the initial values that we start the nonlinear estimation with. Thus, we use the options data to identify ς_1 by finding the value that minimizes the option pricing error of the model. The option pricing error is measured by IVRMSE put forward by Renault (1997):

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

where N denotes the number of observations, and where IV_i^{hist} and IV_i^{mod} denote the i-th observation of historical implied volatility and the implied volatility generated by the model, respectively. Then, ς_1 is computed by maximizing the Gaussian IV option-error likelihood defined as:

$$L^{Op} = -\frac{1}{2}\sum_{i=1}^{N} \left(log(IVRMSE^2) + e_i^2/IVRMSE^2\right) \tag{5.5}$$

where $e_i \equiv IV_i^{hist} - IV_i^{mod}$.

While the returns estimation is straightforward using the conditional Gaussian likelihood, the

[1]Note that to take into account the correction factor $AVar[\sigma_{t+1}^2 \,|I(t)]$ (i.e., $A \neq 0$), 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)).

GMM estimation with the realized variance needs some discussion about moment conditions and weighting matrix.

5.1.3 GMM estimation with realized variance

The purpose of this subsection is to describe the first step with the observations of realized variance. When $e\rho$ 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 the approximated CF by inverting the ARMA(1,1) realized variance to the AR(H) process as given in (5.4).

By plugging this into the conditional CF of volatility given in Section 4.3.1, we get

$$E\left[\exp\left(-uRV_{t+1}\right)|I(t)\right] \approx \exp\left\{-(a(u)-e\rho u)\left(\sum_{k=0}^{H-1}(e\rho)^k RV_{t-k}\right) - b(u) - (a(u)-u)\frac{f\delta c}{1-e\rho}\right\}$$

so that the realized variance process is approximately CAR(H). We construct the unconditional moment restrictions using the instrument $A_t = \exp(-a_1 RV_t - a_2 RV_{t-1})$ for some $(a_1, a_2)' \in \mathbb{C}^2$:

$$E\left[A_t\left(\exp\left(-uRV_{t+1}\right) - \exp\left\{-(a(u)-e\rho u)\left(\sum_{k=0}^{H-1}(e\rho)^k RV_{t-k}\right) - b(u) - (a(u)-u)\frac{f\delta c}{1-e\rho}\right\}\right)\right] \approx 0.$$

This choice of unconditional moments ensures the identification of the parameters of interest. A comprehensive discussion of identification of unknown parameters from the moment conditions is provided in Appendix D.

Let $h_t(\tau;\theta_{rv})$ denote such moments of realized variance with $\tau = (a_1, a_2, u)'$. For our empirical exercise, we use $H = 10$, and 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

$$E\left[g_t(\tau;\theta_{rv})\right] = E\begin{bmatrix} Re\left\{h_t(\tau;\theta_{rv})\right\} \\ Im\left\{h_t(\tau;\theta_{rv})\right\} \end{bmatrix} = 0,$$

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 GMM estimator is then defined as:

$$\hat{\theta}_{rv,T} = \underset{\theta_{rv}}{Argmin}\ \bar{g}(\tau,\theta_{rv})'\hat{W}_T^{-1}\bar{g}(\tau,\theta_{rv}), \tag{5.6}$$

where $\bar{g}(\tau, \theta_{rv}) = (1/T)\sum_{t=1}^{T} g_t(\tau; \theta_{rv})$ 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

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

and $\hat{W}_T = \hat{V}_T = (1/T)\sum_{t=1}^{T} g_t(\tau; \tilde{\theta}_{rv,T}) g_t(\tau; \tilde{\theta}_{rv,T})'$, where $\tilde{\theta}_{rv,T}$ is a preliminary consistent estimator of the true parameter value θ_{rv}^0. In our empirical exercise, it is computed using an identity matrix as a weighting matrix.

However, even with a small dimensional τ, $\hat{W}_T$, 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:

$$W^{-1} = (V^2 + \alpha I)^{-1} V \tag{5.7}$$

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 (5.6) 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.

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

Note that we do not use or propose a data dependent selection method of a tuning parameter α that leads to the efficient estimation since it is beyond the scope of this paper. Our focus is the

[2] (5.7) 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.

consistent estimation and we choose $\alpha = 0.1$ that is big enough to ensure a small enough bias.

5.2 Empirical results

In this section, we present an empirical application of the our SV model applying the estimation method discussed in Section 5.1. We first estimate the parameters using the observations of returns and realized variance on the S&P500 index. Then, using the options data, we examine the option pricing performance of the 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) and the Affine RV model (ARV hereafter) by Christoffersen, Feunou, Jacobs, and Meddahi (2014).

5.2.1 Competitor models

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

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

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).

Another model that we consider is a GARCH-type option pricing model that is developed by Christoffersen et al. (2014) where the realized variance component plays a role in the variance

dynamic of returns. This model is dubbed the ARV model by Christoffersen et al. (2014)[4]. This model assumes the following dynamic model of daily returns:

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

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

$$h_{t+1}^{RV} = \omega + \beta h_t^{RV} + \alpha_2 \left(\epsilon_{2,t+1} - \gamma\sqrt{h_t^{RV}}\right)^2,$$

where $(\epsilon_{1,t+1}, \epsilon_{2,t+1})$ follows jointly a bivariate standard normal distribution with correlation ρ_a. The observations of realized variance are linked to h_t^{RV} as follows

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

The persistence of daily variance is also captured by the form $(\beta + \alpha\gamma^2)$.

Note that λ for both models represents the risk price of returns that is the same as ς_2 in our model. ARV has an additional risk price parameter $\mathcal{X}$ to HN that represents the volatility risk price that is comparable to ς_1 in our model. It is identified with the options data:

$$\gamma^* = \gamma - \mathcal{X},$$

where γ^* is the risk neutral γ that is computed, with other parameter estimates from the observations of returns, by maximizing the Gaussian IV option-error likelihood defined in (5.5).

Both HN and ARV models are estimated using quasi-maximum likelihood (QMLE) techniques. From the observations of returns and realized variance, the moments used for the estimation of ARV in addition to the first moment of realized variance are:

$$\begin{aligned}
E\left[r_{t+1}|I(t)\right] &= \left(\lambda - \frac{1}{2}\right) h_t^{RV} \\
Var\left[r_{t+1}|I(t)\right] &= h_t^{RV} \\
Var\left[h_{t+1}^{RV}|I(t)\right] &= 2\alpha^2\left(1 + 2\gamma^2 h_t^{RV}\right) \\
Cov\left[r_{t+1}, h_{t+1}^{RV}|I(t)\right] &= -2\rho_a\alpha\gamma h_t^{RV},
\end{aligned}$$

[4]The ARV model is a special type of the GARV model (Christoffersen et al. (2014)) where the variance dynamic of returns depends both on realized variance and returns. They show that the GARV model outperforms the ARV model in terms of option pricing but we use the ARV model for comparison to be comparable with our affine model that the latent volatility is identified by the observations of realized variance.

and the conditional correlation between returns and volatility is

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

We estimate the two models using quasi maximum likelihood (QMLE). We use the conditional log-likelihood of returns for HN and the conditional joint log-likelihood of returns and realized variance for ARV. The estimation results are given in Table 1 along with them of some versions of our model. Note that ω for ARV is estimated using the unconditional variance targeting such that:

$$\omega = E\left[h_t^{RV}\right]\left(1-\beta-\alpha\gamma^2\right)-\alpha. \tag{5.8}$$

where we compute $E\left[h_t^{RV}\right]$ as the sample mean of realized variance.

5.2.2 Parameter estimation

In this section, we estimate the model parameters from the returns and realized variance on the S&P 500 index. The dataset was obtained from Oxford-Man Institute[5] 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[6]. The realized variance process $\{RV_t\}$ is computed from 5-minute intraday returns.

The estimation results of the ARG(1)-Normal model (AN henceforth) in different settings from the observations of returns and realized variance are given in Table 1 along with those of the HN and the ARV models[7]. Several remarks are in order. First, both ς_2 (for AN) and λ (for HN and ARV) represent the risk price of the returns. Likewise, both ς_1 (for AN) and $\mathcal{X}$ (for ARV) model are the risk price of volatility. There is no volatility risk price (ς_1) in the HN model since it uses only the risk price of returns. Second, the parameter ρ_a from the ARV model can be roughly interpreted as the leverage effect both in absolute value which is comparable to ϕ for our AN model, although it itself does not directly determine it. The conditional correlation formula between returns and volatility provided in Section 5.2.1 for the ARV model shows that leverage increases in ρ_a given α, γ and volatility. We in fact see in Figure 1 below that the correlation in absolute value fluctuates at around the value close to the estimate of ρ_a given in Table 1.

The first to the third columns of Table 1 present the estimation results of the AN models from the daily observations of returns and realized variance. The first column (ANNL) imposes the zero

[5]Oxford-Man Institute's "realized library," http://realized.oxford-man.ox.ac.uk

[6]This rate is obtained from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data library.html

[7]Note that the parameter ω is directly estimated with other parameters for the HN while it is implied from the unconditional variance formula (see (5.8)) for the ARV. We apply different estimation procedures for ω because they lead to the best performance of those two models from our various exercises with the given data.

leverage effect (i.e. $\phi = 0$). The first and second columns (ANL) imposes the condition for the constant leverage effect

$$f = \frac{1+e}{2},$$

while the third (ANTL) one allows the leverage effect to be time-varying. The ANTL model still uses the restriction[8]

$$k^2 = \frac{1-e}{2c}.$$

Note that the estimation results of ANNL and ANL only differ with the estimation of the returns parameters since f is estimated from the observations of realized variance only. In all estimations of the class of AN model, we use the two step estimation of the mixture of GMM and MLE described in Section 5.1. The Figure 1 below shows the conditional correlation between returns and volatility computed using the formula given in the equation (4.4). The top panel plots the daily realized variance, the second panel plots the conditional correlation for the model with the constant leverage effect, and the third to the last panels present the results of the ANTL, HN, and ARV, respectively. The last two columns report the parameter estimates of the competitor models, the HN and ARV models. Both models positively estimate γ and α and the ARV model positively estimates ρ_a, which confirms the negative correlation between returns and volatility.

Comparing the AN, HN and ARV models, the AN and ARV models produce the similar leverage effect while the HN shows a significant difference. As already expected, the ANL generates a time-invariant correlation estimated at -0.1718. When the time varying feature is allowed, our model generates the conditional correlation which looks close to that of ARV. Both ANTL and ARV fluctuates at around a value between -0.15 and -0.16 with the shapes that are almost identical to each other. In both models, the negative correlation is stronger when the level of volatility is higher. The leverage effect from the HN model is much larger than the other two models with the correlation fluctuating between -0.65 and -1[9]. This result of the high correlation between returns and volatility of the HN model is in fact in line with the restriction of the GARCH model that the innovations of returns and volatility share the same process.

All models present high levels of volatility persistence (close to 0.98 for the class of AN models, and 0.96 for the HN and ARV), which is consistent with the empirical findings in the literature. However, the ARV model differs in the estimate of the return risk premium while the AN models and the HN model provide similar values. The reported estimate of λ is 0.0540 and this is much smaller than the estimates of ς_2's and λ by the HN model that are greater than 1, although they are all statistically not significant.

As we developed in Section 2, the leverage effect is closely related to the shape of the volatility

[8]We also estimate our AN model with k as a free parameter and plot the conditional correlation between returns and volatility in Figure 6. We do not report the parameter estimates because they are dependent on the value of ς_1 which is not well identified and the conditional correlation does not depend on the those different parameter estimates.

[9]The average conditional correlation for the period of January, 2000 to June, 2016 is -0.9260.

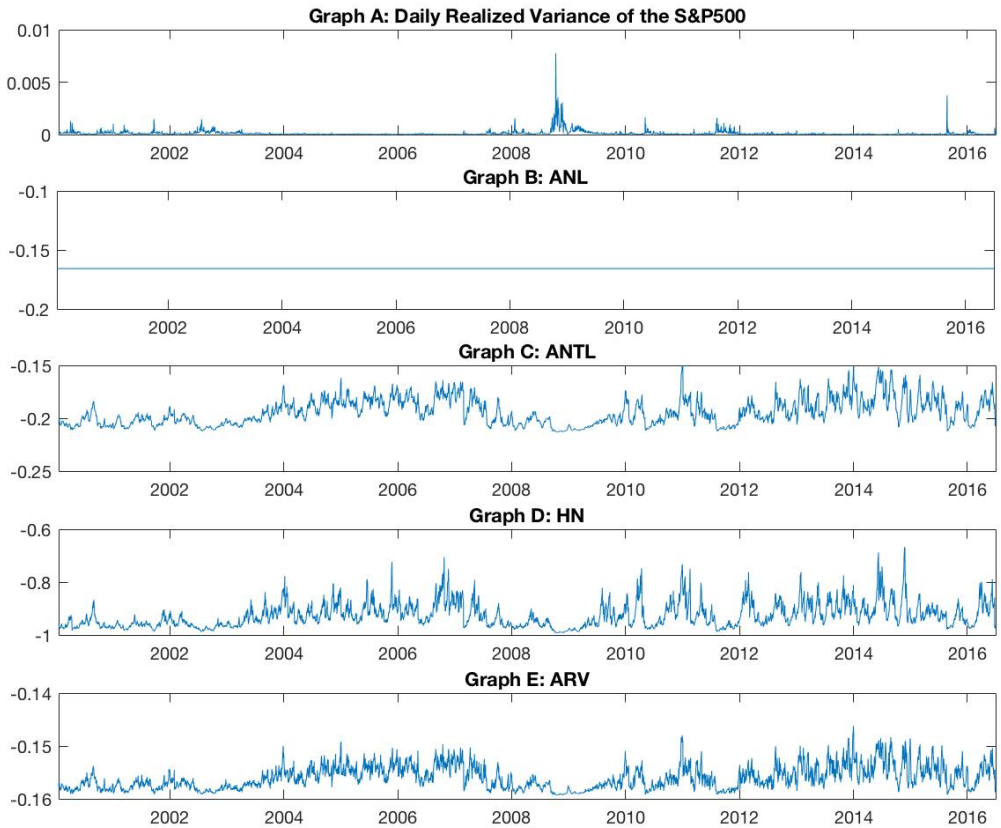
Table 1: Estimates of the parameters on S&P500 index returns and realized variance

	ANNL	ANL	ANTL		HN	ARV
ρ	0.9802	0.9802	0.9737	α	4.06e-6	4.74e-6
	(0.0269)	(0.0269)	(0.0273)		(3.83e-7)	(1.34e-6)
δ	1.1236	1.1236	1.0773	β	0.8015	5.06e-10
	(0.5132)	(0.5132)	(1.0429)		(0.0047)	(0.2738)
c	8.67e-06	8.67e-06	1.14e-05	ω	5.18e-9	2.84e-9
					(1.67e-07)	
e	0.7470	0.7470	0.7557	σ		9.76e-6
	(0.0331)	(0.0331)	(0.0318)			(2.71e-6)
f	0.8735	0.8735	0.0264	ρ_a		0.1593
	(0.0165)	(0.0165)	(2.2060)			(0.01)
ϕ	0	-0.1718	-0.1928	γ	199.70	450.19
		(0.0140)	(0.0163)		(11.4435)	(127.19)
ς_2	0.9286	1.1236	1.2494	λ	1.4934	0.0540
	(1.3420)	(1.2831)	(1.2815)		(1.2768)	(1.3552)
ς_1	-1.46e-6	-0.02	-0.01	$\mathcal{X}$		-5.2125
	(0.0991)	(1.9333)	(1.31e-07)			(0.0843)
Persistence	0.9802	0.9802	0.9737		0.9632	0.9614

* We estimate the models using the daily observations of returns and realized variance for the S&P500 index for the period January 1, 2000, to June 30, 2016. The volatility parameters of the AN models, (ρ, δ, c, e, f), are estimated by GMM with the moment conditions provided in subsection 5.1.3, and the return parameters of the AN models, $(\phi, \varsigma_1, \varsigma_2)$, are estimated by MLE given the volatility parameter estimates. The HN and ARV models are estimated by MLE.
* The standard errors (s.e.) are given in parentheses.
* The s.e. of ς_1 and $\mathcal{X}$ are computed using the outer product of gradient.
* In the first and second columns, f is computed as $(1+e)/2$ and the s.e. is computed as $s.e.(e)/2$.
* In the first to the third columns, k^2 is computed as $(1-e)/2c$.
* In the first to third column, c is computed from the equation $E[RV_{t+1}] = (1-e\rho)\frac{\delta c}{(1-\rho)} - f\delta c$ where $E[RV_{t+1}]$ is estimated by the sample mean of realized variance.
* In the fifth column, ω is computed as $E[h_t]\left(1-\beta-\alpha\gamma^2\right)-\alpha$ where $E[h_t]$ is estimated by the sample mean of realized variance.

smile. We expect a bigger leverage effect when the volatility smile is more pronounced. Figure 2 below plots the average implied volatilities for given intervals of log moneyness of the options data on S&P500 from January 3, 2000 to December 31, 2006 together with those generated from the different models. We see that the volatility smile from the market data is clearly skewed implying the presence of significant leverage effect. In the next subsection, we carry out empirical analysis on the options data in order to see whether the returns data identifies the desired level of leverage effect that produces the skewed smile and how the option pricing error is affected by such identification.

Figure 1: Daily conditional correlation of returns and volatility, 2000-2016



** The top panel plots the daily observations of realized variance from January 2000 to June 2016. For Graph B and C, which show the ANL and ANTL models, we plot the daily conditional correlation between returns and volatility, $Corr_t(r_{t+1}, \sigma^2_{t+1})$. For the ANL, the constant leverage effect restriction, $f = (1+e)/2$, is imposed while it is a free parameter for the ANTL. For Graph D, which shows the HN, we plot the daily conditional correlation between returns and volatility, $Corr_t(r_{t+1}, h_{t+1})$. For Graph E, which shows the ARG, we plot the daily conditional correlation between returns and volatility, $Corr_t(r_{t+1}, h^{RV}_{t+1})$.The conditional correlations are computed using the formula given in (4.4) (for the AN models) and Section 5.2.1 (for the HN and ARV) with the parameter estimates given in Table 1.*

Table 2: Option pricing performances

	ANL	ANNL	HN	ARV	ANL*	ARV*
$IVRMSE$	3.8758	4.2013	4.6155	3.9627	3.3830	3.3998

** The IVRSME is computed with the implied volatilities generated from the models. For ANL, ANNL, HN, and ARV, we get the implied volatilities with risk neutral parameters computed with the parameter estimates in Table 1. All IVRMSE values are in annualized percentage points.*
** The IVRMSE for ANL* is computed with (s.e. in parenthesis) $\phi^* = -0.5137$ (0.0032) which is the value that minimizes the Gaussian IV option-error likelihood, with the other parameter values in the second column of Table 1. The s.e. is computed using the outer product of gradient.*
**The IVRMSE for ARV* is computed with (s.e. in parenthesis) $\rho_a^* = 0.4352$ (0.0044) which are the values that minimizes the Gaussian IV option-error likelihood, with the other parameter values in the fifth column of Table 1. The s.e. is computed using the outer product of gradient.*

5.2.3 Option pricing performance

In order see option pricing performance, we use European options written on the S&P500 index. The data was downloaded from Optionmetrics[10] and the observations range from January 3, 2000 to December 31, 2006. 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 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 12,241.

We compute the prices of each option for given K, S_t and time to maturity following the steps described above for ANL, ANNL, HN and ARV models. In order to compare the option pricing performances of these model, we use the percentage implied volatility (IVRMSE) defined in the previous subsection as a pricing error. The results are presented in Table 3 which shows that the ANL and ARV models have the smallest IVRMSEs with that of ANL is slightly smaller. The option pricing error results are the worst for ANNL and HN which indicates that either no or too big leverage effect does not do a good option pricing.

In order to see whether the leverage effect is adequately identified from the returns data, we plot the volatility smiles of the data together with those generated from the models in Figure 2 below where the implied volatilities are the average of them for a given interval of log moneyness over the period January, 2000, to December, 2006. We see that none of the models with the parameter values estimated from the returns data is able to generate the smiles as pronounced as the smiles of the market options. Either the leverage effect is estimated to be too small or too big to reproduce the shape of the volatility smile from the data.

We then compute the parameter estimates using the options data to see the leverage effect that

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

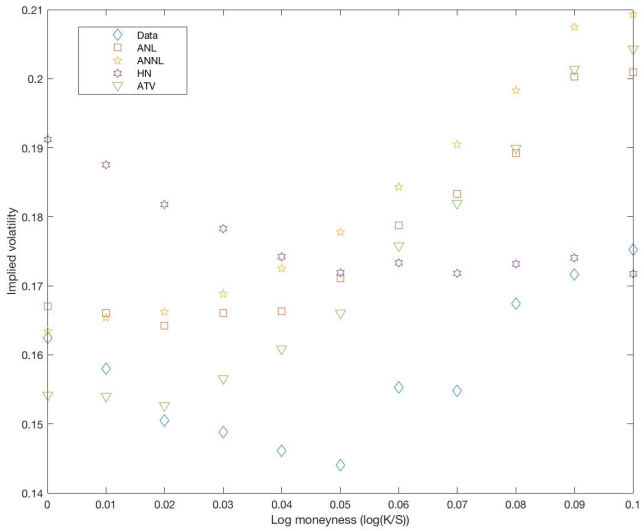
fits the options. Although it is standard in the option pricing literature to estimate the whole risk neutral parameters from the options data, we only estimate parts of the parameters (ϕ for ANL and ρ_a for ARV) given the other parameter estimates presented in Table 1. Those parameters are tightly related to the correlation between returns and volatility even though we cannot show how they are determined jointly with the other parameters. We do not carry out this exercise for the HN because γ is linked to the correlation in HN but it is also closely related to the persistence of volatility. It is the same for ARV as well but ARV has an additional degree of freedom for the leverage effect, ρ_a, which is the correlation parameter between the innovations of returns and volatility.

We estimate ϕ and ρ_a by maximizing the Gaussian IV option-error likelihood defined in (5.5). The resulting IVRMSEs are given in Table 2 under ANL* and ARV* with the parameter estimates in the caption of Table 2. This exercise identifies a much bigger leverage effect for both ANL and ARV, with $\phi = -0.5137$ and $\rho_a = 0.4352$, than the ones computed from the returns data. The option pricing performance for both models improves significantly.

Figure 3 plots the average implied volatilities of ANL and ARV with the new estimates of ϕ and ρ_a, respectively. We see that the shape of volatility smile with this larger value of ϕ is much closer to that of the data but ARV fails to generate the large skewness of the smile even with a large value of ρ_a (thus, a bigger leverage effect). Although the two models show different abilities in producing the volatility smile, the IVRMSEs computed from them are similar. However, ANL seems to over-estimate implied volatilities which is largely attributable to the over-estimation of the latent volatility σ_t^2. The volatility is filtered from the observations of realized variance using a simple two-step procedure but this method is obviously not efficient. It seems that the option pricing performance of the ANL could improve if we apply a more efficient estimation method that can decrease the bias of filtering the volatility.

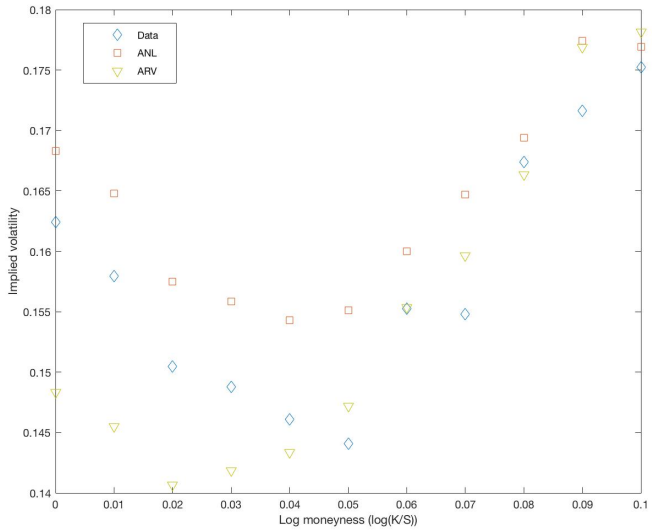
We categorize options according to their time to maturity and moneyness where moneyness is defined as $log(K/S_t)$ with K and S_t denoting a strike price and a price of the underlying asset at time t. Table 3 presents some descriptions of the options data and IVRMSE for each maturity and moneyness category of the ANL, HN and ARV models. In terms of IVRMSE, the option pricing performance of the HN model is dominated by ANL in all categories. ARV also dominates HN in almost all categories. Comparing the ANL and ARV models, ANL perform better for the options that are not relatively deep out-of-money (OTM) and have relatively short maturities. This result confirms our discussion in Section 2 that SV models are more flexible to produce the skewed volatility smiles than the GARCH models where the volatility is not latent. Also, from the generated volatility smiles in Figure 3, the outperformance of ANL over ARV with smaller moneyness is resulted from its ability to produce the skewness closer to the actual data.

Figure 2: Volatility smiles from the options data



** For this figure, we use out of money call options with time to maturity between 15 and 180 days and implied volatility less than 0.7, over the period of January, 2000, to December, 2006.*
**The implied volatilities are the average implied volatilities of the data and the ones implied by different models, with certain values of log moneyness. For example, the implied volatility for log moneyness 0.01 is the average implied volatility of the options with log moneyness between 0.005 and 0.015. The implied volatilities for log moneyness 0 and 0.1 are the average implied volatility of the options with log moneyness between 0 and 0.005, and between 0.095 and 0.1, respectively.*

Figure 3: Volatility smiles from the options data



** For this figure, we use out of money call options with time to maturity between 15 and 180 days and implied volatility less than 0.7, over the period of January, 2000, to December, 2006.*
**The implied volatilities are the average implied volatilities of the data and the ones implied by ANL and ARV, with certain values of log moneyness. For example, the implied volatility for log moneyness 0.01 is the average implied volatility of the options with log moneyness between 0.005 and 0.015. The implied volatilities for log moneyness 0 and 0.1 are the average implied volatility of the options with log moneyness between 0 and 0.005, and between 0.095 and 0.1, respectively.*
** The implied volatilities for ANL are computed with $\phi = -0.5137$ and other parameter values given in the second column of Table 1.*
** The implied volatilities for ARV are computed with $\rho_a = 0.4352$ and other parameter values given in the sixth column of Table 1.*

6 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 $\alpha'(0)$),

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

Table 3: 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	2,351	4,914	3,345	738	867
Ave. IV(%)	16.87	15.65	16.46	16.36	16.79
ANL					
IVRMSE	3.4421	3.7644	3.9858	4.3789	4.6447
ANL*					
IVRMSE	3.5365	3.2242	3.3368	3.5854	3.8159
HN					
IVRMSE	4.2668	4.7610	4.6446	4.4970	4.6808
ARV					
IVRMSE	3.8564	4.0088	3.9303	3.9237	4.1405
ARV*					
IVRMSE	3.5997	3.3775	3.3372	3.2722	3.3078

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	2,687	2,408	2,081	2,672	2,367
Ave. IV(%)	15.83	14.82	14.53	16.20	19.64
ANL					
IVRMSE	3.1334	3.4233	3.8466	4.1392	4.7013
ANL*					
IVRMSE	3.1160	3.1092	3.2054	3.4676	3.9491
HN					
IVRMSE	4.7081	4.7022	4.4774	4.0752	5.0897
ARV					
IVRMSE	3.1539	3.1755	3.7574	4.1441	5.2539
ARV*					
IVRMSE	3.3849	3.1623	3.1940	3.3902	3.8095

* We use out of money call options with time to maturity between 15 and 180 days and implied volatility less than 0.7, over the period of January, 2000, to December, 2006. We report IVRMSE from the models by moneyness and maturity. The ANL, HN, and ARV models are estimated in Table 1. ANL* and ARV* models are estimated with $\phi^* = -0.5137$ and $\rho_a^* = 0.4352$ together with the other parameter estimates in Table 1. All IVRMSE values are in annualized percentage points.
* Moneyness is K/S.
* IV stands for Implied Volatility.

ς_2),

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, the empirical exercise on the S&P500 in index and options still shows that the leverage effect is under-estimated from the returns data and 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., B. Feunou, K. Jacobs, and N. Meddahi (2014): "The Economic Value of Realized Volatility: Using High-Frequency returns for Option Valuation," *Journal of Financial and Quantitative Analysis*, 49, 663–697.

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.

Gagliardini, P., C. Gourieroux, and E. Renault (2011): "Efficient derivative pricing by the extended method of moments," *Econometrica*, 79, 1181–1232.

Garcia, R., E. Ghysels, and E. Renault (2010): "The econometrics of option pricing," *Handbook of Financial Econometrics,* ed. by Y. Ait-Sahalia and L.P. Hansen, North Holland, Vol. 1, 479–552.

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,* ed. by R.F. Engle and D.L. McFadden, Elsevier Science, Vol. 4, 2111–2245.

Renault, E. (1997): "Econometric models of option pricing errors," *Advances in Economics and Econometrics: Theory and Application,* ed. by D.M. Kreps and K.F. Wallis, Cambridge University Press, Vol. 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.

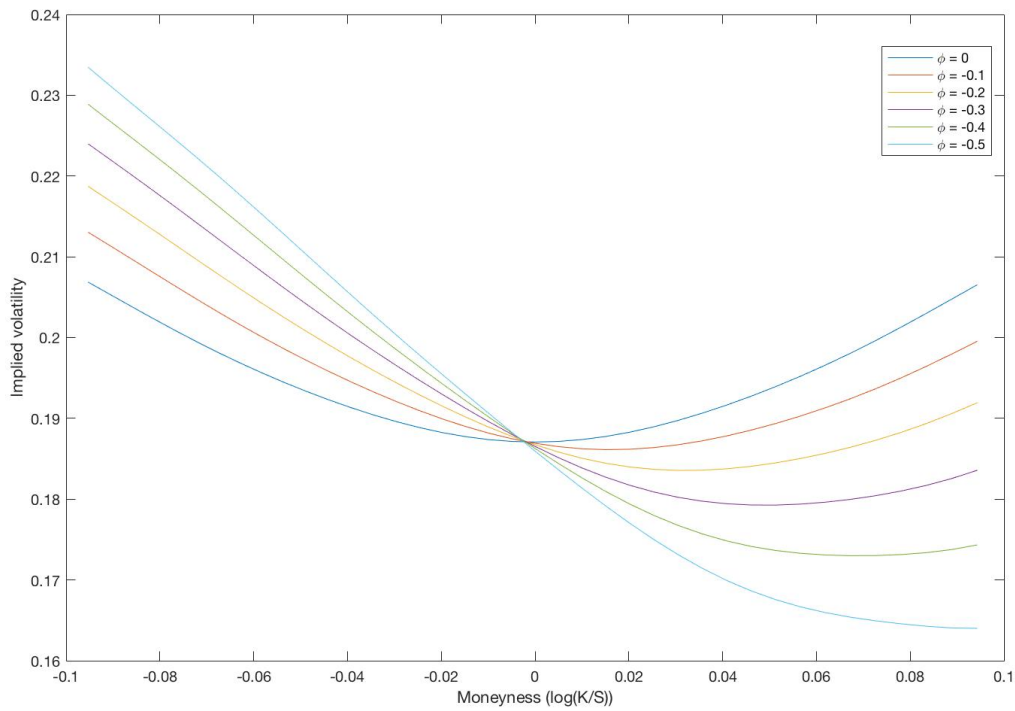
Romano, M. and N. Touzi (1997): "Contingent claims and market completeness in a stochastic volatility model," *Mathematical Finance*, 7, 399–410.

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

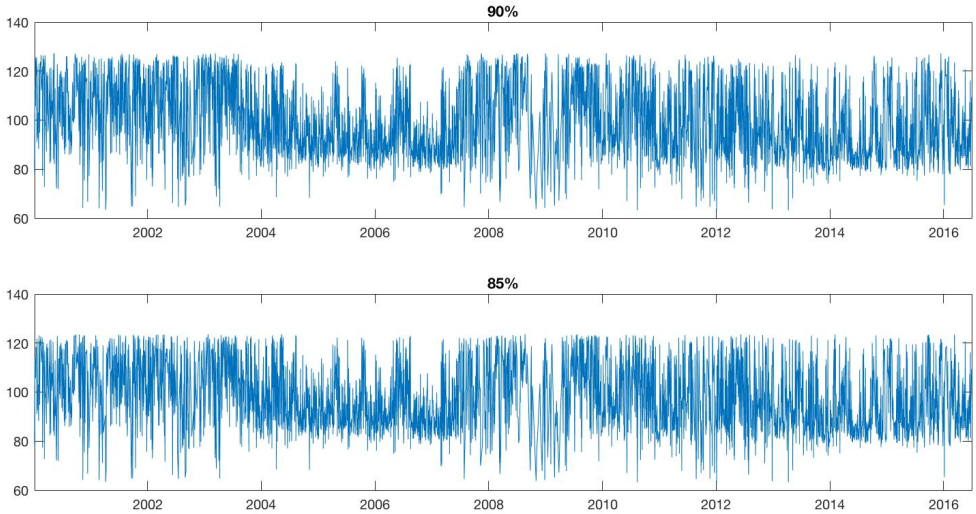
Appendix A

Figure 4: Volatility smiles from the SV model



** The implied volatilities are computed by inverting the model option price using the Black-Scholes formula where the model option prices are generated from the joint model of conditionally normal returns and ARG(1) volatility given in Section 4, with the risk-neutral parameters* $(\rho, \delta, c, e)' = (0.9, 1.1, 9.96e-6, 0)'$ *and different values of* ϕ*. The time to maturity is 30 days and the volatility* σ_t^2 *is set to be* $(0.2)^2/365$*.*

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

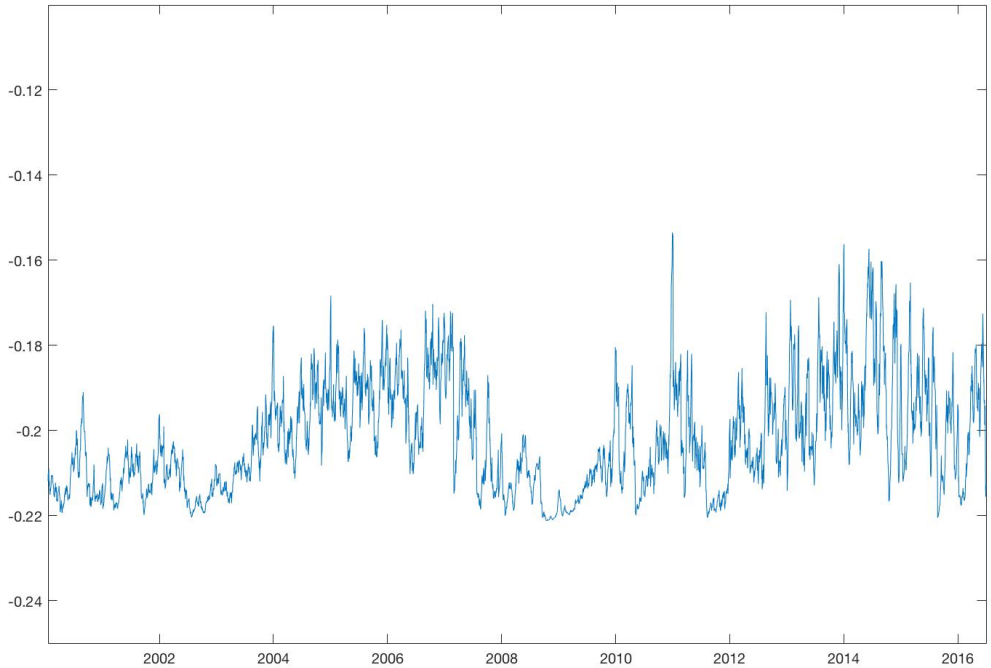


** We compute fitted values of the time series $E[RV_{t+1}|I(t)]$ and $Var[RV_{t+1}|I(t)]$ that are based on the estimation of an AR(1) model for the process RV_t with ARCH(1) innovations:*

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

on daily data over 16 years (realized variance of the S&P500 from January 2000 to June 2016).
** The first panel excludes the 5% largest and 5% smallest values.*
** The second panel excludes the 10% largest and 5% smallest values.*

Figure 6: AN model: Daily conditional correlation between returns and volatility, 2000-2016



** This graph plots the daily conditional correlation between returns and volatility from January 2000 to June 2016 generated from the AN model. The conditional correlations are computed using the formula given in* (4.4)*. In this version of the model,* f *and* k^2 *are not restricted to be* $(1+e)/2$ *and* $(1-e)/2c$*, respectively. The parameter estimates of* (ρ, δ, c, e, f) *are the same as the ones reported for ANTL in Table 1. The parameter estimates of* $(\phi, k^2, \varsigma_1, \varsigma_2) = (-0.0042, 5.79e+03, -4.7807, 1.1760)$*, which is sensitive to the initial values of* ς_1 *for the nonlinear estimation. However, the daily conditional correlation does not seem to depend on the parameter estimates of* $(\phi, k^2, \varsigma_1, \varsigma_2)$ *computed from different initial values.*

Appendix B

Proof of Proposition 2.2

1st step: We first prove an intermediary result that has its own interest:

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

Proof:

$$
\begin{aligned}
C_t(K) &= E^*[BS_{(t)}\left(K, \tilde{S}_t, \sigma^{*2}\left[J(t)\right]\right)|I(t)] \\
&\Longrightarrow \frac{\partial C_t(K)}{\partial \lambda_1} = E^*\left[\Phi\left[d_{1,t}\left(K, \sigma^{*2}[J(t)]\right)\right]\frac{\partial \tilde{S}_t}{\partial \lambda_1}|I(t)\right]
\end{aligned}
$$

with:

$$
\frac{\partial \tilde{S}_t}{\partial \lambda_1} = S_t \frac{\partial \xi_t}{\partial \lambda_1}
$$

and:

$$
\begin{aligned}
\xi_t &= \frac{\exp\left\{A\left[\lambda, J(t)\right]\right\}}{E^*[\exp\left\{A\left[\lambda, J(t)\right]\right\}|I(t)]} \\
&\Longrightarrow \frac{\partial \xi_t}{\partial \lambda_1} = \xi_t \tilde{Z}_t - \exp\left\{A\left[\lambda, J(t)\right]\right\}\frac{E^*[\tilde{Z}_t \exp\left\{A\left[\lambda, J(t)\right]\right\}|I(t)]}{\left\{E^*[\exp\left\{A\left[\lambda, J(t)\right]\right\}|I(t)]\right\}^2} \\
&= \xi_t\left[\tilde{Z}_t - \frac{E^*[\tilde{Z}_t \exp\left\{A\left[\lambda, J(t)\right]\right\}|I(t)]}{\left\{E^*[\exp\left\{A\left[\lambda, J(t)\right]\right\}|I(t)]\right\}}\right] \\
&\Longrightarrow \frac{\partial \xi_t}{\partial \lambda_1}(\lambda = 0) = \tilde{Z}_t - E^*[\tilde{Z}_t|I(t)] \\
&\Longrightarrow \frac{\partial C_t(K)}{\partial \lambda_1}(\lambda = 0) = S_t Cov^*[\tilde{Z}_t, \Phi\left(d_{1,t}\left(K, \sigma^{*2}\left[J(t)\right]\right)\right)|I(t)].
\end{aligned}
$$

2nd step: Sign of $\frac{\partial C_t(K)}{\partial \lambda_1}(\lambda = 0)$:

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

Hence:

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

Then:

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

with, given $I(t)$, both $\tilde{Z}_t$ and $\Phi\left(d_{1,t}\left(K, \sigma^{*2}\left[J(t)\right]\right)\right)$ are increasing functions of $\sigma^{*2}\left[J(t)\right]$ when $x_t(K) > -\frac{V}{2}$. Thus, for any out-of-the money call option ($x_t(K) > 0$), we can conclude that:

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

Moreover, the above formulas show that the more out-of-the money the call option (the larger

$x_t(K)$), the steeper is the slope of the option price as a function of λ_1.

Proof of Proposition 3.1.

For any real numbers u and v:

$$
\begin{aligned}
& E[M_{t+1}(\varsigma)\exp\left(-u\sigma_{t+1}^2 - vr_{t+1}\right)|I(t)] && (6.1)\\
= \; & M_{0,t}(\varsigma)\, E[\exp(-r_{f,t})\exp\left(-(\varsigma_1+u)\,\sigma_{t+1}^2 - (\varsigma_2+v)\,r_{t+1}\right)|I(t)] && (6.2)\\
= \; & \frac{M_{0,t}(\varsigma)\exp(-r_{f,t})}{M_{0,t}(\varsigma_1+u,\varsigma_2+v)} && (6.3)
\end{aligned}
$$

However, if $\varsigma = (\varsigma_1, \varsigma_2)$ characterizes the true value of the risk premium parameters, we also have by definition of the risk-neutral distribution:

$$
\begin{aligned}
& E[M_{t+1}(\varsigma)\exp\left(-u\sigma_{t+1}^2 - vr_{t+1}\right)|I(t)]\\
= \; & \exp(-r_{f,t})E^*[\exp\left(-u\sigma_{t+1}^2 - vr_{t+1}\right)|I(t)]\\
= \; & \exp(-r_{f,t}))E^*[\exp\left(-u\sigma_{t+1}^2\right)E^*[\exp(-vr_{t+1})\,|I^\sigma(t)]\,|I(t)]\\
= \; & \exp(-r_{f,t}))E^*[\exp\left(-u\sigma_{t+1}^2\right))\exp\left(-v\mu^*\left[I^\sigma(t)\right] + \frac{v^2}{2}\sigma^{*2}\left[I^\sigma(t)\right]\right)|I(t)]
\end{aligned}
$$

By identifying the two above formulas for (6.1) we get:

$$
\begin{aligned}
E^*[\exp\left(-u\sigma_{t+1}^2 - vr_{t+1}\right)|I(t)] &= \frac{M_{0,t}(\varsigma)}{M_{0,t}(\varsigma_1+u,\varsigma_2+v)}\\
&= \frac{\exp\left[l(\varsigma_1,\varsigma_2)\sigma_t^2 + g(\varsigma_1,\varsigma_2)\right]}{\exp\left[l(\varsigma_1+u,\varsigma_2+v)\sigma_t^2 + g(\varsigma_1+u,\varsigma_2+v)\right]}\\
&= \exp\left[-l^*(u,v)\sigma_t^2 - g^*(u,v)\right]
\end{aligned}
$$

This proves part (i) of the proposition.

Now we assume that $\mu^*[I^\sigma(t)]$ and $\sigma^{*2}[I^\sigma(t)]$ are affine in σ_{t+1}^2 and σ_t^2:

$$
\begin{aligned}
\mu^*[I^\sigma(t)] &= \omega_{1,1}\sigma_{t+1}^2 + \omega_{1,2}\sigma_t^2 + \omega_{1,3}\\
\sigma^{*2}[I^\sigma(t)] &= -\omega_{2,1}\sigma_{t+1}^2 - \omega_{2,2}\sigma_t^2 - \omega_{2,3}.
\end{aligned}
$$

for some constant $\omega_{i,j}, i = 1,2, j = 1,2,3$. We define quadratic functions $\alpha^*(\cdot)$, $\beta^*(\cdot)$, $\gamma^*(\cdot)$ such

that

$$\begin{aligned}
\alpha^*(v) &= \omega_{1,1}v + \frac{1}{2}\omega_{2,1}v^2 = \alpha^{*\prime}(0)v + \frac{1}{2}\alpha^{*\prime\prime}(0)v^2 \\
\beta^*(v) &= \omega_{1,2}v + \frac{1}{2}\omega_{2,2}v^2 = \beta^{*\prime}(0)v + \frac{1}{2}\beta^{*\prime\prime}(0)v^2 \\
\gamma^*(v) &= \omega_{1,3}v + \frac{1}{2}\omega_{2,3}v^2 = \gamma^{*\prime}(0)v + \frac{1}{2}\gamma^{*\prime\prime}(0)v^2.
\end{aligned}$$

Then,

$$\begin{aligned}
E^*[\exp\left(-u\sigma_{t+1}^2 - vr_{t+1}\right)|I(t)] &= E^*[\exp\left(-u\sigma_{t+1}^2\right)\exp\left(-v\mu^*\left[I^\sigma(t)\right] + \frac{v^2}{2}\sigma^{*2}\left[I^\sigma(t)\right]\right)|I(t)] \\
&= E^*\left[\exp\left[-\left(u + \omega_{1,1}v + \frac{1}{2}\omega_{2,1}v^2\right)\sigma_{t+1}^2\right]|I(t)\right] \times \\
&\exp\left[-\left(\omega_{1,2}v + \frac{1}{2}\omega_{2,2}v^2\right)\sigma_t^2 - \left(\omega_{1,3}v + \frac{1}{2}\omega_{2,3}v^2\right)\right] \\
&= \exp\left[-\beta^*(v)\sigma_t^2 - \gamma^*(v)\right]E^*\left[\exp\left[-\left(u + \alpha^*(v)\right)\sigma_{t+1}^2\right]|I(t)\right] \\
&= \exp\left[-\beta^*(v)\sigma_t^2 - \gamma^*(v)\right]\exp\left[-l^*\left(u + \alpha^*(v), 0\right)\sigma_t^2 - g^*(u + \alpha^*(v), 0)\right)\right] \\
&= \exp\left[-\left[l^*\left(u + \alpha^*(v), 0\right) + \beta^*(v)\right]\sigma_t^2 - \left[g^*(u + \alpha^*(v), 0) + \gamma^*(v)\right]\right] \\
&= \exp\left[-\left[a^*\left(u + \alpha^*(v)\right) + \beta^*(v)\right]\sigma_t^2 - \left[b^*(u + \alpha^*(v)) + \gamma^*(v)\right]\right].
\end{aligned}$$

We have the CAR models since:

$$\begin{aligned}
E^*\left[\exp\left(-u\sigma_{t+1}^2\right)|I(t)\right] &= \exp\left[-l^*(u,0)\sigma_t^2 - g^*(u,0)\right] \\
&= \exp\left[-a^*(u)\sigma_t^2 - b^*(u)\right] \\
E^*\left[\exp\left(-vr_{t+1}\right)|I^\sigma(t)\right] &= \exp\left(-v\mu^*\left[I^\sigma(t)\right] + \frac{v^2}{2}\sigma^{*2}\left[I^\sigma(t)\right]\right) \\
&= \exp\left[-\alpha^*(v)\sigma_{t+1}^2 - \beta^*(v)\sigma_t^2 - \gamma^*(v)\right].
\end{aligned}$$

This proves the second part of the proposition.

Proof of Corollary 3.3.

We have, by the CAR model given in proposition 3.1.,

$$\begin{aligned}
E^*\left[\exp(r_{t+1})|I(t)\right] &= \exp\left[-l^*(0,-1)\sigma_t^2 - g^*(0,-1)\right] \\
&= \exp\left[-\left[a^*(\alpha^*(-1)) + \beta^*(-1)\right]\sigma_t^2 - \left[b^*(\alpha^*(-1)) + \gamma^*(-1)\right]\right]
\end{aligned}$$

Thus, the risk-neutral pricing of the underlying asset:

$$E^* \left[\exp(r_{t+1})|I(t)\right] = 1,$$

is equivalent to

$$a^*(\alpha^*(-1)) + \beta^*(-1) = 0 \text{ and } b^*(\alpha^*(-1)) + \gamma^*(-1).$$

Proof of Proposition 3.2.

We have first from Proposition 3.1.:

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

Let us admit for the moment that , as checked below:

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

Then we deduce from above that:

$$l\left(u,v\right) = a\left[u+\alpha^*\left(v-\varsigma_2\right)+\alpha\left(\varsigma_2\right)\right] - a\left(\varsigma_1+\alpha\left(\varsigma_2\right)\right) + \beta^*\left(v-\varsigma_2\right) + l\left(\varsigma_1,\varsigma_2\right)$$

Then, if we define functions $\alpha(.), \beta(.), \gamma(.)$ as in Proposition 3.2, we first note that these functions are quadratic (since the functions $\alpha^*(.), \beta^*(.), \gamma^*(.)$ are quadratic) and we can rewrite $l(u,v)$ as :

$$l\left(u,v\right) = a\left[u+\alpha(v)+\alpha^*\left(-\varsigma_2\right)+\alpha\left(\varsigma_2\right)\right] - a\left(\varsigma_1+\alpha\left(\varsigma_2\right)\right) + \beta(v) + \beta^*\left(-\varsigma_2\right) + l\left(\varsigma_1,\varsigma_2\right)$$

However, we have by definition:

$$\begin{aligned} \alpha\left(\varsigma_2\right) &= \alpha^*\left(\varsigma_2-\varsigma_2\right) - \alpha^*\left(-\varsigma_2\right) = -\alpha^*\left(-\varsigma_2\right) \\ \beta\left(\varsigma_2\right) &= \beta^*\left(\varsigma_2-\varsigma_2\right) - \beta^*\left(-\varsigma_2\right) = -\beta^*\left(-\varsigma_2\right) \end{aligned}$$

Therefore:

$$l\left(u,v\right) = a\left[u+\alpha(v)\right] - a\left(\varsigma_1+\alpha\left(\varsigma_2\right)\right) + \beta(v) - \beta\left(\varsigma_2\right) + l\left(\varsigma_1,\varsigma_2\right)$$

which is consistent with:

$$\begin{aligned} l(u,v) &= a[u+\alpha(v)]+\beta(v) \\ l(\varsigma_1,\varsigma_2) &= a(\varsigma_1+\alpha(\varsigma_2))+\beta(\varsigma_2) \end{aligned}$$

By symmetry, we can obviously prove in an analogous way the formula of Proposition 3.2. for $g(u,v)$.

It remains to check that these formulas are consistent with the formulas announced for $a^*(u)$ and $b^*(u)$. We have:

$$\begin{aligned} a^*(u) &= l^*(u,0) = l(u+\varsigma_1,\varsigma_2) - l(\varsigma_1,\varsigma_2) \\ &= a[u+\varsigma_1+\alpha(\varsigma_2)]+\beta(\varsigma_2) - a[\varsigma_1+\alpha(\varsigma_2)]-\beta(\varsigma_2) \\ &= a[u+\varsigma_1+\alpha(\varsigma_2)] - a[\varsigma_1+\alpha(\varsigma_2)] \end{aligned}$$

By symmetry, we can obviously prove in an analogous way the formula of Proposition 3.2. for $b^*(u)$.

Appendix C: Heston model as a continuous time limit

We want to check that the affine specification introduced in Section 3 and 4 for the joint dynamics of $\left(r_{t+1},\sigma^2_{t+1}\right)$ 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^2_t for its first two conditional moments

$$\begin{aligned} E[\sigma^2_{t+1}|I(t)] &= \omega+\rho\sigma^2_t \\ Var[\sigma^2_{t+1}|I(t)] &= \bar{\omega}+\bar{\rho}\sigma^2_t \end{aligned} \tag{6.4}$$

obviously amounts to a vector auto-regressive VAR(1) specification for the bivariate process $\left(\sigma^2_t,\sigma^4_t\right)$ 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:

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

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

For the sake of getting the instantaneous analog of $\sigma^2_{t,H}(N)$, 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 C.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 C.2: (Continuous-time limit) If the equations in Lemma C.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 $\tilde{I}(t) = \left\{\sigma^2_{t-kH,H}(N), k \geq 1\right\}$.

The proof is given below. 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 $\omega = 2\bar{\omega}(\rho/\bar{\rho})$ as in our example of the ARG(1) volatility model, 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[12], $\rho, \bar{\omega}$, 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, if we consider a general affine volatility model, 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 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.

Proof

Before we prove Proposition C.2, we first provide and prove Proposition C.1 below.

Proposition C.1:

The volatility factor $\sigma^2_{t,H}(N)$ given above 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 (6.5), (6.16), and (6.17),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 C.1

Everywhere below we use the following notation:

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

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

for any random variable X.

From the first moment of volatility (see Section 3.2) 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

$$\begin{aligned} E_t[\sigma_{t+H}^2] &= \rho^H E_t[\sigma_t^2] + \omega(1+\rho+\cdots+\rho^{H-1}) \\ &= \rho^H E_t[\sigma_t^2] + \omega\frac{1-\rho^H}{1-\rho} \\ &= \rho^H E_t[\sigma_t^2] + \omega(H) \end{aligned} \tag{6.5}$$

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 3.2) 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

$$\begin{aligned} a &= 2\rho\omega + \bar{\rho} \\ b &= \omega^2 + \bar{\omega}, \end{aligned} \tag{6.6}$$

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 (6.5) 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

$$C_1 = \omega\sum_{h=0}^{H-1}\rho^{2(H-1-h)}\frac{1-\rho^H}{1-\rho} = \omega\frac{1-\rho^{H-1}(1-\rho^H)}{(1-\rho)(1-\rho)^2} = \omega(H)\frac{1-\rho^{H-1}}{1-\rho}. \tag{6.7}$$

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

$$C_2 = aC_1 + b\sum_{h=0}^{H-1}\rho^{2h} = aC_1 + b\frac{1-\rho^{2H}}{1-\rho^2}. \tag{6.8}$$

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

$$E_t\left[(1-\rho^{2H}L^H)\sigma_{t+H+h}^4\right] = a\rho^{H-1}\frac{1-\rho^H}{1-\rho}E_t[\sigma_{t+h}^2] + C_2. \tag{6.9}$$

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

$$E_t\left[(1-\rho^{2H}L^H)\frac{1}{HN}\sum_{n=0}^{HN}\sigma^4_{t+H+n/N}\right] = a\rho^{H-1}\frac{1-\rho^H}{1-\rho}E_t\left[\sigma^2_{t,H}(N)\right] + C_2. \tag{6.10}$$

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

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

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

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

Making the corresponding substitution in (6.10) and dividing by HN, we can write

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

where

$$a_0(H;N) = \frac{1}{HN}a\rho^{H-1}\frac{1-\rho^H}{1-\rho}. \tag{6.11}$$

By the law of iterated expectations and (6.5), 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 (6.9) 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 (6.5) 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

$$\begin{aligned}E_t\left[\left(1-\rho^{2H}L^H\right)\sigma^2_{t+H+n/N}\sigma^2_{t+H+(n+j)/N}\right] &= \rho^{j/N}E_t\left[\left(1-\rho^{2H}L^H\right)\sigma^4_{t+H+n/N}\right]\\ &\quad+\omega(j/N)E_t\left[\left(1-\rho^{2H}L^H\right)\sigma^2_{t+H+n/N}\right]\\ &= \left[a\rho^{j/N}\rho^{H-1}\frac{1-\rho^H}{1-\rho}+\omega(j/N)\rho^H(1-\rho^H)\right]E_t\left[\sigma^2_{t+n/N}\right]\\ &\quad+C_3(j),\end{aligned}\tag{6.12}$$

where

$$\begin{aligned}C_3(j) &= \rho^{j/N}C_2+\omega(j/N)\omega(H)\\ &= \rho^{j/N}\left[a\omega(H)\frac{1-\rho^{H-1}}{1-\rho}+b\frac{1-\rho^{2H}}{1-\rho^2}\right]+\omega^2\frac{1-\rho^{j/N}}{1-\rho}\frac{1-\rho^H}{1-\rho}.\end{aligned}\tag{6.13}$$

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 (6.5) 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

$$C_4 = \frac{1}{HN}\sum_{n=1}^{HN}\omega(n/N) = \frac{\omega}{HN}\frac{(1-\rho^{1/N}) - \rho^{1/N}(1-\rho^H)}{(1-\rho)(1-\rho^{1/N})}. \tag{6.14}$$

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

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

and

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

Substituting this result to the expression (6.12) for the cross-terms, we obtain that

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

is equal to

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

where

$$\begin{aligned} C_5(j,n) = &\left(a\rho^{j/N}\rho^{H-1}\frac{1-\rho^H}{1-\rho} + \omega(j/N)\rho^H(1-\rho^H)\right) \\ &\times\left(\omega(n/N) - HN\rho^{(n-1)/N}\frac{1-\rho^{1/N}}{1-\rho^H}C_4\right) + C_3(j). \end{aligned} \tag{6.15}$$

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

$$\begin{aligned} a(H;N) &= a_0(H;N) \\ &+ \frac{2}{HN}\frac{\rho^H}{1-\rho^H}(1-\rho^{1/N})\sum_{j=1}^{HN-1}\sum_{n=1}^{HN-j}\left(\rho^{j/N}\frac{a}{\rho}\frac{1-\rho^H}{1-\rho} + \omega(H)(1-\rho^{n/N})\right)\rho^{(n-1)/N}, \end{aligned} \tag{6.16}$$

and

$$b(H;N) = \frac{2}{H^2N^2} \sum_{j=1}^{HN-1} \sum_{n=1}^{HN-j} C_5(j,n) \tag{6.17}$$

with $a_0(H;N)$ defined above in (6.11), $\omega(H)$ defined in (6.5), the coefficients C_1 through C_5 defined in (6.7), (6.8), (6.13), (6.14), and (6.15).

Proof of Proposition C.2

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

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

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 C.2.

For the second part, we see from the definition of variance and the equation (6.5) that the conditional moment restrictions given in Proposition C.1 are the same as

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

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 C.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 (6.11) and (6.6) . 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

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

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

$$\begin{aligned} \lim_{N\to\infty} a(H;N) = &-\frac{2}{H}\frac{a}{\rho}\frac{\rho^H}{1-\rho}\left(\frac{1-\rho^H}{log(\rho)} + \rho^H H\right) \\ &+ 2\frac{\rho^H}{1-\rho^H}\frac{\omega(H)}{H}\left(H + \frac{1-\rho^H}{log(\rho)}\right) \\ &- \frac{\rho^H}{1-\rho^H}\frac{\omega(H)}{H}\left(H + \frac{1-\rho^{2H}}{log(\rho^2)}\right). \end{aligned} \tag{6.18}$$

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 (6.18) 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[13]

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

with b defined in (6.6).

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

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

In case of affine first two moments as in (6.4), we have

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

[13]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.

Hence, the limit becomes

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

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

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

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

Appendix D: Choice of instruments and identification

While the GMM estimation is based on realized variance data as we see in Section 5, 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:

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

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 some finite sample bias-variance trade off.

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

$$\left[\frac{\partial\Psi'_{t,\theta_\sigma}(u_1)}{\partial\theta}, ..., \frac{\partial\Psi'_{t,\theta_\sigma}(u_J)}{\partial\theta}\right]\Sigma_{t,\theta_\sigma}^{-1}\left(u_1, ..., u_J\right) \tag{6.19}$$

where $\Sigma_{t,\theta_\sigma}^{-1}\left(u_1, ..., u_J\right)$ 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 (proof given below) that:

Proposition D.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.

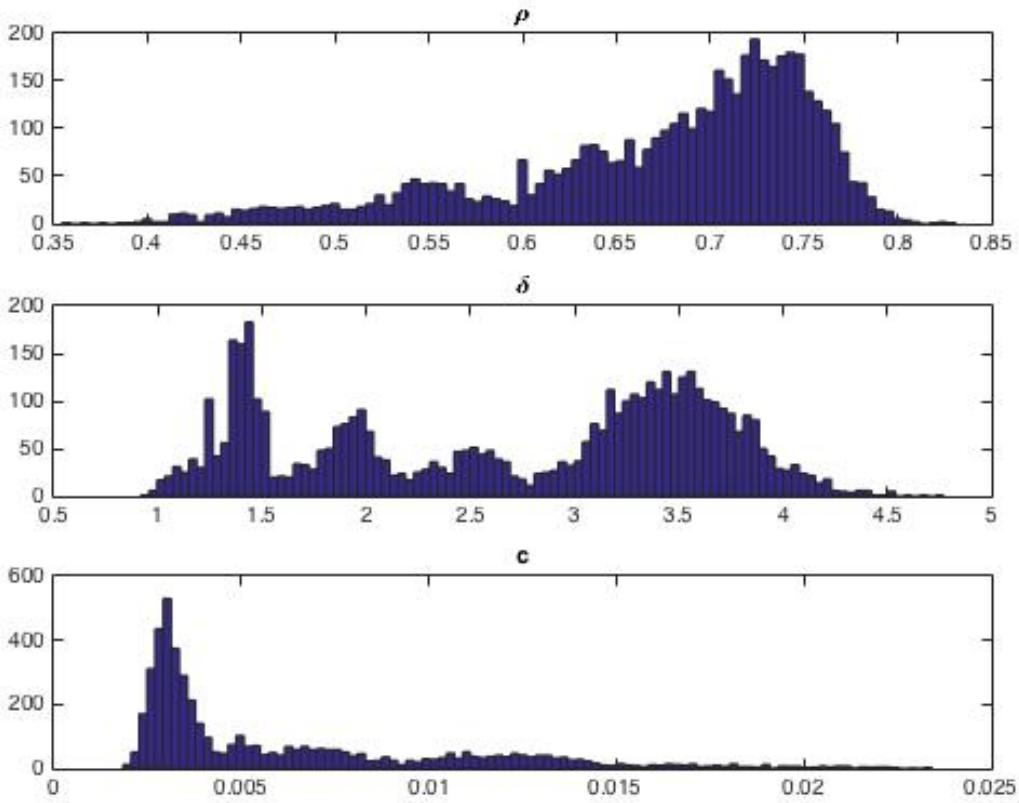
In spite of the positive result of Proposition D.1., we do not expect very accurate GMM estimators since the above unconditional moment conditions are only about the marginal distribution of σ_t^2 and $\left(\sigma_t^2+\sigma_{t+1}^2\right)$. While the marginal distribution of σ_t^2 is of course unable 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 7 and 8 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 (6.19) that each optimal instrument should include all the indices $u_j, j=1,...,J$.

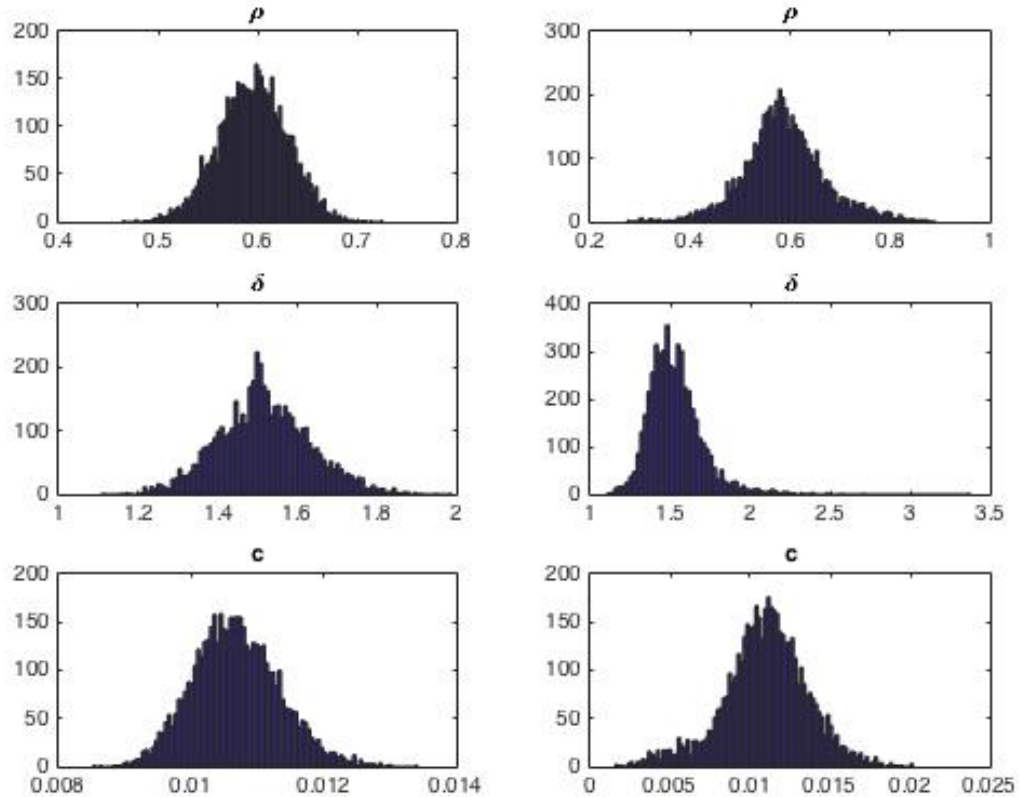
Figures

Figure 7: 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* $[1i, 10i]$ *on which the conditional characteristic function and the instrument are evaluated.*
** An identity is used as a weighting matrix.*
** 10 randomly generated values were used as initial values for each* ρ, δ, *and* c.
** We used 5000 replications.*

Figure 8: 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 $[1i, 10i]$ on which the conditional characteristic function.*
** Let v be a complex number on which the DI-instrument is evaluated. The right hand side panel used $v = 1i$ and the left hand side one used $v = 1i$ and $v = 10i$.*
** An identity is used as a weighting matrix.*
** 10 randomly generated values were used as initial values for each ρ, δ, and c.*
** We used 5000 replications*

Proof of Proposition D.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:

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

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

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

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:

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

Therefore:

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