Identifying the Volatility Risk Price Through the Leverage Effect

Xu Cheng*, Eric Renault†, Paul Sangrey‡

This Version: October 29, 2023

Abstract

In asset pricing models with stochastic volatility, uncertainty about volatility affects risk premia through two channels: aversion to decreasing returns and aversion to increasing volatility. We analyze the identification of and robust inference for structural parameters measuring investors' aversions to these risks: the return risk price and the volatility risk price. We show that the leverage effect (instantaneous causality between the asset return and its volatility) implies that the variance risk premium does not identify the volatility risk price, but that price data can identify the volatility risk price without additional options data. We analyze this identification challenge in a nonparametric discrete-time exponentially affine model, complementing the continuous-time approach of Bandi and Renò (2016). We then specialize to a parametric model and derive the implied minimum distance criterion relating the risk prices to the asset return and volatility's joint distribution. This criterion is almost flat when the leverage effect is small, and we introduce identification-robust confidence sets for both risk prices regardless of the magnitude of the leverage effect.

Keywords: leverage effect, nonparametric identification, stochastic volatility, volatility factor, volatility risk price, weak identification
JEL Classification: C12, C14, C38, C58, G12.

*University of Pennsylvania, Department of Economics, 133 South 36th Street, Philadelphia, PA 19104; Email: xucheng@upenn.edu

†University of Warwick, Department of Economics, Coventry, West Midlands, CV4 7AL, UK; Email: eric.renault@warwick.ac.uk

‡Noom Inc, 450 W 33rd St, New York, NY, 10001; Email: paul@sangrey.io

1 Introduction

Risk aversion functions extracted from observed stock and option prices can be negative as shown by Ait-Sahalia and Lo (2000) and Jackwerth (2015). While this phenomenon was initially viewed as a puzzle, Chabi-Yo, Garcia, and Renault (2008) rationalize it by a lack of conditioning on state variables. They further argue that the presence of state variables may explain the so-called pricing kernel puzzle, i.e., the pricing kernel, when projected onto prices, often exhibits a U-shape pattern rather than the expected decreasing shape. This so-called "nonmonotonic pricing kernel" is now the focus of a large strand of literature.

Chabi-Yo, Garcia, and Renault (2008) show that conditioning on state variables encompasses many explanations put forward in the extant literature.[1] In particular, Christoffersen, Heston, and Jacobs (2013) promote a "variance-dependent pricing kernel" to reconcile the time-series properties of stock returns with the cross-section of option prices. This approach amounts to using the conditional variance of the return as a specific state variable in a discrete time context. Bandi and Renò (2016) develop a continuous time version of this model where the volatility factor is an increasing function of the spot volatility process in the context of a jump diffusion model of return and volatility. These models all have a Stochastic Discount Factor (SDF) depending on two risk factors – future return and future volatility, whose risk compensation are proportional to the risk aversion coefficients ζ_r and ζ_σ, respectively.

The starting point of this paper is that, to interpret asset prices based on two risk factors, a key concept is the instantaneous causality between these two factors in the sense of Pierce and Haugh (1977). When the two factors are the future asset return and its volatility, as in this paper, the instantaneous causality relationship is usually dubbed the "leverage effect" in the sense of Black (1976). The main theoretical message of this paper is on the impact of the leverage effect on the accuracy of the common belief that identification of risk aversion to volatility of volatility must be based on observations from the derivative markets. We recall that the variance risk premium is measured as the difference between the risk-neutralized expected return variation and the realized return variation. Bollerslev, Tauchen, and Zhou (2009) refer to the work of Britten-Jones and Neuberger (2000) to measure the market's risk neutral expectation of the total return variation between time t and $t+1$ conditional on the time t information from a large portfolio of European calls. They note that this model-free measure provides a natural empirical analog to the risk-neutral expectation of future return variation.

This paper stresses that the presence of the leverage effect reverses both sides of the

[1] Examples include heterogeneity of beliefs following regime shifts (Ziegler, 2007), state-dependent preferences (Melino and Yang, 2003), and external habits with state dependence in beliefs (Veronesi, 2001).

common belief that identification of risk aversion to volatility of volatility must be based on observations from the derivative markets. On the negative side, we argue that the variance risk premium does not unambiguously identify the risk aversion to volatility variation. Due to the leverage effect, the volatility factor on the time interval $[t, t+1]$ is conditionally correlated with the return itself, given the time t information. Therefore, the variance risk premium combines the effect of two risk aversion parameters: $\zeta_r > 0$ for aversion to negative variation in the return and $\zeta_\sigma < 0$ for aversion to positive variation in the volatility. On the positive side, we show that, precisely because the volatility factor on the time interval $[t, t+1]$ is correlated with the return itself conditional on the time t information, observations of the return alone, without any observation from the derivative markets, are sufficient to identify both risk aversion parameters. The asset pricing equation in the presence of a leverage effect depends on both risk aversion parameters, which allows econometric identification of these parameters from the joint distribution of the return and the volatility.

We propose a valid inferential methodology to materialize the positive side of the message above: learning both risk aversion parameters with observations of the return. This is challenging for the following reason: On the one hand, although we expect the leverage effect to be negative, it is difficult to quantify empirically and its estimate usually is small (Aït-Sahalia, Fan, and Li, 2013). On the other hand, in the absence of a leverage effect, risk aversion to the volatility of volatility is not identified from the time series of the asset only. Therefore, when the true leverage effect is small, the return data alone provide a limited amount of information about the aversion to the volatility risk, compared to the finite-sample noise in the data. This low signal-to-noise ratio, as modeled by (nearly) weak identification, may invalidate standard inference based on the Generalized Method of Moments (GMM) estimator, see Stock and Wright (2000), Andrews and Cheng (2012), Antoine and Renault (2012).

We provide an identification-robust confidence set for the structural parameters that measure the return risk price, the volatility risk price, and the leverage effect. The robust confidence set provides correct asymptotic coverage, uniformly over a large set of models and allows for any magnitude of the leverage effect. This uniform validity is crucial for the confidence set to have good finite-sample coverage (Mikusheva, 2007; Andrews and Guggenberger, 2010). In contrast, standard confidence sets based on the GMM estimator and its asymptotic normality do not have uniform validity in the presence of a small leverage effect. This issue affects all the structural parameters because they are estimated simultaneously.

We construct the robust confidence set in two steps. First, we establish a minimum distance criterion using link functions between the structural parameters and a set of reduced-form parameters that determine the joint distribution of the return and volatility. The structural model implies that the link functions are zero when evaluated at the true values of the structural parameters and the reduced-form parameters. Identification and estimation of these reduced form parameters are standard and are not affected by the presence of a small leverage effect. However, the link functions are almost flat in some structural parameters when the leverage effect is small, resulting in weak identification. Second, given this minimum distance criterion, we invert the conditional quasi-likelihood ratio (QLR) test of Andrews and Mikusheva (2016) to construct a robust confidence set. The key feature of this test is that it treats the flat link functions as an infinite-dimensional nuisance parameter. The critical value is constructed by conditioning on a sufficient statistic for this nuisance parameter, and it is known to yield a valid test regardless of the nuisance parameter's value. Andrews and Mikusheva (2016) develop this test in a GMM framework. We show it works in minimum distance contexts such as the one considered here and provide conditions for its asymptotic validity. For practitioners, we provide a detailed algorithm for the construction of this simulation-based robust confidence set. It is worth noting that that this strategy of identification-robust inference in a minimum-distance framework may be applied in other contexts such as Magnusson (2010) and Magnusson and Mavroeidis (2010).

One may question the relevance of this robust inference method given that a natural solution to the low signal-to-noise issue in returns is to use derivatives data. We follow in this respect a research agenda first put forward by Bandi and Renò (2016). As they compellingly show, the benefit of using the stock return data alone and not resorting to option prices is robustness to misspecification. Bandi and Renò (2016) emphasize that their result "does not hinge on sudden changes in risk premia associated with market downturns (as possibly yielded by the use of the VIX) or implied volatility smirks (as given by cross-sectional option prices). Said differently, the effect is solely revealed by the dynamic properties of stock prices, once volatility is filtered effectively and a sufficiently rich specification is adopted, without the need for the, arguably economically confounding (due to risk premia), information contained in traded or synthetic options."

Our asset pricing framework is germane to Bandi and Renò (2016) with a couple of differences. First, like Bandi and Renò (2016), we resort to a variance-dependent exponentially affine SDF, while discussing identification in a nonparametric setting for both the historical and the risk-neutral distributions. While Bandi and Renò (2016) remain nonparametric in the context of a jump diffusion model of return and volatility, we choose to work in discrete time. An advantage is that our filtered value of volatility, based on high frequency data like that in Bandi and Renò (2016), is much less noisy. Our estimate is based on the average value of the instantaneous variance over a day, whereas Bandi and Renò (2016) estimate the spot volatility process. It is also worth noting that, when the time interval between observations becomes infinitesimal, one may argue that our discrete time empirical model converges to a diffusion model(see Han, Khrapov, and Renault, 2020) germane to Bandi and Renò (2016).

Second, Duffie, Pan, and Singleton (2000) have shown that in the context of an exponentially affine SDF, derivative asset prices can be computed from "extended transforms", meaning conditional Laplace transforms of the joint vector of payoffs and state variables at some horizon, given the information available at time t. We extend Duffie, Pan, and Singleton (2000) to nonparametric historical and risk-neutral conditional Laplace transforms of the pair of daily return and filtered volatility. As such, our general identification result with the unrestricted Laplace transforms generalizes that in Bandi and Renò (2016) on a jump diffusion process. Furthermore, we define a general concept of "local zero-leverage", more general than zero-leverage, and show that it is sufficient to impair the identification of the volatility risk aversion.

As in Bandi and Renò (2016), we illustrate our general nonparametric identification result by estimating a specific parametric model. Our empirical model is an extension of that in Corsi, Fusari, and La Vecchia (2013). The proposed extension is motivated by the need to properly accommodate the leverage effect as well as the need to illustrate our identification statement. We parameterize the impact of the leverage effect in order to capture the implications of the usual continuous time models in the literature of option pricing with stochastic volatility. On the one hand, as in the continuous time models, knowledge of the contemporaneous volatility innovation reduces the conditional variance of the return innovation since the two innovations are correlated, i.e., the instantaneous causality effect. On the other hand, this also introduces an additional term in the return drift because knowledge of the future volatility has an impact on the expected return. As stressed by Bollerslev, Litvinova, and Tauchen (2006), this effect of leverage on the expected return is hard to disentangle in discrete time from the volatility feedback effect due to the risk compensation. Bollerslev, Litvinova, and Tauchen (2006) claim that only continuous time observations would allow us to clearly disentangle the two effects by detecting the direction of causality. This potential identification issue is carefully taken into account in the definition of our three structural parameters: the two risk aversion parameters, and the leverage effect parameter that is identified in the risk neutral world.

Our empirical model also extends Corsi, Fusari, and La Vecchia (2013) by introducing additional state variables for identification. By introducing additional state variables in the conditional mean and variance of the conditionally Gaussian log-return, we obtain identification of the two risk aversion parameters by only using observations of the equity return (jointly with its realized variance), in view of the non-arbitrage condition equilibrium. In contrast, Corsi, Fusari, and La Vecchia (2013) acknowledge that in their model the parameter of risk aversion for the volatility of volatility "must be calibrated".

The rest of the paper is organized as follows. Section 2 discusses theoretical identification in a model-free framework where the conditional probability distributions of interest, both in the historical and risk-neutral worlds, are characterized by their conditional Laplace transforms. In section 3, we specify a general exponentially affine pricing model for the joint historical distribution of the return and the volatility factor. Our bivariate discrete time model is inspired by the continuous time model of Heston (1993) and belongs to the general class of Compound AutoRegressive Models (CAR) of Darolles, Gouriéroux, and Jasiak (2006). Although we assume a CAR structure, we remain in this section nonparametric about the conditional Laplace transforms that define the conditional distribution of the volatility factor given the past, and the conditional Laplace transforms that define the current return given the current volatility factor and the past information. Section 4 studies the empirical model. We introduce a dynamic model for the volatility factor and specify a parametric model for the joint conditional distribution of the return and the volatility factor given the past. We also provide the link function between the reduce form parameters that characterize the joint conditional distribution of the return and the volatility factor and the structural parameters that characterize the risk aversions and the leverage effect. Section 5 provides an identification-robust inference method for the structural parameters. The proposed confidence set is uniformly valid regardless of the magnitude of the leverage effect. In particular, it is robust to a small leverage effect. Section 6 and Section 7 provide Monte Carlo simulation results and an empirical application. Section 8 concludes.

2 General Framework

In this section, we consider general conditional Laplace transforms that describes the joint distribution of an asset return and a volatility factor. The historical and risk-neutral distributions are connected by an exponentially affine SDF. We show that, as in continuous time with the Girsanov theorem, the change of measure provided by the exponentially affine SDF is leverage effect preserving: there is instantaneous causality between the two variables in the risk-neutral world if and only if this causality exists in the historical world. We prove our identification claims in this general setting. We also define a general concept of "local zero-leverage", more general than zero-leverage, and show that it is sufficient to impair the identification of volatility risk aversion.

2.1 Variance-Dependent Pricing Kernel

We specify a SDF that characterizes the compensation both for the risk on equity (with the risk price parameter ζ_r) and for the uncertainty of the volatility factor (with the risk price parameter ζ_σ) implied by the stochastic volatility process between dates t and $t+1$. The general study of affine option pricing by Pan (2002) (see formula (A1) on page 34 of Pan, 2002, for the state-price density) shows that this SDF must compensate two kinds of risk: The first is the risk on equity as carried by the future random return r_{t+1}. The second is the risk on volatility as carried by what will be expected at time $t+1$ about future integrated variance $\int_{t+1}^{T} \sigma_u^2 du, T > t+1$.

For econometric identification of these two kinds of risk, two approaches are sensible. The first approach is to consider a framework where the time to maturity is infinitely small. We might then consider extensions of the nonparametric infinitesimal method of moments proposed by Bandi and Renò (2016). The alternative approach is to choose some horizon and work in discrete time, as in the present paper. Our model is essentially robust to time aggregation, and it can be interpreted through its continuous time limit. The cost of focusing on a discrete time version is limited. Moreover, the discrete time approach is better suited to address the weak identification issue we are interested in, without further complications due to the nonparametric rates of convergence for estimators of the spot volatility. As acknowledged by Bandi and Renò (2016), "variance measures integrated over a longer horizon are expected to be relatively less noisy than less integrated ones". We consider hereafter (see the next subsection) that the relevant volatility factor μ_t is produced by an optimal forecast of the daily integrated variance, which is accurately estimated with high frequency data.

W consider the following SDF:

$$M_{t+1}(\zeta) = \exp(-r_{f,t}) M_{0,t}(\zeta) \exp\left\{-\zeta_r r_{t+1} - \zeta_\sigma \mu_{t+1}\right\}, \tag{2.1}$$

where $r_{f,t}$ stands for the continuously compounded risk free interest rate between t and $t+1$, r_{t+1} is the stock log-return in excess of the risk-free rate $r_{f,t}$, and μ_{t+1} is a volatility factor defined more precisely in the next subsection. Our general model in (2.1) can be interpreted in both continuous time and discrete time. Christoffersen, Heston, and Jacobs (2013) use a Heston and Nandi (2000) GARCH process for a discrete time approximation of the continuous time model. However, this approximation cannot be exact since the GARCH model is known to be non-robust to temporal aggregation. In contrast, we employ the affine stochastic volatility (SV) model, which has been shown (see Meddahi and Renault, 2004) to be the SV version of weak-GARCH that is robust to temporal aggregation.

Several authors (see e.g., Chernov, Gallant, Ghysels, and Tauchen, 2003; Christoffersen, Heston, and Jacobs, 2009) have emphasized the importance of a multifactor variance specification, meaning that the aforementioned function of volatility in the pricing kernel would be a sum of various Markov factors potentially correlated with the innovation of the return process. Since the focus of this paper is the leverage effect, i.e., the instantaneous correlation between the return process and the variance process, the multifactor approach may be useful to generate stochastic correlation between the return and the volatility, due to the fact that the factors have different degrees of correlation with the return process and the weights of the different factors vary over time. As in Bandi and Renò (2016), we avoid the statistical challenge to estimate a multifactor model by assuming (see section 2.3. below) that what matters for the variance-dependent SDF specification is a volatility factor computed from high-frequency data.

2.2 Conditional Laplace Transforms

In this section, we use the Laplace transforms to characterize the relationship between the historical and risk-neutral conditional distributions at time $t+h$ given information available at time t with the SDF (2.1) that bridges the gap between the two distributions.

Let $I(t)$ stand for the information available at time t, which contains at least the past and present observations of the joint process $W_t = (r_t, \mu_t)$ such that $\{W_\tau, M_{0,\tau}(\zeta), \tau \leq t\} \subset I(t)$. The historical joint dynamics of the process W_t is defined by the conditional Laplace transform $\mathcal{L}$:

$$\mathcal{L}_t(u, v) = E[\exp\left(-ur_{t+1} - v\mu_{t+1}\right)|I(t)], \forall u \in \mathbb{R}, v \in \mathbb{R}.$$

Plugging the formula (2.1) for $M_{t+1}(\zeta)$ into this conditional Laplace transform, we note that the non-arbitrage condition

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

is tantamount to

$$M_{0,t}(\zeta) = \frac{1}{\mathcal{L}_t(\zeta_r, \zeta_\sigma)}.$$

Therefore, the SDF can be written as

$$M_{t+1}(\zeta) = \frac{\exp\left(-r_{f,t}\right)\exp\left\{-\zeta_r r_{t+1} - \zeta_\sigma \mu_{t+1}\right\}}{\mathcal{L}_t(\zeta_r, \zeta_\sigma)}. \tag{2.2}$$

The risk neutral joint dynamics of the process W_t is defined by the conditional Laplace transform $\mathcal{L}^*$:

$$\mathcal{L}_t^*(u,v) = E^*[\exp\left(-ur_{t+1} - v\mu_{t+1}\right)|I(t)]\,.$$

The bridge between the two, historical and risk-neutral, conditional distributions is given by the SDF $M_{t+1}(\zeta)$ through the identity

$$\exp(-r_{f,t})E^*[\exp\left(-ur_{t+1} - v\mu_{t+1}\right)|I(t)] = E[M_{t+1}(\zeta)\exp\left(-ur_{t+1} - v\mu_{t+1}\right)|I(t)]\,.$$

By plugging in the formula (2.2) for the SDF $M_{t+1}(\zeta)$, we obtain

$$\mathcal{L}_t^*(u,v) = \frac{\mathcal{L}_t(u+\zeta_r, v+\zeta_\sigma)}{\mathcal{L}_t(\zeta_r,\zeta_\sigma)}. \tag{2.3}$$

Formula (2.3) is the fundamental relationship between the risk-neutral and historical distributions that allows us to discuss identification of the risk aversion parameters ζ_r and ζ_σ from the historical distribution when assuming that the risk neutral distribution does not depend on these parameters.

Moreover, as enhanced by Duffie, Pan, and Singleton (2000), the formula in (2.3) can be interpreted as providing the market price π_t at time t of any time $t+1$ exponential payoff $\exp\left(-ur_{t+1} - v\mu_{t+1}\right)$ from the Laplace transform:

$$\pi_t = E[M_{t+1}(\zeta)\exp\left(-ur_{t+1} - v\mu_{t+1}\right)|I(t)] = \frac{\mathcal{L}_t(u+\zeta_r, v+\zeta_\sigma)}{\mathcal{L}_t(\zeta_r,\zeta_\sigma)}.$$

Note that this statement is more general than that in Duffie, Pan, and Singleton (2000), because here only the SDF is assumed to be exponentially affine without any additional assumption about the historical and the risk-neutral distributions.

2.3 The Volatility Factor

In this subsection, we focus on the specification of the volatility factor μ_{t+1} in the variance-dependent SDF (2.2) during the period $[t, t+1]$. The characterization of the state-price density by Pan (2002) suggests that our volatility factor should be

$$\mu_t = E[IV_{t+1}\,|I(t)]\,, \tag{2.4}$$

where IV_{t+1} is the integrated variance over the period, i.e., $IV_{t+1} = \int_t^{t+1}\sigma_u^2 du$.

It is worth noting that this model is to a large extent a generalization of both Bandi and Renò (2016) and Christoffersen, Heston, and Jacobs (2013). First, suppose σ_u^2 is a diffusion process with a linear drift:

$$\mu_t = E[IV_{t+1}\,|I(t)] = \int_t^{t+1} E[\sigma_u^2\,|I(t)]\,du = A + B\sigma_t^2.$$

In this case, up to a nonlinear transformation of the process σ_u^2 and inclusion of volatility jumps, the continuous time approach of Bandi and Renò (2016) that takes the spot volatility process σ_u^2 as the volatility factor is nested in (2.4). Moreover, as mentioned above, we take advantage of the fact that the daily realized variance computed from high frequency data provides a much less noisy, consistent estimator of the integrated variance than estimators of the spot volatility. For our empirical application, we assume that this estimator accuracy allows us to approximate the forecast μ_t of IV_{t+1} by the forecast of the realized variance RV_{t+1} computed as sum of consecutive squared returns within "day" $[t, t+1]$.

Second, in their discrete time GARCH framework, Christoffersen, Heston, and Jacobs (2013) choose the conditional variance of return $Var[r_{t+1}\,|I(t)]$ as a volatility factor. We encompass their setting by assuming that $\mu_t = Var[r_{t+1}\,|I(t)]$. The maintained assumption

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

has been thoroughly discussed in the literature on "HEAVY models" (High-frEquency-bAsed VolatilitY models) as developed by Shephard and Sheppard (2010). For the sake of specifying a feasible model, they focus on the realized variance RV_{t+1}. We maintain the approximation

$$\mu_t = E[IV_{t+1}\,|I(t)] = E[RV_{t+1}\,|I(t)]\,.$$

As explained by Shephard and Sheppard (2010), $Var[r_{t+1}\,|I(t)]$ may be interpreted as a "close-to-close conditional variance", whereas $E[RV_{t+1}\,|I(t)]$ is the "conditional expectation of the open-to-close variation". Brownlees and Gallo (2010) have empirically checked more generally whether the two measures are related by an exact affine relationship. Not only this restriction is not rejected, but they do not find compelling empirical evidence against the identity $Var[r_{t+1}\,|I(t)] = E[RV_{t+1}\,|I(t)]$. It is worth stressing that these maintained assumptions are about the historical distribution and are unrelated to the specification of the risk-neutral distribution discussed below. In particular, we see the process $\mu_t = E[RV_{t+1}\,|I(t)]$ as a well-defined statistical object unrelated to the risk-neutralization distribution for the purpose of asset pricing.

2.4 Leverage Effect Preserving Change of Measure

In continuous time models, the leverage effect is defined as the instantaneous correlation between the asset return and its volatility process. By virtue of the Girsanov theorem, the leverage effect is unchanged when moving from the historical measure to the risk neutral measure. Interestingly, the fact that the pricing kernel is exponentially affine is sufficient (and necessary to a large extent) for a similar property in discrete time, regarding the absence of a leverage effect.

Proposition 1. *Given $I(t)$, r_{t+1} and μ_{t+1} are conditionally independent for the historical distribution if and only if they are conditionally independent for the risk neutral distribution.*

When the conditional independence condition is violated, we say that there exists a leverage effect. The proof of Proposition 1 is a straightforward implication of (2.3) since independence is characterized by the factorization of the Laplace transform function. Combining independence and (2.3), we have

$$\mathcal{L}_{r,t}(u+\zeta_r)\mathcal{L}_{\sigma,t}(v+\zeta_\sigma) = \mathcal{L}^*_{r,t}(u)\mathcal{L}^*_{\sigma,t}(v)\mathcal{L}_{r,t}(\zeta_r)\mathcal{L}_{\sigma,t}(\zeta_\sigma), \forall u, v \in \mathbb{R},$$

where $\mathcal{L}_{r,t}(.)$ (resp. $\mathcal{L}_{\sigma,t}(.)$) stands for the historical conditional Laplace transform of r_{t+1} (resp. μ_{t+1}) given $I(t)$, and $\mathcal{L}^*_{r,t}(u)$ and $\mathcal{L}^*_{\sigma,t}(v)$ are their risk-neutral counterparts. In particular, by considering the above formula for $u = 0$ or $v = 0$, we obtain

$$\begin{aligned}
\mathcal{L}_{\sigma,t}(u+\zeta_\sigma) &= \mathcal{L}^*_{\sigma,t}(u)\mathcal{L}_{\sigma,t}(\zeta_\sigma), \forall u \in \mathbb{R}, \\
\mathcal{L}_{r,t}(v+\zeta_r) &= \mathcal{L}^*_{r,t}(v)\mathcal{L}_{r,t}(\zeta_r), \forall v \in \mathbb{R}.
\end{aligned}$$

Since the risk-neutral distribution does not depend on the preference parameters ζ_r and ζ_σ, we have the following conclusion in Proposition 2 below.

Proposition 2. *When there is no leverage effect,*
(i) the historical conditional distribution of r_{t+1} given $I(t)$ depends on the pricing kernel only through the parameter ζ_r;
(ii) the historical conditional distribution of μ_{t+1} given $I(t)$ depends on the pricing kernel only through the parameter ζ_σ.

Proposition 2 explains why it is a common belief that one can identify ζ_σ, price of the volatility risk, by comparing the risk neutral and the historical distribution of the volatility factor. A common practice is to draw this comparison through the variance premium.

However, in contrast to the common belief, observations of the risk neutral expectation of the future variance, through the VIX, does not allow retrieving the price ζ_σ of the volatility risk in the presence of a leverage effect. To see this, note that, from (2.3), we can characterize the two risk neutral distributions by

$$\begin{aligned} \mathcal{L}^*_{r,t}(u) &= \frac{\mathcal{L}_t(u+\zeta_r,\zeta_\sigma)}{\mathcal{L}_t(\zeta_r,\zeta_\sigma)}, \\ \mathcal{L}^*_{\sigma,t}(v) &= \frac{\mathcal{L}_t(\zeta_r,v+\zeta_\sigma)}{\mathcal{L}_t(\zeta_r,\zeta_\sigma)}, \end{aligned}$$

and the risk neutral expectation of the volatility factor is given by

$$E^*[\mu_{t+1}\,|I(t)] = \frac{\partial \mathcal{L}^*_{\sigma,t}(0)}{\partial v} = \frac{1}{\mathcal{L}_t(\zeta_r,\zeta_\sigma)}\frac{\partial \mathcal{L}_t(\zeta_r,\zeta_\sigma)}{\partial v}. \tag{2.5}$$

In the presence of a leverage effect, $(\mathcal{L}_t(\zeta_\sigma,\zeta_r) \neq \mathcal{L}_{\sigma,t}(\zeta_\sigma)\mathcal{L}_{r,t}(\zeta_2))$, $\mathcal{L}_{r,t}(\zeta_r)$ cannot be factorized out for the purpose of simplification between the numerator and the denominator in (2.5). Thus, the risk neutral expectation of the future variance should depend (for a given historical Laplace transform) on both ζ_σ and ζ_r and it is not clear how the two parameters could be disentangled. This lack of identification confirms what Pan (2002) proves with her computation of the state-price density, see formula (A2) on page 34 of Pan (2002). The state-price density formula shows that the market price of the volatility Brownian shock depends not only on ζ_σ but also on $\rho\zeta_r$, where ρ is the instantaneous correlation between the two Brownian shocks corresponding to the return and the volatility. The same observation is made in discrete time by Christoffersen, Heston, and Jacobs (2013), whose formula (4) shows that the risk premium on the volatility risk depends on ζ_r in the presence of a leverage effect.

2.5 Leverage-Effect-Based Identification of Risk Prices

In contrast to another common belief, observations of the price of the asset may actually provide identification of ζ_σ, price of the volatility risk. By definition, the absence of arbitrage implies that the risk neutral expectation of the excess return $\exp(r_{t+1})$ equals to unity, i.e.,

$$\mathcal{L}^*_{r,t}(-1) = E^*[\exp(r_{t+1})\,|I(t)] = 1. \tag{2.6}$$

Combining this equality with the Laplace transform in (2.3) with $u = -1$ and $v = 0$, the pricing equation of the underlying asset satisfies

$$\mathcal{L}_t(\zeta_r, \zeta_\sigma) = \mathcal{L}_t(\zeta_r - 1, \zeta_\sigma). \tag{2.7}$$

In the presence of a leverage effect, i.e., $\mathcal{L}_t(\zeta_\sigma, \zeta_r) \neq \mathcal{L}_{r,t}(\zeta_r)\mathcal{L}_{\sigma,t}(\zeta_\sigma)$, $\mathcal{L}_{\sigma,t}(\zeta_\sigma)$ cannot be factorized out for simplification of (2.7). Thus, (2.7) provides identification information for both ζ_r and ζ_σ for a given historical Laplace transform. We set the focus of this paper on a model where the bivariate process $W_t = (r_t, \mu_t)$ is Markov of order one. In this case, identification of the coefficients of the two state variables in equation (2.7) provides in general two independent equations that allow for identification of both risk price parameters ζ_r and ζ_σ.

This result can be seen as a generalization of the result first put forward by Bandi and Renò (2016). They stress that their exponentially affine variance-dependent pricing kernel (see their specification (17) on page 125) outlines some restrictions on the risk aversion coefficients (see their Proposition 8.1) ζ_σ and ζ_r induced by the absence of arbitrage. Even though both the historical and risk-neutral distributions are viewed as nonparametric, their equation (19) shows that the two risk aversion parameters can be identified from asset prices when there is a leverage effect, see coefficient $\psi\rho(\sigma)\sigma$ in their equation (19). We find a similar result, again with the exponentially affine variance-dependent pricing kernel, while keeping nonparametric specifications for the historical and risk-neutral distributions. We are arguably even more nonparametric since we consider unrestricted Laplace transforms instead of diffusion processes. Moreover, this nonparametric framework allows us to be more precise about the possibility of identification of both risk price parameters ζ_r and ζ_σ from observations of the asset return. The following Proposition 3 shows that while the presence of a leverage effect is a necessary condition, it is not fully sufficient. Not only the conditional distribution of r_{t+1} given $I(t)$ and μ_{t+1} must depend on μ_{t+1}, but also this dependence must not be hidden by the value of ζ_r the parameter of risk aversion for equity.

Let $\mathcal{L}_{(r|\sigma),t}(u) = E[\exp(-ur_{t+1})\,|I^\sigma(t)]$ stands for the conditional Laplace transform of r_{t+1} given the information set $I^\sigma(t) = I(t) \vee \{\mu_{t+1}\}$.

Proposition 3. *Suppose $\mathcal{L}_{(r|\sigma),t}(u)$ can be factorized for all u as*

$$\mathcal{L}_{(r|\sigma),t}(u) = A\left(u, \mu_{t+1}\right) B\left(u, I(t)\right)$$

for some given (independent of risk aversion parameters) deterministic functions A and B

and the risk aversion parameter ζ_r *satisfies*

$$A\left(\zeta_r, \mu_{t+1}\right) = A\left(\zeta_r - 1, \mu_{t+1}\right). \tag{2.8}$$

Then,

$$\mathcal{L}_t(\zeta_r, \zeta) = \mathcal{L}_t(\zeta_r - 1, \zeta), \forall \zeta \geq 0 \Longleftrightarrow B\left(\zeta_r, I(t)\right) = B\left(\zeta_r - 1, I(t)\right).$$

Hence, the condition (2.8) implies that the arbitrage condition (2.6) (or equivalently (2.7)) does not identify the risk aversion parameter ζ_σ but only possibly ζ_r. The condition (2.8) is obviously implied by the absence of a leverage effect (see $A\left(u, \mu_{t+1}\right) \equiv 1$) and is dubbed hereafter the "local zero-leverage". We show in Section 3 that the model of Corsi, Fusari, and La Vecchia (2013) is a case where there is a leverage effect but a local zero-leverage.

To sum up, the presence of a leverage effect should reverse both sides of common beliefs. Not only do observations of variance premium not provide identification of the price of volatility risk ζ_σ, but also, on the contrary, observations of the return of the asset of interest provide in general identification of both risk prices ζ_r and ζ_σ. However, we understand that this identification is fragile. It may in practice be quite weak, either because the leverage effect is small or the risk aversion for equity is close to local zero-leverage.

3 An Exponentially Affine Pricing Model

While our theoretical result about identification is model-free (up to the maintained assumption of an exponential affine pricing kernel), the link functions for the minimum distance estimation of the risk prices must be built from moment restrictions about the joint process of the return and the volatility factor. As already mentioned, it has been known since Duffie, Pan, and Singleton (2000) that exponentially affine models of return and volatility are very convenient to get closed form formulas of the conditional Laplace transforms and associated moments. We propose a discrete time version of this approach.

Our bivariate discrete time model belongs to the general class of Compound AutoRegressive Models (CAR) of Darolles, Gouriéroux, and Jasiak (2006). Although we assume a CAR structure, we remain in this section nonparametric about the functions $a(\cdot), b(\cdot), \alpha(\cdot), \beta(\cdot), \gamma(\cdot)$ (see below) that define the CAR model. The stochastic process (r_t, μ_t) is stationary Markov of order one. The conditional joint distribution of (r_{t+1}, μ_{t+1}) given $I(t)$ is factorized as the product of (i) the conditional distribution of μ_{t+1} given $I(t)$, and (ii) the conditional distribution of r_{t+1} given $I^\sigma(t) = I(t) \vee \{\mu_{t+1}\}$.

We first define the historical distribution and then deduce the risk-neutral distribution using the SDF $M_{t+1}(\zeta)$.

3.1 Historical Distribution

Following Renault (1997) and Garcia and Renault (1998), we preclude any Granger causality relationship from the asset return to its volatility. Moreover, as in all popular models, consecutive asset returns are assumed to be serially independent given the path of the volatility factor. Hence, we specify the historical distributions as follows. (i) The conditional distribution of μ_{t+1} given $I(t)$ depends only on the last past value μ_t. We characterize this distribution by its exponentially affine Laplace transform

$$\mathcal{L}_{\sigma,t}(v) = E[\exp\left(-v\mu_{t+1}\right)|I(t)] = \exp\left\{-a(v)\mu_t - b(v)\right\}. \tag{3.1}$$

(ii) The conditional distribution of r_{t+1} given $I^{\sigma}(t) = I(t) \vee \{\mu_{t+1}\}$ depends only on current and the last past values μ_{t+1} and μ_t of the volatility factor. Similarly, we have an exponentially affine conditional Laplace transform

$$\mathcal{L}_{(r|\sigma),t}(u) = E[\exp\left(-ur_{t+1}\right)|I^{\sigma}(t)] = \exp\left\{-\alpha(u)\mu_{t+1} - \beta(u)\mu_t - \gamma(u)\right\}. \tag{3.2}$$

The bivariate CAR model for (r_{t+1}, μ_{t+1}) can be deduced from (3.1) and (3.2) by the Law of Iterated Expectations:

$$\mathcal{L}_t(u, v) = E[\exp\{-v\mu_{t+1}\}\mathcal{L}_{(r|\sigma),t}(u)|I(t)]. \tag{3.3}$$

Note that by definition,

$$a(0) = b(0) = \alpha(0) = \beta(0) = \gamma(0) = 0.$$

Several remarks are in order. First, the model (3.1) assumes that the volatility factor μ_t is a Markov process of order one. We note that this Markov property is fulfilled in particular by the model in Heston (1993) as well as the model in Bandi and Renò (2016) in the case (with their notations) $\xi\left(\sigma_t^2\right) = \sigma_t^2$. In these models, μ_t is an affine function of the diffusion process σ_t^2.

Second, the realized variance RV_{t+1} is the sum of the $AR(1)$ process μ_t and a forecast error. The forecast error by definition is a martingale difference sequence. Therefore, the process RV_t is $ARMA(1,1)$, rather than the more restrictive $AR(1)$ process sometimes employed for simplification. While Han, Khrapov, and Renault (2020) estimate an unconstrained $ARMA(1,1)$ model, we resort in this paper (see section 4.1) to a HEAVY model.

Third, as in Heston (1993), the affine structure of the volatility model applies not only to the conditional expectation ($AR(1)$ structure) but also to the conditional variance:

$$\begin{aligned} E[\mu_{t+1} \,|I(t)] &= a'(0)\mu_t + b'(0), \\ Var[\mu_{t+1} \,|I(t)] &= -a''(0)\mu_t - b''(0). \end{aligned}$$

Similarly, for the asset return conditional on $I^{\sigma}(t) = I(t) \vee \{\mu_{t+1}\}$,

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

Fourth, the leverage effect, defined as the instantaneous causality between r_{t+1} and μ_{t+1}, is present if and only if the function $\alpha(\cdot)$ is not identically zero (or equivalently not constant). When there is no leverage, we deduce from (3.3) the factorization

$$\mathcal{L}_t(u,v) = \mathcal{L}_{\sigma,t}(v)\mathcal{L}_{(r|\sigma),t}(u) = \mathcal{L}_{\sigma,t}(v)\mathcal{L}_{r,t}(u).$$

3.2 Risk-Neutral Distribution

A desirable feature of the exponentially affine SDF is that it delivers risk-neutral distributions with patterns analogous to those of the historical ones, with adjusted parameter values incorporating the two risk aversion parameters. We prove in the appendix that the conditional Laplace transform that governs the risk-neutral distributions can be written as follows.

Proposition 4. *(i) The risk-neutral conditional distribution of μ_{t+1} given $I(t)$ is characterized by its exponentially affine Laplace transform*

$$\mathcal{L}^*_{\sigma,t}(v) = E^*[\exp\left(-v\mu_{t+1}\right) |I(t)] = \exp\left\{-a^*(v)\mu_t - b^*(v)\right\}$$

with

$$\begin{aligned} a^*(v) &= a\left(v + \zeta_\sigma + \alpha\left(\zeta_r\right)\right) - a\left(\zeta_\sigma + \alpha\left(\zeta_r\right)\right), \\ b^*(v) &= b\left(v + \zeta_\sigma + \alpha\left(\zeta_r\right)\right) - b\left(\zeta_\sigma + \alpha\left(\zeta_r\right)\right). \end{aligned}$$

(ii) The risk-neutral conditional distribution of r_{t+1} given $I^{\sigma}(t) = I(t) \vee \{\mu_{t+1}\}$ is

characterized by its exponentially affine Laplace transform

$$\mathcal{L}^*_{(r|\sigma),t}(u) = E^*[\exp(-ur_{t+1})\,|I^\sigma(t)] = \exp\left\{-\alpha^*(u)\mu_{t+1} - \beta^*(u)\mu_t - \gamma^*(u)\right\}$$

with

$$\begin{aligned}
\alpha^*(u) &= \alpha(u+\zeta_r) - \alpha(\zeta_r), \\
\beta^*(u) &= \beta(u+\zeta_r) - \beta(\zeta_r), \\
\gamma^*(u) &= \gamma(u+\zeta_r) - \gamma(\zeta_r).
\end{aligned}$$

Note that the function $\alpha^*(\cdot)$ is identically zero if and only if the function $\alpha(\cdot)$ is identically zero (or equivalently identically constant). As discussed in Proposition 1, there is no leverage in the risk-neutral world if and only if there is no leverage in the historical world. However, we argue in the next subsection that the correct measure of the leverage effect must be computed in the risk-neutral world. This is consistent with the common practice to assess the amount of leverage from the shape of the volatility smirk (see e.g., Renault, 1997).

3.3 Leverage Effect Characterized by Risk-Neutral Distribution

As explained in the introduction, the leverage effect in discrete time is a parameter, denoted by LEV, that must be defined from the risk-neutral distribution, to avoid confusion with the volatility feedback effect due to risk compensation. Following the intuition of Black (1976), it should measure the negative instantaneous impact of the volatility factor in the risk-neutral return forecast

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

As such, the amount of leverage is

$$LEV = -\alpha^*(-1) \leq 0.$$

Note that Proposition 4 implies $LEV = -\alpha^*(-1) = \alpha(\zeta_r) - \alpha(\zeta_r - 1)$. Therefore, the leverage effect parameter LEV is zero if and only if there is local zero-leverage.

We also note that the historical world analog of the parameter LEV, $-\alpha(-1)$, underestimates (in absolute value) the true leverage in general, i.e., $-\alpha^*(-1) < -\alpha(-1)$. This holds because $-\alpha^*(-1) = \alpha(\zeta_r - 1) - \alpha(\zeta_r)$, viewed as a function of ζ_r, is expected to be an increasing function when ζ_r is positive and near zero, and $\alpha(\cdot)$ is approximately a quadratic function. To see this, we approximate the function $\alpha(x)$ by its quadratic expansion around zero and obtain

$$\alpha(x-1) - \alpha(x) \approx -\alpha'(0) - \frac{\alpha''(0)}{2}\left[x^2 - (x-1)^2\right] = -\alpha'(0) + \frac{\alpha''(0)}{2}[1-2x].$$

The right hand side is an increasing function of x because $\alpha''(0) \leq 0$, as seen from the formula for $Var[r_{t+1}\,|I^\sigma(t)]$.

This result is a striking confirmation of the difficulty to identify the leverage effect in discrete time as documented by Bollerslev, Litvinova, and Tauchen (2006). There is an attenuation effect, due to the risk aversion ζ_r for the return risk. We actually expect that in the historical expectation of the asset return, namely in the quantity $-\alpha(-1)$, there is not only the negative term LEV, but also a positive risk compensation term increasing with the risk aversion parameter ζ_r. This positive addition to the negative term LEV, dubbed the volatility feedback, attenuates its magnitude.

3.4 Identification of Volatility Risk Price

As discussed in Section 2, identification of the risk aversion parameter ζ_σ from asset returns, if possible, should come from the arbitrage identity

$$\mathcal{L}_t(\zeta_r, \zeta_\sigma) = \mathcal{L}_t(\zeta_r - 1, \zeta_\sigma). \tag{3.4}$$

Following the definition of the historical distribution above, we obtain the following results.

Proposition 5. The arbitrage identity (3.4) is equivalent to the conjunction of two identities

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

Remark 1. Proposition 5 provides two different equations that should allow us, in general, to identify both risk aversion parameters ζ_r and ζ_σ, in contrast with the common belief that only the risk aversion to equity could be identified by asset return data. There is, however, an important exception, which is the case of local zero-leverage:

$$\alpha(\zeta_r) = \alpha(\zeta_r - 1). \tag{3.5}$$

In the case of local zero-leverage, the two equations collapse into $\beta(\zeta_r) = \beta(\zeta_r - 1)$ and $\gamma(\zeta_r) = \gamma(\zeta_r - 1)$, and they do not depend on the parameter ζ_σ anymore.

Remark 2. If the functions $\beta(\cdot)$ and $\gamma(\cdot)$ are constant (identical to zero), the two equations above collapse into

$$\begin{aligned} a\left[\zeta_\sigma + \alpha\left(\zeta_r\right)\right] &= a\left[\zeta_\sigma + \alpha\left(\zeta_r - 1\right)\right], \\ b\left[\zeta_\sigma + \alpha\left(\zeta_r\right)\right] &= b\left[\zeta_\sigma + \alpha\left(\zeta_r - 1\right)\right]. \end{aligned}$$

These two equations are fulfilled if $\alpha\left(\zeta_r\right) = \alpha\left(\zeta_r - 1\right)$. This is the case for the model of Corsi, Fusari, and La Vecchia (2013), and it is the reason why they cannot identify ζ_σ aversion to the volatility risk. It is worth noting that in our case this difficulty is solved by the specification of a non-zero function $\beta(\cdot)$. We make the conditional distribution $\mathcal{L}_{(r|\sigma),t}(\cdot)$ depending not only on the current value μ_{t+1} of the volatility factor (the leverage effect) but also on the state variable μ_t. Without this dependence, there is a leverage effect but it is the local zero-leverage.

Remark 3. Next, we discuss the variance premium. As stressed by Drechsler and Yaron (2011), the variance premium is non-zero because of two effects. First, the conditional variance of the return differs in the historical world and the risk-neutral one, i.e.,

$$\begin{aligned} Var^*[r_{t+1}\,|I^\sigma(t)] &= -\alpha^{*\prime\prime}(0)\mu_{t+1} - \beta^{*\prime\prime}(0)\mu_t - \gamma^{*\prime\prime}(0) \\ &= -\alpha^{\prime\prime}(\zeta_r)\mu_{t+1} - \beta^{\prime\prime}(\zeta_r)\mu_t - \gamma^{\prime\prime}(\zeta_r) \\ &\neq -\alpha^{\prime\prime}(0)\mu_{t+1} - \beta^{\prime\prime}(0)\mu_t - \gamma^{\prime\prime}(0) \\ &= Var[r_{t+1}\,|I^\sigma(t)], \end{aligned}$$

following Proposition 4. It turns out that, for reasons explained later, we assume in Section 4 and hereafter that the functions $\alpha(\cdot), \beta(\cdot), \gamma(\cdot)$ are quadratic. At least, we consider them as well approximated by their quadratic expansions around zero such that the functions $\alpha''(\cdot), \beta''(\cdot), \gamma''(\cdot)$ are approximately constant, and $Var^*[r_{t+1}\,|I^\sigma(t)] \approx Var[r_{t+1}\,|I^\sigma(t)]$.

Second, the variance premium shows up because the forecast of the volatility factor differs in the historical world and the risk-neutral one, i.e.,

$$E\{Var[r_{t+1}\,|I^\sigma(t)]\,|I(t)\} \neq E^*\{Var^*[r_{t+1}\,|I^\sigma(t)]\,|I(t)\},$$

because $E[\mu_{t+1}\,|I(t)] \neq E^*[\mu_{t+1}\,|I(t)]$. Actually, following Proposition 4,

$$\begin{aligned} E^*[\mu_{t+1}\,|I(t)] &= a^{*\prime}(0)\mu_t + b^{*\prime}(0) \\ &= a'\left[\zeta_\sigma + \alpha\left(\zeta_r\right)\right]\mu_t + b'\left[\zeta_\sigma + \alpha\left(\zeta_r\right)\right], \end{aligned}$$

and it differs in general from

$$E[\mu_{t+1}\,|I(t)] = a'(0)\mu_t + b'(0).$$

However, we see that this difference, which is the main source of the variance premium, in general does not allow for separate identification of the two risk aversion parameters ζ_σ and ζ_r. We only obtain identification of the risk aversion ζ_σ in the absence of a leverage effect, i.e., the function $\alpha(\cdot)$ is identically zero. In this case,

$$E^*[\mu_{t+1}\,|I(t)] = a'(\zeta_\sigma)\,\mu_t + b'(\zeta_\sigma)\,,$$

and the risk-neutral conditional expectation of the realized variance (as observed from VIX) is rightly used, as it is as a common practice, to identify ζ_σ.

4 The Empirical Model

The first two subsections maintain a semiparametric framework with unspecified functions $a(\cdot), b(\cdot), \alpha(\cdot), \beta(\cdot), \gamma(\cdot)$, focusing on the restrictions between the quadratic expansions of these functions as well as a dynamic model for the volatility factor. Subsections 4.3 specifies a parametric model for the joint conditional distribution of return and volatility factor given the past. This means a parametric model for functions $a(\cdot), b(\cdot), \alpha(\cdot), \beta(\cdot), \gamma(\cdot)$. The last two subsections provide the link functions between several reduced form parameters and three structural ones, namely the two risk aversion parameters ζ_r and ζ_σ and one parameter ϕ summarizing the leverage effect.

4.1 HEAVY Model for the Volatility Factor

As discussed in subsection 3.1. the volatility factor μ_t is an $AR(1)$ process, whose transition distribution is given by the conditional Laplace transform

$$\mathcal{L}_{\sigma,t}(v) = E[\exp\left\{-v\mu_{t+1}\right\}|I(t)] = \exp\left\{-a(v)\mu_t - b(v)\right\}.$$

Unfortunately, we have no direct observations of the volatility factor. We only observe daily realized variance (RV_{t+1}) from high-frequency data over the day $[t, t+1]$ and by definition

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

Therefore, RV_t is assumed to be an $ARMA(1,1)$ process.

To understand the HEAVY model, it is worth sketching an analogy with the $GARCH(1,1)$ model. When a return process r_t has a zero conditional mean and its conditional variance $h_t = E[r_{t+1}^2 | I(t)]$ is an $AR(1)$ process, r_t^2 is an $ARMA(1,1)$ process. However, there is a general agreement not to estimate directly this $ARMA$ process but to write a $GARCH(1,1)$ equation

$$\begin{aligned} h_{t+1} &= \omega + (\alpha+\beta)h_t + \alpha\left(r_{t+1}^2 - h_t\right) \\ &= \omega + \beta h_t + \alpha r_{t+1}^2. \end{aligned}$$

In other words, we constrain the innovation of the $AR(1)$ conditional variance process h_t to be proportional to the martingale difference sequence $r_{t+1}^2 - h_t$ to get a simple model that can be estimated by Gaussian QMLE. Such an estimator is likely to be much more accurate than a linear estimator based on a general $ARMA(1,1)$ representation. The key idea is that the QMLE estimator works better than the linear estimator in taking into account the conditional heteroskedasticity of the $ARMA(1,1)$ process. The HEAVY model works similarly. Instead of only using the $ARMA(1,1)$ model of the realized variance process RV_t, we specify the following model for the volatility factor μ_t:

$$\begin{aligned} \mu_{t+1} &= \varpi + \rho\mu_t + \varkappa\left(RV_{t+1} - \mu_t\right) \\ &= \varpi + (\rho - \varkappa)\mu_t + \varkappa RV_{t+1}. \end{aligned} \tag{4.1}$$

This model provides a neat representation of the conditional distribution of the return r_{t+1} given the past information $I(t)$ and the current realized variance RV_{t+1}. This is the focus of Corsi, Fusari, and La Vecchia (2013). It actually coincides, by virtue of (4.1), with the conditional distribution of r_{t+1} given past information $I(t)$ and the current volatility factor μ_{t+1}. In other words, by replacing RV_{t+1} with μ_{t+1}, we have not changed the conditioning information (for the "realizing smiles"), but only made it well summarized by (r_{t+1}, μ_{t+1}), a bivariate Markov process of order one.

This model is easy to estimate. Following Shephard and Sheppard (2010), we estimate it with observations $RV_t, t = 1, 2, \ldots, T$ and the following recursion:

$$\begin{aligned} \mu_{t+1}(\lambda) &= \varpi + \lambda_1 RV_{t+1} + \lambda_2 \mu_t(\lambda), t = 1, \ldots, T-1, \\ \lambda &= (\varpi, \lambda_1, \lambda_2)', \mu_1 = \frac{1}{\sqrt{T}} \sum_{t=1}^{[\sqrt{T}]} RV_t. \end{aligned}$$

This recursion allows us to compute the quasi-log likelihood function

$$L_T(\lambda) = \sum_{t=1}^{T-1} l[RV_{t+1} \,| I(t), \lambda] \,,$$

where

$$l[RV_{t+1} \,| I(t), \lambda] = -\frac{1}{2} \log \left[\mu_t (\lambda)\right] - \frac{RV_{t+1}}{2\mu_t (\lambda)}$$

is computed as if the conditional distribution of RV_{t+1} given $I(t)$ is the square of a normally distributed random variable with zero mean and variance $\mu_t (\lambda)$. Maximization of the quasi-log likelihood delivers the QMLE estimator $\hat{\lambda} = (\hat{\varpi}, \hat{\lambda}_1, \hat{\lambda}_2)'$ and the recursion

$$\begin{aligned} \hat{\mu}_{t+1} &= \hat{\varpi} + \hat{\lambda}_1 RV_{t+1} + \hat{\lambda}_2 \hat{\mu}_t, t = 1, \ldots, T-1, \\ \hat{\mu}_1 &= \frac{1}{\sqrt{T}} \sum_{t=1}^{[\sqrt{T}]} RV_t. \end{aligned}$$

For the purpose of estimating our structural pricing model, we assume hereafter that the filtered values $\hat{\mu}_t, t = 1, \ldots, T$ of the volatility factor μ_t are accurate proxies of the true values. We ignore the random errors $\hat{\mu}_t - \mu_t, t = 1, \ldots, T$ in subsequent analysis.

4.2 Bridging Conditional Variance and Realized Variance

As discussed in section 2.3., the volatility factor μ_t, besides following the HEAVY model, is connected to the conditional variance of the return through the constraint

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

From the CAR model for the historical distribution of r_{t+1} given $I(t)$ (see section 3.1), this constraint is characterized as

$$\begin{aligned} \mu_t &= E\left\{Var[r_{t+1} \,| I^{\sigma}(t)] \,| I(t)]\right\} + Var\left\{E[r_{t+1} \,| I^{\sigma}(t)] \,| I(t)]\right\}, \text{ where} \\ E[r_{t+1} \,| I^{\sigma}(t)] &= \alpha'(0)\, \mu_{t+1} + \beta'(0)\, \mu_t + \gamma'(0)\,, \\ Var[r_{t+1} \,| I^{\sigma}(t)] &= -\alpha''(0)\, \mu_{t+1} - \beta''(0)\, \mu_t - \gamma''(0)\,, \end{aligned}$$

such that (4.2) is equivalent to

$$\mu_t = -\alpha''(0)\, E[\mu_{t+1} \,| I(t)] - \beta''(0)\, \mu_t - \gamma''(0) + [\alpha'(0)]^2\, Var[\mu_{t+1} \,| I(t)] \,.$$

Moreover, from the historical distribution of μ_{t+1} given $I(t)$, we have

$$\begin{aligned} E[\mu_{t+1} | I(t)] &= a'(0)\mu_t + b'(0), \\ Var[\mu_{t+1} | I(t)] &= -a''(0)\mu_t - b''(0). \end{aligned}$$

As such, the constraint above is equivalent to

$$\mu_t + \alpha''(0) [a'(0)\mu_t + b'(0)] + \beta''(0) \mu_t + \gamma''(0) = [\alpha'(0)]^2 [-a''(0)\mu_t - b''(0)] .$$

Both the coefficients of μ_t and the constant term must be identically zero, implying that (4.2) is equivalent to the conjunction of the following two restrictions:

$$\begin{aligned} 1 + \alpha''(0) a'(0) + \beta''(0) + [\alpha'(0)]^2 a''(0) &= 0, \\ \alpha''(0) b'(0) + \gamma''(0) + [\alpha'(0)]^2 b''(0) &= 0. \end{aligned}$$

It is worth interpreting these restrictions in terms of the historical moments of the return process. First, consider the condition mean $E[r_{t+1} | I^\sigma(t)] = \alpha'(0) \mu_{t+1} + \beta'(0) \mu_t + \gamma'(0)$. We focus on the parameter

$$\psi = \alpha'(0),$$

which encapsulates two effects: the leverage effect and the volatility feedback. The leverage effect is expected to be negative. It characterizes the instantaneous negative causality between the asset return and its current volatility. The volatility feedback is expected to be a positive increasing function of the parameter ζ_r, measuring the aversion to the risk of variation in the return. Second, consider the conditional variance $Var[r_{t+1} | I^\sigma(t)] = -\alpha''(0) \mu_{t+1} - \beta''(0) \mu_t - \gamma''(0)$. We focus on the parameter

$$\xi = -\alpha''(0) ,$$

which is another leverage effect parameter expected to be between zero and one. The observation of the current volatility factor must reduce the expected variation of the return.

We summarize the message of this section in the following Proposition.

Proposition 6. *The conditional variance of the return* can be written as $Var[r_{t+1} | I^\sigma(t)] = \xi\mu_{t+1} - \beta''(0) \mu_t - \gamma''(0)$, where the coefficients $\beta''(0)$ and $\gamma''(0)$ satisfy the constraints

$$\begin{aligned} -\beta''(0) &= 1 - \xi a'(0) + \psi^2 a''(0), \\ -\gamma''(0) &= -\xi b'(0) + \psi^2 b''(0), \end{aligned}$$

and $\psi = \alpha'(0)$, $\xi = -\alpha''(0)$ are two reduced form parameters that are both related to the leverage effect (and the volatility feedback for ψ).

Remark. As discussed above, we plan to eventually approximate the function $\alpha(\cdot)$ by its quadratic approximation $\alpha^R(\cdot)$. It is worth noting that the two reduced form parameters ψ and ξ, both related to the leverage effect, are sufficient to characterize this quadratic approximation with $\alpha^R(u) = \psi u - \frac{\xi}{2}u^2$. When the function $\alpha(\cdot)$ coincides with its quadratic approximation $\alpha^R(\cdot)$, the joint nullity of the two parameters ψ and ξ characterizes the nullity of the quadratic approximation $\alpha^R(\cdot)$ and the absence of a leverage effect.

4.3 A Fully Parametric Model

4.3.1 Conditional Normality of the Return Given the Volatility Factor

Ané and Geman (2000) (see also Clark, 1973, for a seminal contribution) argue that for log-returns, conditional normality can be recovered when conditioning on a measure of the market activity. We follow Corsi, Fusari, and La Vecchia (2013) and Han, Khrapov, and Renault (2020) to adopt the realized variance RV_{t+1} over the time interval $[t, t+1]$ as a measure of the market activity. Moreover, following (4.1), we know that, in the framework of our HEAVY model, this is tantamount to conditioning on the current volatility factor μ_{t+1}. In other words, this conditional normal distribution is fully characterized by the conditional Laplace transform $\mathcal{L}_{(r|\sigma),t}(u) = E[\exp(-ur_{t+1})\,|I^{\sigma}(t)] = \exp\left\{-\alpha(u)\mu_{t+1} - \beta(u)\mu_t - \gamma(u)\right\}$. Moreover, the normality implies that

$$E[\exp(-ur_{t+1})\,|I^{\sigma}(t)] = \exp\left\{-um_t\left(\mu_{t+1}\right) + \frac{u^2}{2}\omega_t\left(\mu_{t+1}\right)\right\},$$

where $m_t\left(\mu_{t+1}\right)$ and $\omega_t\left(\mu_{t+1}\right)$ stand respectively for the mean and variance of the conditional normal distribution of r_{t+1}. Therefore, the conditional normality holds if and only if the functions $\alpha(\cdot), \beta(\cdot), \gamma(\cdot)$ are all quadratic. Following Proposition 6, we specify

$$\begin{aligned} \alpha(u) &= \psi u - \frac{\xi}{2}u^2, \\ \beta(u) &= \beta u + \left[\xi a'(0) - 1 - \psi^2 a''(0)\right]\frac{u^2}{2}, \\ \gamma(u) &= \gamma u + \left[\xi b'(0) - \psi^2 b''(0)\right]\frac{u^2}{2}, \end{aligned} \tag{4.3}$$

which implies that

$$\begin{aligned} m_t\left(\mu_{t+1}\right) &= \psi\mu_{t+1} + \beta\mu_t + \gamma, \\ \omega_t\left(\mu_{t+1}\right) &= \xi\mu_{t+1} - \left[\xi a'(0) - 1 - \psi^2 a''(0)\right]\mu_t - \left[\xi b'(0) - \psi^2 b''(0)\right]. \end{aligned}$$

Remark 1. We know from section 3.3 that

$$LEV = \alpha\left(\zeta_r\right) - \alpha\left(\zeta_r - 1\right) = \psi - \frac{\xi}{2}\left[\zeta_r^2 - \left(\zeta_r - 1\right)^2\right] = \psi - \xi\left(\zeta_r - \frac{1}{2}\right)$$

Therefore,

$$\psi = LEV + \xi\left(\zeta_r - \frac{1}{2}\right) \tag{4.4}$$

characterizes the instantaneous impact of the volatility factor μ_{t+1} on the expected return through both the leverage effect, by the term $LEV\ \mu_{t+1}$, and the volatility feedback, by the term $\xi\mu_{t+1}\left(\zeta_r - \frac{1}{2}\right)$. Note that, as already explained, the volatility feedback looks spuriously instantaneous in a discrete time framework and is proportional to both the expected current return variation $\xi\mu_{t+1}$ and the risk aversion ζ_r (up to a Jensen effect). We note that, when there is no leverage effect, i.e., $\psi = \xi = 0$, not only does the leverage measure LEV equal zero but there is no volatility feedback.

Remark 2. Of course, the absence of a leverage effect does not eliminate the risk compensation. In this case, the risk compensation is only carried by μ_t, the value of the volatility factor at the beginning of the period $[t, t+1]$. More precisely, when $\psi = \xi = 0$, we know from (4.3) that $\beta(u) = \beta u - \frac{u^2}{2}$, and we know from the equilibrium pricing relationship of Proposition 5 that $\beta\left(\zeta_r\right) - \beta\left(\zeta_r - 1\right) = 0$. Together, they imply that $\beta = \zeta_r - \frac{1}{2}$ in the absence of a leverage effect. It can be easily checked that in this case the function $\gamma(\cdot)$ is identically zero. As such, in the absence of a leverage effect, the expected return is

$$E[r_{t+1}\,|I^\sigma(t)] = E[r_{t+1}\,|I(t)] = \left(\zeta_r - \frac{1}{2}\right)\mu_t,$$

and we are back to the standard scenario where the risk premium is proportional to the risk aversion, which can be easily estimated.

Remark 3. It is worth stressing that the conditional normality of the return given $I(t)$ and some latent variable μ_{t+1} is hardly a restrictive assumption. It allows for skewness and fat tails not only in the unconditional distribution of the returns but also in the conditional distribution given the past observations that define the investor's information set $I(t)$. The mixture variable μ_{t+1}, which we dub the volatility factor, maintains some randomness in the conditional mean $m_t(\mu_{t+1})$ and variance $\omega_t(\mu_{t+1})$ of the Gaussian distribution. In particular, it is much more general than the conditional log-normality given the past returns, a model that is traditionally used for the QMLE estimation of GARCH models.

4.3.2 ARG(1) Model for the Volatility

We specify a discrete time model inspired by the continuous time model of Heston (1993). Following Gouriéroux and Jasiak (2006), we consider the simplest version where the transition dynamics are driven by the Gamma distributions as in the affine model of Heston (1993) and its precursor Feller (1951)'s square root process. More precisely, we use the ARG(1) model defined by Gouriéroux and Jasiak (2006) as follows: (i) The conditional distribution of μ_{t+1} given some mixing variable U_t is the Gamma distribution with the shape parameter $\delta + U_t$ and a scale parameter c. (ii) The conditional distribution of U_t given μ_t is the Poisson distribution with the parameter $\rho\mu_t/c$. This parametric model is nested in the general affine model defined in (3.1) with the specification

$$a(u) = \frac{\rho u}{1 + cu},\ b(u) = \delta \log(1 + cu). \tag{4.5}$$

In this parametric model, the reduced form parameters $\omega_1 = (\rho, c, \delta)'$ are identified by the first two conditional moments

$$\begin{aligned} E[\mu_{t+1} | I(t)] &= \rho\mu_t + \delta c, \\ Var[\mu_{t+1} | I(t)] &= 2\rho c\mu_t + \delta c^2. \end{aligned} \tag{4.6}$$

The interpretation of these three parameters in ω_1 is as follows: (i) ρ is the autocorrelation parameter in the heteroskedastic $AR(1)$ process μ_t, (ii) c is a scale parameter, and (iii) δ is the location parameter. Together, they determine the unconditional mean $E[\mu_{t+1}] = \delta c/(1-\rho)$.

We recall that together with the additional reduced form parameters $\omega_2 = (\gamma, \beta, \psi, \xi)'$, the vector of reduced form parameters $\omega = (\omega_1', \omega_2')'$ also defines the first two conditional moments of the return

$$\begin{aligned} E[r_{t+1} | I^\sigma(t)] &= \psi\mu_{t+1} + \beta\mu_t + \gamma, \\ Var[r_{t+1} | I^\sigma(t)] &= \xi\mu_{t+1} + \left[1 - \rho\xi - 2\rho c\psi^2\right]\mu_t - \left[\delta c^2\psi^2 + \xi\delta c\right]. \end{aligned} \tag{4.7}$$

4.4 One Structural Parameter for Two Channels of Leverage Effect

As explained in the comments introducing Proposition 6, we have introduced two channels for the leverage effect so far. On the one hand, the parameter $\xi \in [0,1]$ captures the idea that, once the value μ_{t+1} of the volatility factor is known, there remains less uncertainty about the current return due to the leverage effect. On the other hand, the parameter $LEV \leq 0$ captures the idea that, due to the negative instantaneous correlation between the volatility and the return, the knowledge of μ_{t+1} leads to a reduction of the return forecast by $LEV \cdot \mu_{t+1}$.

4.4.1 Continuous Time Intuition

To understand our discrete-time model with a leverage effect, we sketch an analogous standard continuous time model with stochastic volatility:

$$\begin{aligned} \frac{dS_t}{S_t} &= \vartheta_t dt + \sigma_t dW_t^S, \\ d\sigma_t &= \varsigma_t dt + \varkappa_t dW_t^\sigma, \\ Corr\left[dW_t^S, dW_t^\sigma\right] &= \phi \in [-1,0]. \end{aligned}$$

In this model, the orthogonal decomposition

$$dW_t^S = \phi dW_t^\sigma + \sqrt{1-\phi^2} dW_t^{\sigma\perp}$$

implies that

$$\begin{aligned} Var[dW_t^S | dW_t^\sigma, I(t)] &= \left(1-\phi^2\right) dt, \\ E\left[\left.\frac{dS_t}{S_t}\right| dW_t^\sigma, I(t)\right] &= \vartheta_t dt + \phi \sigma_t dW_t^\sigma. \end{aligned} \tag{4.8}$$

Therefore, our two leverage effect parameters ξ and LEV, the discrete-time analogs of the impact of leverage on the instantaneous variance and the instantaneous expectation respectively in (4.8), must be connected by their common dependence on the instantaneous correlation between the return and the volatility factor. In other words, we must define a structural parameter ϕ in the discrete-time model such that it is well-suited to produce the analogs of $\left(1-\phi^2\right) dt$ and $\phi \sigma_t dW_t^\sigma$.

4.4.2 Instantaneous Correlation Between Return and Volatility Factor

Following $E[r_{t+1}\,|I^\sigma(t)] = \alpha'(0)\mu_{t+1} + \beta'(0)\mu_t + \gamma'(0)$ and $\psi = \alpha'(0)$, we have

$$Cov[r_{t+1}, \mu_{t+1}\,|I(t)] = Cov\{E[r_{t+1}|I^\sigma(t)], \mu_{t+1}|I(t)\} = \psi Var[\mu_{+1}|I(t)]$$

using $\mu_{t+1} \in I^\sigma(t) = I(t) \vee \{\mu_{t+1}\}$. Because $\mu_t = Var[r_{t+1}|I(t)]$, we have

$$Corr^2[r_{t+1}, \mu_{t+1}\,|I(t)] = \psi^2 \frac{Var^2[\mu_{+1}\,|I(t)]}{\mu_t Var[\mu_{t+1}\,|I(t)]} = \frac{\psi^2}{k_t^2} \tag{4.9}$$

with

$$k_t^2 = \frac{\mu_t}{Var[\mu_{t+1}\,|I(t)]} = \frac{E[RV_{t+1}\,|I(t)]}{\varkappa^2 Var[RV_{t+1}\,|I(t)]},$$

where the second equality follows from (4.1) for the HEAVY model. Han, Khrapov, and Renault (2020) document empirically that there is approximately a time-invariant proportionality between the two time series $E[RV_{t+1}\,|I(t)]$ and $Var[RV_{t+1}\,|I(t)]$ with observations on S&P500. Hence, we maintain hereafter the approximation that the coefficient k_t is constant:

$$k_t^2 = \frac{\mu_t}{Var[\mu_{t+1}\,|I(t)]} \approx \frac{E\left[\mu_t\right]}{E\left\{Var[\mu_{t+1}\,|I(t)]\right\}} = k^2.$$

For the $ARG(1)$ model in (4.6), we have

$$k^2 = \frac{1}{c\,(1+\rho)}$$

following

$$E\left[\mu_t\right] = \frac{\delta c}{1-\rho} \quad \text{and} \quad E\left\{Var[\mu_{t+1}\,|I(t)]\right\} = 2\rho c \frac{\delta c}{1-\rho} + \delta c^2.$$

Following (4.9), $\phi = \psi/k$ could be interpreted as a correlation coefficient similar to the correlation coefficient ϕ in the continuous time context (4.8). However, in discrete time, this assessment of the leverage effect is contaminated by the volatility feedback. Therefore, we write $\phi = LEV/k$ instead.[2] As such, we have

$$\begin{aligned} \xi &= 1-\phi^2, \quad \phi \le 0, \\ LEV &= k\phi, \quad k = \frac{1}{\sqrt{c\,(1+\rho)}}, \end{aligned} \tag{4.10}$$

and the identity (4.4) becomes

$$\psi = k\phi + \xi\left(\zeta_r - \frac{1}{2}\right). \tag{4.11}$$

Our focus of interest is the possible weakness of identification when the leverage effect is small, which is tantamount to a small absolute value of both LEV and ϕ. The bottom line is that the coefficient $\phi \in [-1, 0]$ is a structural parameter that encapsulates the two channels of leverage effect according to (4.10), which are discrete time analogs of (4.8).

4.5 Link Functions

Moment conditions in (4.6) and (4.7) clearly identify the reduced form parameters $\omega = (\omega_1, \omega_2)'$ with $\omega_1 = (\rho, c.\delta)'$ and $\omega_2 = (\gamma, \beta, \psi, \xi)'$. GMM estimation and inference about these reduced form parameters are discussed in the next section. Here, we provide the link function $g(\theta, \omega)$ that allows us to identify the structural parameters $\theta = (\zeta_r, \zeta_\sigma, \phi)'$ from the reduced form parameters ω. Given the true value of the reduced form parameters, denoted by ω^0, the true value of the structural parameters, denoted by θ^0, is the unique solution to the estimating equations $g\left(\theta, \omega^0\right) = 0$. Therefore, we can conduct inference of θ^0 by the minimum distance approach given a consistent estimator of ω.

The link function $g(\theta, \omega)$ consists of four estimating equations: the two equations provided by the arbitrage pricing, see Proposition 5 above, and the two equations (4.10) and (4.11) provided by the two channels of leverage effect. In this section, we consider the case of a small leverage effect, i.e., a small value of ϕ^0 in absolute value. We argue that, in this case, identification of the structural parameters is fragile and their inference is non-standard.

[2] By overlooking the volatility feedback effect as if $\psi = LEV = k\phi$, we have slightly twisted the interpretation of $k\phi$. The higher the leverage effect is, i.e., a larger absolute value of ϕ, the better is this interpretation of $k\phi$ following (4.11), since ξ is small in this case.

4.5.1 Two Estimating Equations Provided by Arbitrage Pricing

When making explicit the functions $a(\cdot)$ and $b(\cdot)$ following the $ARG(1)$ model, the two arbitrage identities of Proposition 5 can be written as

$$\begin{aligned} g_1(\theta,\omega) &= \tilde{\beta}\left\{1+c\left[\zeta_\sigma+\alpha\left(\zeta_r\right)\right]\right\}\left\{1+c\left[\zeta_\sigma+\alpha\left(\zeta_r-1\right)\right]\right\}+\rho LEV=0, \qquad (4.12)\\ g_2(\theta,\omega) &= \tilde{\gamma}-\delta\log\left\{1-c\frac{LEV}{1+c\left[\zeta_\sigma+\alpha\left(\zeta_r\right)\right]}\right\}=0, \end{aligned}$$

where

$$\begin{aligned} \tilde{\beta} &= \beta\left(\zeta_r\right)-\beta\left(\zeta_r-1\right),\\ \tilde{\gamma} &= \gamma\left(\zeta_r\right)-\gamma\left(\zeta_r-1\right),\\ LEV &= k\phi. \end{aligned}$$

Identification weakness for the structural parameter ζ_σ is obvious from (4.12) when the leverage effect parameter ϕ is close to zero. We see that $LEV = k\phi$ being close to zero implies that $\tilde{\beta}$ and $\tilde{\gamma}$ are close to zero. As discussed in Remark 2 below Proposition 5, $\tilde{\beta} = \tilde{\gamma} = 0$ implies that only ζ_r is identified (strongly in general) by

$$LEV = 0 = \alpha\left(\zeta_r\right)-\alpha\left(\zeta_r-1\right).$$

The bottom line is that among the three structural parameters $\zeta_r, \zeta_\sigma, \phi$, only ζ_r is strongly identified by the two estimating functions $g_1(\theta,\omega)$ and $g_2(\theta,\omega)$.

4.5.2 Two Estimating Equations Provided by Channels for Leverage Effect

The first estimating equation about the leverage effect comes from (4.10):

$$g_3(\theta,\omega) = \xi-(1-\phi^2) = 0, \qquad (4.13)$$

a relation between the reduced form parameter ξ and the structural parameter ϕ. A consistent estimator $\hat{\xi}$ of ξ can easily be obtained from estimating the parametric model that characterizes the joint dynamics of (r_{t+1}, μ_{t+1}), see the next section. As such, a simple consistent estimator of the structural parameter ϕ is $\check{\phi} = -\sqrt{1-\min(\hat{\xi},1)}$, where the negative sign is based on our prior knowledge that the leverage effect is negative. However, there are two nonstandard inference issues here if ϕ is close to 0, and we do not recommend using $\check{\phi}$ in practice. First, in the link function (4.13), the derivative with respect to ϕ is close to zero when ϕ is close to zero. Hence, similarly to situations described by Dufour (1997), ϕ can be arbitrarily poorly identified when the reduced form parameter ξ is arbitrarily close to 1. Second, asymptotic normality is a poor approximation to the finite-sample distribution of $\check{\phi}$ when ϕ is close to zero and it is near the boundary of the parameter space.

The second estimating equation about the leverage effect comes from (4.11):

$$g_4(\theta, \omega) = \psi - k\phi - \xi\left(\zeta_r - \frac{1}{2}\right). \tag{4.14}$$

Since ψ is a reduced form parameter that is strongly identified and ξ is not close to zero in the case of a small leverage effect, the estimating function $g_4(\theta, \omega)$ provides a strongly identifying relationship between structural parameters ϕ and ζ_r.[3]

5 Inference Robust to a Small Leverage Effect

In this section, we first consider the GMM estimation of the reduced-form parameters. Then, we plug this consistent GMM estimator of the reduced-form parameters in the link function and provide an robust confidence set for the structural parameters. We show that the proposed confidence set has correct asymptotic coverage uniformly over the parameter space that allows for both a large leverage effect and an arbitrarily small leverage effect. In particular, the confidence set is robust even if the structural parameters are weakly identified when the leverage effect is small.

5.1 Estimation of Reduced Form Parameters

As discussed above, we have two sets of reduced-form parameters that characterize the joint distribution of (r_{t+1}, μ_{t+1}): (i) $\omega_1 = (\rho, c, \delta)'$ that characterizes the conditional distribution of μ_{t+1} given $I(t)$, and (ii) $\omega_2 = (\gamma, \beta, \psi, \xi)'$, which jointly with ω_1, defines the conditional distribution of the return r_{t+1} given I_t^σ. As such, the reduced-form parameter $\omega = (\omega_1, \omega_2)'$ is identified by the conditional mean and variance of the volatility factor and the return. Specifically, following (4.6), the reduced-form parameter ω satisfies

$$\begin{aligned} E[u_{1,t+1}(\omega)|I(t)] &= 0, \text{ with } u_{1,t+1}(\omega) = \mu_{t+1} - [\rho\mu_t + c\delta], \quad (5.1) \\ E[u_{2,t+1}(\omega)|I(t)] &= 0, \text{ with } u_{2,t+1}(\omega) = \mu_{t+1}^2 - [\rho^2\mu_t^2 + 2\rho c(1+\delta)\mu_t + c^2\delta(1+\delta)], \end{aligned}$$

[3] We do acknowledge that this equation holds with some approximations due to the need to accommodate the volatility feedback in discrete time.

and, following (4.7),

$$
\begin{aligned}
E[\varepsilon_{1,t+1}(\omega)|I^{\sigma}(t)] &= 0, \text{ with } \varepsilon_{1,t+1}(\omega) = r_{t+1} - (\psi\mu_{t+1} + \beta\mu_t + \gamma), \\
E[\varepsilon_{2,t+1}(\omega)|I^{\sigma}(t)] &= 0, \text{ with } \varepsilon_{2,t+1}(\omega) = r_{t+1}^2 - \left(\psi\mu_{t+1} + \beta\mu_t + \gamma\right)^2 - Var[r_{t+1}|I^{\sigma}(t)],
\end{aligned}
\tag{5.2}
$$

where

$$
Var[r_{t+1}|I^{\sigma}(t)] = \xi\mu_{t+1} + \left[1 - \rho\xi - 2\rho c\psi^2\right]\mu_t - \left[\delta c^2\psi^2 + \xi\delta c\right].
$$

These conditional moments are based on the martingale differences sequences $u_{t+1}(\omega) = (u_{1,t+1}(\omega), u_{2,t+1}(\omega))'$ and $\varepsilon_{t+1}(\omega) = (\varepsilon_{1,t+1}(\omega), \varepsilon_{2,t+1}(\omega))'$. Furthermore, they form a set of sequential moments as in Ai and Chen (2012) since $u_{t+1}(\omega) \in I_t^{\sigma}$. It is obvious from (5.1) and (5.2) that the reduced-form parameter ω can be easily identified. Without loss of generality, we assume that they are strongly identified such that the standard GMM estimator has an asymptotic normal distribution.

We estimate ω based on the unconditional moment restrictions

$$
\begin{aligned}
E\left[h_t(\omega)\right] = 0, \text{ where } h_t(\omega) &= [h_{1,t}'(\omega), h_{2,t}'(\omega)]', \\
h_{1,t}(\omega) &= \left[x_{1,t} \otimes u_{1,t+1}(\omega_1), z_{1,t} \otimes u_{2,t+1}(\omega_1)\right]', \\
x_{1,t} &= (1, \mu_t), z_{1,t} = (1, \mu_t, \mu_t^2), \\
h_{2,t}(\omega) &= \left[x_{2,t} \otimes e_{1,t+1}(\omega), z_{2,t} \otimes e_{2,t+1}(\omega)\right]', \\
x_{2,t} &= (1, \mu_t, \mu_{t+1}), \;\; z_{2,t} = (x_{2,t}, \mu_t^2, \mu_t\mu_{t+1}, \mu_{t+1}^2).
\end{aligned}
\tag{5.3}
$$

Let $\hat{\omega}$ denote the efficient two-step GMM estimator based on the moments in (5.3). Let P denote the distribution of the data $\mathcal{W} = \{(r_t, \mu_t) : t \geq 1\}$ and $\mathcal{P}$ denote the parameter space of P. We make the following high-level assumption on the reduced-form parameters.

Assumption R. The following conditions hold uniformly over $P \in \mathcal{P}$. For some positive definite matrix Ω, its estimator $\hat{\Omega}$, and some fixed constant $0 < C < \infty$.
(i) $T^{1/2}\Omega^{-1/2}(\hat{\omega} - \omega) \to_d N(0, I)$.
(ii) $\hat{\Omega} - \Omega \to_p 0$.
(iii) $\lambda_{\min}(\Omega) \geq C^{-1}$ and $\lambda_{\max}(\Omega) \leq C$.

5.2 Identification-Robust Inference for Structural Parameters

The true values of the structural parameter $\theta = (\zeta_r, \zeta_\sigma, \phi)'$ and the reduced-form parameter ω satisfy the link function $g(\theta, \omega) = 0$, where this four-dimensional link function are given by the two arbitrage pricing conditions in (4.12) and the two channels for the leverage effect in (4.13) and (4.14). Once the reduced-form parameter ω is identified and estimated, we rely on this link function for the identification and inference of the structural parameter θ.

In a standard problem without any identification issues of the structural parameter, we can estimate θ by the minimum distance estimator and construct tests and confidence sets for θ using an asymptotically normal approximation for $T^{1/2}(\hat{\theta}-\theta)$. However, this standard method does not work here. This link function only provides weak identification of ζ_σ, the price of volatility risk, under a small leverage effect, as discussed in section 4.5. In this case, $g(\theta,\hat{\omega})$ is almost flat in ζ_σ and the minimum distance estimator of $\hat{\zeta}_\sigma$ may not even be consistent, see Stock and Wright (2000). To make the problem even more complicated, the inconsistency of $\hat{\zeta}_\sigma$ has a spillover effect on $\hat{\zeta}_r$ and $\hat{\phi}$, making their distribution non-normal even in large samples, as demonstrated in Andrews and Cheng (2012).

Let $g_0(\theta)$ denote the link function $g(\theta,\omega)$ evaluated at the true value of ω and $\hat{g}(\theta)$ denote its counterpart evaluated at the GMM estimator $\hat{\omega}$. Let $G(\theta,\omega)$ denote the partial derivative of $g(\theta,\omega)$ wrt ω, abbreviated as $G_0(\theta)$ when evaluated at the true value of ω and as $\hat{G}(\theta)$ when evaluated at $\hat{\omega}$.

We construct a confidence set for $\theta\in\Theta\coloneqq[-M_1,0]\times[0,M_2]\times[-1+\epsilon,0]$ by inverting the test $H_0:\theta=\theta_0$ vs $H_1:\theta\neq\theta_0$, where M_1 and M_2 are large positive constants and ϵ is a small positive constant. The test statistic is a QLR statistic that takes the form

$$QLR(\theta_0)=T\hat{g}(\theta_0)'\hat{\Sigma}(\theta_0,\theta_0)^{-1}\hat{g}(\theta_0)-\min_{\theta\in\Theta}T\hat{g}(\theta)'\hat{\Sigma}(\theta,\theta)^{-1}\hat{g}(\theta), \tag{5.4}$$

where $\hat{\Sigma}(\theta_1,\theta_2)=\hat{G}(\theta_1)\hat{\Omega}\hat{G}(\theta_2)'$ and $\hat{\Omega}$ is a consistent estimator of Ω.

Before presenting the robust confidence set based on the QLR statistic, we first introduce some useful quantities and provide a heuristic discussion of the identification problem and its consequences. Define

$$\eta_T(\theta)=T^{1/2}\left[\hat{g}(\theta)-g_0(\theta)\right]=G_0(\theta)\Omega^{1/2}\cdot\xi_T+o_p(1), \tag{5.5}$$

where $\xi_T\rightarrow_d N(0,I)$ following Assumption R. Thus, $\eta_T(\cdot)$ weakly converges to a Gaussian process $\eta(\cdot)$ with covariance function $\Sigma(\theta_1,\theta_2)=G_0(\theta_1)\Omega G_0(\theta_2)'$. Following (5.5), we can write $T^{1/2}\hat{g}(\theta)=\eta_T(\theta)+T^{1/2}g_0(\theta)$, where $\eta_T(\theta)$ is the noise from the reduced-form parameter estimation and $T^{1/2}g_0(\theta)$ is the signal from the link function. Under weak identification, $g_0(\theta)$ is almost flat in θ, modeled as the signal $T^{1/2}g_0(\theta)$ being finite even for $\theta\neq\theta_0$ and $T\rightarrow\infty$. Thus, the signal and the noise are of the same order of magnitude, yielding an inconsistent minimum distance estimator $\hat{\theta}$.

The identification strength of θ_0 is determined by the function $T^{1/2}g_0(\theta)$. However, this function is unknown and cannot be consistently estimated (due to $T^{1/2}$). Thus, we take the conditional inference procedure as in Andrews and Mikusheva (2016) and view $T^{1/2}g_0(\theta)$ as an infinite dimensional nuisance parameter for the inference of θ_0. The goal is to construct robust confidence set for θ_0 that has correct size asymptotically regardless of this unknown nuisance parameter.

Andrews and Mikusheva (2016) provide the conditional QLR test in a nonlinear GMM problem, where $\hat{g}(\theta)$ is replaced by a sample moment. The same method can be applied to the present nonlinear minimum distance problem. Following Andrews and Mikusheva (2016), we first project $\hat{g}(\theta)$ onto $\hat{g}(\theta_0)$ and construct a residual process

$$\hat{r}(\theta) = \hat{g}(\theta) - \hat{\Sigma}(\theta, \theta_0)\hat{\Sigma}(\theta_0, \theta_0)^{-1}\hat{g}(\theta_0). \tag{5.6}$$

The limiting distributions of $\hat{r}(\theta)$ and $\hat{g}(\theta_0)$ are Gaussian and independent. Thus, conditional on $\hat{r}(\theta)$, the asymptotic distribution of $\hat{g}(\theta)$ no longer depends on the nuisance parameter, $T^{1/2}g_0(\theta)$, making the procedure robust to any identification strength.

Specifically, we obtain the $1-\alpha$ conditional quantile of the QLR statistic, denoted by $c_{1-\alpha}(r, \theta_0)$, as follows. For $b = 1, \ldots, B$, we take independent draws $\eta_b^* \sim N(0, \hat{\Sigma}(\theta_0, \theta_0))$ and produce a simulated process,

$$g_b^*(\theta) = \hat{r}(\theta) + \hat{\Sigma}(\theta, \theta_0)\hat{\Sigma}(\theta_0, \theta_0)^{-1}\eta_b^*, \tag{5.7}$$

and a simulated statistic,

$$QLR_b^*(\theta_0) = T\eta_b^{*\prime}\hat{\Sigma}(\theta_0, \theta_0)^{-1}\eta_b^* - \min_{\theta\in\Theta} T g_b^*(\theta)'\hat{\Sigma}(\theta, \theta)^{-1}g_b^*(\theta). \tag{5.8}$$

Let $b_0 = \lceil(1-\alpha)B\rceil$, the smallest integer greater than or equal to $(1-\alpha)B$. Then the critical value $c_{1-\alpha}(r, \theta_0)$ is the b_0^{th} smallest value among $\{QLR_b^*, b = 1, \ldots, B\}$. Finally, we construct a robust confidence set for θ by collecting the null values that are not rejected, i.e., the nominal level $1-\alpha$ confidence set is

$$CS_T = \{\theta_0 : QLR_T(\theta_0) \leq c_{1-\alpha}(r, \theta_0)\}. \tag{5.9}$$

Assumption S. The following conditions hold over $P \in \mathcal{P}$, for any θ in its parameter space, and any ω in some fixed neighborhood around its true value, for some fixed $0 < C < \infty$.

1. $g(\theta, \omega)$ is partially differentiable in ω, with partial derivative $G(\theta, \omega)$ that satisfies $||G(\theta_1, \omega) - G(\theta_2, \omega)|| \leq C||\theta_1 - \theta_2||$ and $||G(\theta, \omega_1) - G(\theta, \omega_2)|| \leq C||\omega_1 - \omega_2||$.
2. $C^{-1} \leq \lambda_{\min}(G(\theta, \omega)'G(\theta, \omega)) \leq \lambda_{\max}(G(\theta, \omega)'G(\theta, \omega)) \leq C$.

Theorem 1. *Suppose Assumption R and Assumption S hold. Then,*

$$\liminf_{T\to\infty} \inf_{P\in\mathcal{P}} \Pr\left(\theta_0 \in CS_T\right) \geq 1-\alpha.$$

This theorem states that the confidence set constructed by the conditional QLR test has correct asymptotic size. Uniformity is important for this confidence set to cover the true parameter with a probability close to $1-\alpha$ in finite-samples. This uniform result is established over a parameter space $\mathcal{P}$ that allows for weak identification of the structural parameter θ. Therefore, the proposed confidence set is robust to a small leverage effect.

6 Simulations

In this section, we investigate the finite-sample performance of the proposed test and show that the asymptotic approximations derived above work well in practice. We also compare it with the standard test that assumes all parameters are strongly identified. The standard test is known to be invalid under weak identification but its degree of distortion is unknown in general. We simulate the data using the parametric model in subsection 4.3, where the true values of the parameters are given in Table 1 based on the values used by Han, Khrapov, and Renault (2020).[4] To investigate the robustness of the procedure with respect to various identification strengths, we vary both ϕ and T. Specifically, we consider $\phi \in \{-0.40, -0.01\}$ and $T \in \{2,000; 10,000\}$. The number of data points in the empirical section is approximately $5,200$ for comparison.

Table 1: Simulation Set-up

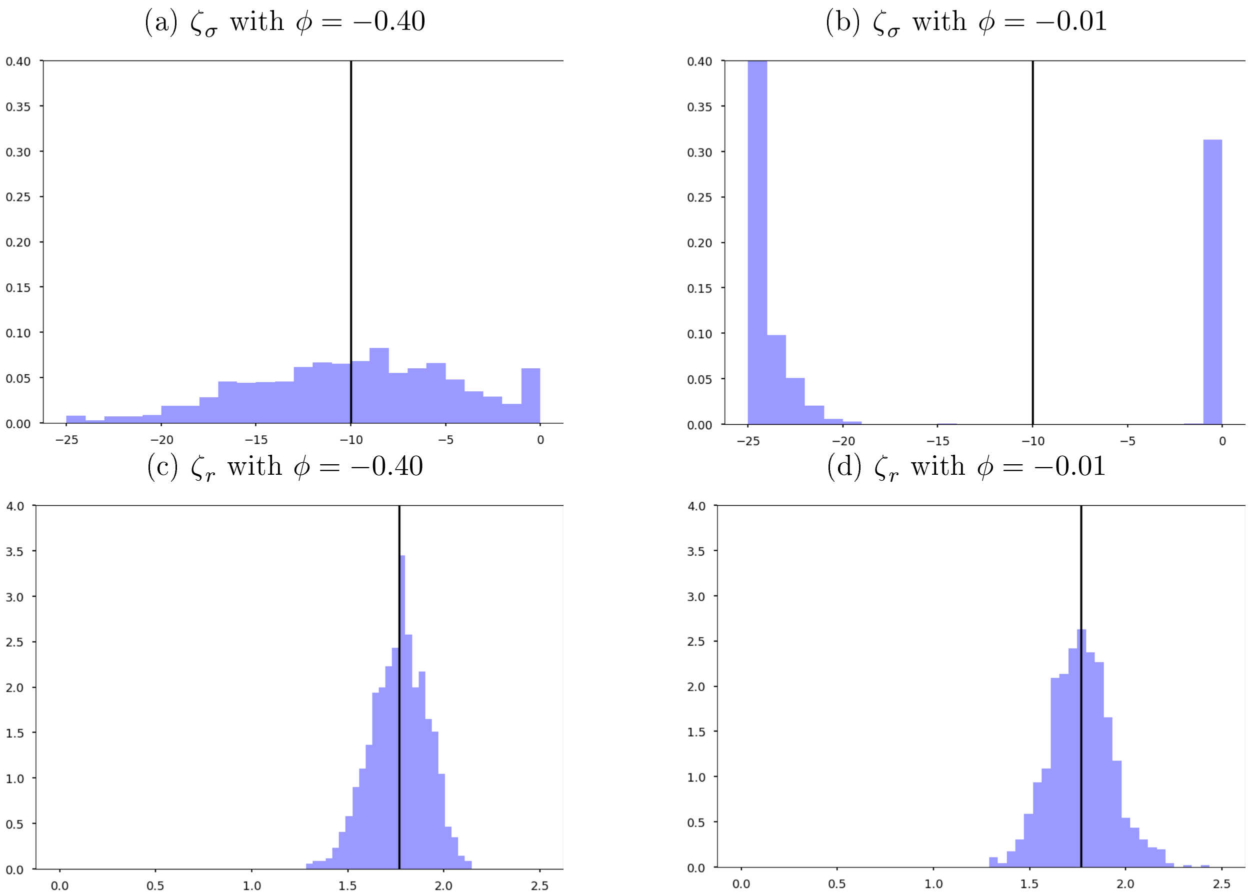
δ	ρ	c	ζ_σ	ζ_r
Parameter Values used by Han, Khrapov, and Renault (2020)				
0.65	0.50	3.94×10^{-3}	-10	1.77

To show the effect of various identification strengths, we first vary the true value of ϕ and plot the distribution of $\widehat{\zeta_r}$ and $\widehat{\zeta_\sigma}$ in Figure 1. The reported result is based on $10,000$ observations and 2000 simulation repetitions. The black lines in the middle of the figures are the true parameter values. Clearly, the estimators sometimes pile up at the boundaries

[4] To avoid boundary issues with respect to the estimate of c and δ in finite-sample, we reparameterize the moment conditions and link functions in terms of $\log(c)$, $\log(c)+\log(\delta)$, and $\text{logit}(\rho)$. This reparameterization forces the scale parameters to be positive and ρ to lie in $(0,1)$. We find that the resulting estimates for the transformed reduced-form parameters are better approximated by the Gaussian distribution for a given finite sample.

of the parameter space. As expected, this simulation shows that the Gaussian distribution is not a good approximation for the finite-sample distribution of either of the estimators, especially for ζ_σ.

Figure 1: Parameter Estimates' t-Statistics



(a) ζ_σ with $\phi = -0.40$

(b) ζ_σ with $\phi = -0.01$

(c) ζ_r with $\phi = -0.40$

(d) ζ_r with $\phi = -0.01$

Next, we study the finite-sample size of in the standard QLR test and the proposed conditional QLR test for a joint test for the three structural parameters. The nominal level of the test is 5 %. The critical value of the standard QLR test is the 95 % quantile of the χ^2-distribution with 3 degree of freedom. The critical value of the conditional QLR test is obtained by the stimulation-based procedure in section 5.2, with 1000 simulation repetitions to approximate the quantile of the conditional distribution. The finite-sample size is based on 1000 simulation repetitions.

Simulation results show that the standard QLR test under-rejects in finite-sample. This is most severe when the identification is weak. The proposed test, however, has finite-sample coverage probabilities close to the nominal level in all cases.

Table 2: Finite-Sample Size of the Standard and Proposed Tests

	Standard %	Proposed %	Standard %	Proposed %
ϕ	T = 2,000		T = 10,000	
-0.01	1.70	4.80	2.10	4.90
-0.40	3.70	5.70	3.90	4.00

7 Empirical Application

For the empirical application, we use the daily return on the S&P 500 for r_{t+1} and the associated realized volatility computed with high-frequency data for RV_{t+1}. The data is obtained from SPY (SPDR S&P 500 ETF Trust), an exchange-traded fund that mimics the S&P 500. This gives us a market index whose risk is not easily diversifiable and can be used to estimate the prices of risk that investors face in practice. We use the procedure Sangrey (2019) develops to estimate the integrated total volatility, i.e., the instantaneous expectation of the price variance. This measure reduces to the integrated diffusion volatility if prices have continuous paths and it works well in the presence of market microstructure noise.

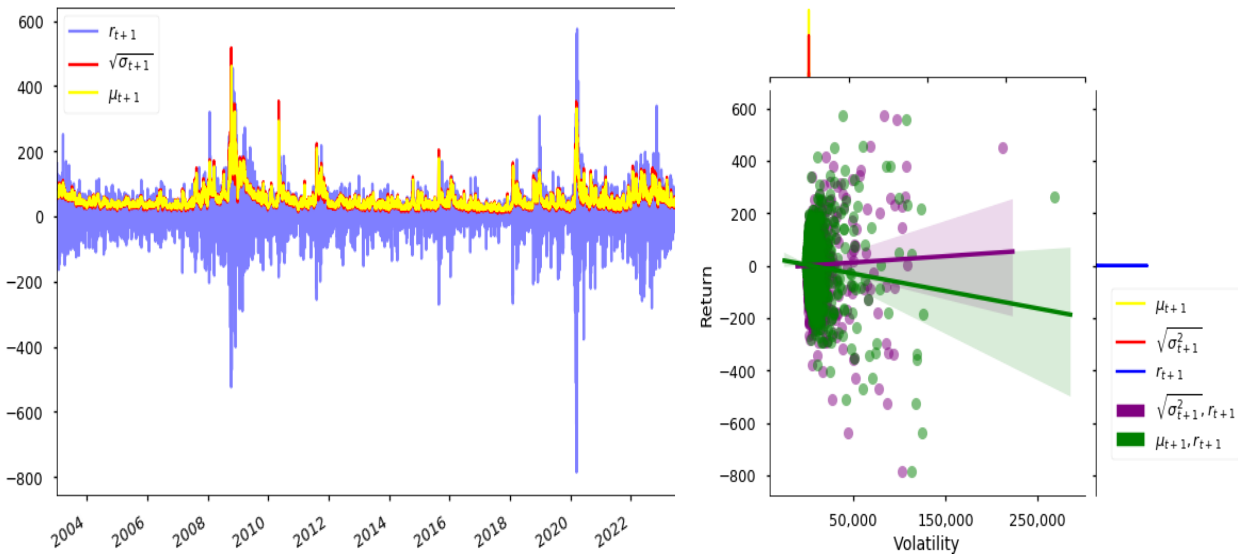
Since SPY is one of the most liquid assets traded, we can choose the frequency at which we sample the underlying price. To balance market-microstructure noise, computational cost, and efficiency of the resultant estimators, we sample at the 15-second frequency. We annualize the data by multiplying r_{t+1} by 252 and RV_{t+1} by 252^2. The data starts in 2003 and ends in June 2023. Since the asset is only traded during business hours, this leads to 5159 days of data with an average of approximately 160 observations per day. We compute r_{t+1} as the daily return from the open to the close of the market, the interval over which we can estimate the volatility. This avoids specifying the relationship between overnight and intra-day returns. We preprocess the data using the pre-averaging approach as in Podolskij and Vetter (2009) and Aït-Sahalia, Jacod, and Li (2012).

Once we have computed the realized volatility, we apply the HEAVY estimation procedure of Shephard and Sheppard (2010) to compute μ_{t+1}. We find $\varpi = 0.002$, $\lambda_1 = 0.669$, and $\lambda_2 = 0.312$. As we would expect, μ_{t+1} is very persistent, but not quite a unit root: $\lambda_1 + \lambda_2 = 0.981$. The original Shephard and Sheppard (2010) paper estimates $\lambda_1 = 0.564$ and $\lambda_2 = 0.417$. The increased weight on RV_{t+1} is likely due to a combination of having more recent data with substantially less market microstructure noise and a more precise estimate of RV_{t+1} due to Sangrey (2019).

To see how the data move over time, we plot their time series in Figure 2. We also plot the joint unconditional distribution in Figure 2 to see the static relationship between

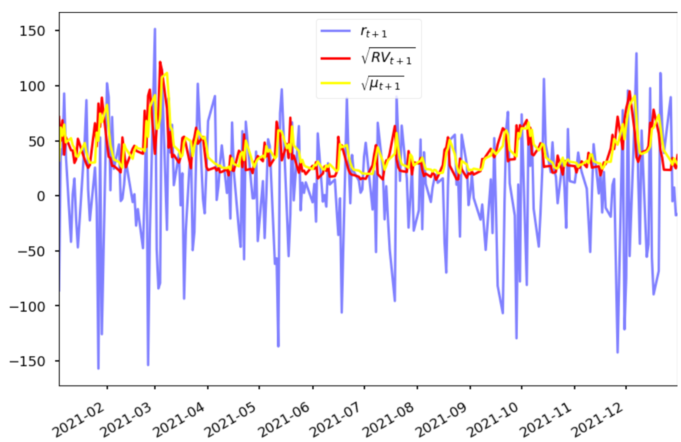
Figure 2: S&P 500 Volatility and Log-Return

(a) Time Series (b) Joint Distribution



the two series. The volatility has a long right tail, a typical gamma-type distribution. The returns have a bell-shaped distribution. The RV_{t+1} and r_{t+1} processes appear slightly negatively correlated (-0.09), as shown by the regression line in the joint plot, corroborating Bandi and Renò (2012) and Aït-Sahalia, Fan, and Li (2013). Conversely, the $\hat{\mu}_{t+1}$ and r_{t+1} processes appear slightly positively correlated (0.03), The theory predicts that the instantaneous correlation between shocks to μ_{t+1} and shocks to r_{t+1} is negative, but the lagged relationship is positive.

Figure 3: S&P 500 Volatility and Log Return



Because Figure 2 covers a long period and is therefore hard to see, we also zoom in and examine the behavior in 2021. We choose 2021 because it contains neither a lull nor huge spikes in volatility. The RV_{t+1} and $\hat{\mu}_{t+1}$ are strongly correlated, with μ_{t+1} lagging the behavior of RV_{t+1} as we would expect. The strong average correlation (0.80 over 2003-2023) of the two processes mitigates against potential misspecification from the HEAVY process because the vast majority of movement in $\hat{\mu}_{t+1}$ arises from variation in RV_{t+1} and $\hat{\mu}_{t+1}$ is only slightly more persistent than RV_{t+1}, which is consistent with the literature. We can exploit the different correlations to separately identify ϕ and the volatility risk price ζ_σ in an appropriately specified structural model. To better understand the volatility and return process we report a series of summary statistics in Table 3.

Table 3: Summary Statistics

	r_{t+1}	RV_{t+1}	μ_{t+1}
Mean	1.12	3698.59	3758.36
Standard Deviation	66.94	9098.23	8311.23
Skewness	-0.58	11.25	9.57
Kurtosis	16.81	205.43	9.57

We report the estimates and confidence intervals for the reduced-form parameters c, δ, and ρ. The confidence intervals reported here use the Gaussian limiting theory, i.e., the point estimates ± 1.96 standard errors.[5]

Table 4: Parameters that Govern the Volatility Process

	Point Estimate	95.00 % Confidence Interval
c	8.21	(4.84, 13.91)
δ	79.88	(46.47, 137.31)
ρ	0.81	(0.73, 0.87)

For confidence intervals of the three structural parameters, we first compute their joint confidence set based on the conditional QLR test and then project it to each of the components. We use 2000 simulations to compute the quantile for the QLR statistic.

The results in Table 5 have a few notable features. First, and most importantly, we reject the hypothesis that the price of volatility risk equals zero with the identification-robust test. The largest value lies within ζ_σ's confidence interval is -11.54. The data are unable to reject the combination of large negative values of ζ_σ and zero values of ζ_r. In other words, we cannot reject that the aversion to volatility is capable of explaining the entire equity premium. We can, however, reject the idea that the aversion to equity risk

[5] We first obtain confidence intervals for $\log(c)$ and $\log(c) + \log(\delta)$, and transform them into confidence intervals for c and δ. Similarly, we create the confidence interval for ρ by inverting the interval for $\text{logit}(\rho)$.

Table 5: Structural Parameters

	95 % Confidence Interval
ϕ	(-0.29 -0.33)
ζ_σ	(-23.85, -11.54)
ζ_r	(0.00, 0.33)

is capable of explaining the entire equity premium. Our confidence interval for ϕ and the upper bound for ζ_σ are consistent with findings in the literature.

8 Conclusion

It is commonly believed that option price data alone allows for the identification of risk aversion to volatility of volatility. We prove in this paper that, in the presence of a leverage effect, observations on the underlying asset return allow for identification without any option data. Moreover, in contrast with a widespread practice, the presence of a leverage effect implies that the variance risk premium does not separately identify the risk aversion to volatility of volatility and the standard risk aversion to volatility level.

While our general identification result is model-free, we provide a simple parametric model for the sake of numerical illustration of our identification robust strategy. This novel inference strategy applies the approach put forward by Andrews and Mikusheva (2016) in the context of GMM to minimum distance problems. This extension may be useful in other fields of structural econometrics. For the sake of clarity, our empirical illustration is developed in a simplified parametric model. This model could be enriched by more sophisticated dynamics of the volatility factor. An additional useful extension would be the introduction of two volatility factors, one for short term volatility and the other for long term volatility. This extension is natural because the economic uncertainty that we want to capture with volatility of volatility is well understood by the long run risk model (Bansal and Yaron, 2004).

Moreover, following an argument put forward by Bandi and Renò (2016), we suspect that our identification strategy based on the leverage effect would be more compelling with a model allowing for jumps both in return and volatility. While accommodating jumps in a discrete time model in general requires the Markov switching regime models, we could resort to the model proposed by Augustyniak, Bauwens, and Dufays (2019).

Of course, although our focus is the possibility of identification without data from the derivative markets, it does not mean that option price data should be wasted when they are available. They obviously in general allow us to obtain much tighter confidence sets for structural parameters, see e.g., the empirical results of Han, Khrapov, and Renault (2020). Our result must rather be understood as a possibility result. As in other contexts, the possibility of nonparametric identification of a structural model is worth studying even though empirical implementations generally resort to a parametric specification. It could be argued that our results are more robust to misspecification than those obtained with additional option data. For instance, it is well documented that stochastic volatility processes often feature some long memory, see e.g., Comte and Renault (1998). Although long memory is arguably observationally equivalent to structural breaks (or switching regimes) as far as underlying asset returns are concerned, option pricing is much more sensitive to the model choice. Our method avoids this potential model misspecification issue by using asset return data only for inference on the structural parameters.

Appendix

Proof of Proposition 3. By the Law of Iterated Expectations,

$$\begin{aligned}
\mathcal{L}_t(\zeta_r, \zeta) &= E[\exp\left\{-\zeta \mu_{t+1}\right\} \mathcal{L}_{(r|\sigma),t}(\zeta_r) \,|I(t)] \\
&= B\left(\zeta_r, I(t)\right) E[\exp\left\{-\zeta \mu_{t+1}\right\} A\left(\zeta_r, \mu_{t+1}\right) |I(t)] \,,
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{L}_t(\zeta_r - 1, \zeta) &= E[\exp\left\{-\zeta \mu_{t+1}\right\} \mathcal{L}_{(r|\sigma),t}(\zeta_r - 1) \,|I(t)] \\
&= B\left(\zeta_r - 1, I(t)\right) E[\exp\left\{-\zeta \mu_{t+1}\right\} A\left(\zeta_r - 1, \mu_{t+1}\right) |I(t)] \,.
\end{aligned}$$

From the condition in (2.8), we have

$$A\left(\zeta_r, \mu_{t+1}\right) = A\left(\zeta_r - 1, \mu_{t+1}\right).$$

Therefore,

$$\mathcal{L}_t(\zeta_r, \zeta) = \mathcal{L}_t(\zeta_r - 1, \zeta) \Longleftrightarrow B\left(\zeta_r, I(t)\right) = B\left(\zeta_r - 1, I(t)\right).$$

■

Proof of Proposition 4. We first consider the risk-neutral distribution of the volatility factor. The risk-neutral conditional Laplace transform of μ_{t+1} given $I(t)$ is

$$\mathcal{L}^*_{\sigma,t}(v) = \frac{\mathcal{L}_t(\zeta_r, v+\zeta_\sigma)}{\mathcal{L}_t(\zeta_r, \zeta_\sigma)}. \tag{A.1}$$

The numerator of (A.1) can be made explicit with the conditional Laplace transforms defined in (3.1) and (3.2):

$$\begin{aligned}
\mathcal{L}_t(\zeta_r, v+\zeta_\sigma) &= E[\exp\left(-\zeta_r r_{t+1} - (v+\zeta_\sigma)\mu_{t+1}\right)|I(t)] \\
&= E[\exp\left\{-(v+\zeta_\sigma)\mu_{t+1}\right\} E[\exp\left\{-\zeta_r r_{t+1}\right\}|I^\sigma(t)]\,|I(t)] \\
&= E[\exp\left\{-(v+\zeta_\sigma)\mu_{t+1}\right\} E[\exp\left\{-\alpha(\zeta_r)\mu_{t+1} - \beta(\zeta_r)\mu_t - \gamma(\zeta_r)\right\}|I(t)] \\
&= \exp\left\{-\beta(\zeta_r)\mu_t - \gamma(\zeta_r)\right\}\exp\left\{-a\left[v+\zeta_\sigma+\alpha(\zeta_r)\right]\mu_t - b\left[v+\zeta_\sigma+\alpha(\zeta_r)\right]\right\}.
\end{aligned}$$

We get the denominator of (A.1) by setting $v = 0$ and simplify the ratio to obtain

$$\begin{aligned}
\mathcal{L}^*_{\sigma,t}(v) &= \exp\left\{-a\left[v+\zeta_\sigma+\alpha(\zeta_r)\right]\mu_t + a\left[\zeta_\sigma+\alpha(\zeta_r)\right]\mu_t\right\} \\
&\quad \exp\left\{-b\left[v+\zeta_\sigma+\alpha(\zeta_r)\right] + b\left[\zeta_\sigma+\alpha(\zeta_r)\right]\right\}.
\end{aligned}$$

Analogous to (3.1), the risk-neutral conditional Laplace transform of the volatility factor can be written as

$$E^*[\exp\left(-v\mu_{t+1}\right)|I(t)] = \exp\left\{-a^*(v)\mu_t - b^*(v)\right\},$$

where

$$\begin{aligned}
a^*(v) &= a\left[v+\zeta_\sigma+\alpha(\zeta_r)\right] - a\left[\zeta_\sigma+\alpha(\zeta_r)\right], \\
b^*(v) &= b\left[v+\zeta_\sigma+\alpha(\zeta_r)\right] - b\left[\zeta_\sigma+\alpha(\zeta_r)\right].
\end{aligned}$$

Next, we consider the risk-neutral distribution of the return conditional on the volatility factor. We use the following Lemma.

Lemma A.1 *For all* (u, v),

$$\mathcal{L}^*_t(u,v) = E^*[\exp\left(-v\mu_{t+1}\right) g_t(u\left|\mu_{t+1}\right)|I(t)],$$

where

$$g_t(u\left|\mu_{t+1}\right) = \exp\left[-\alpha^*(u)\mu_{t+1} - \beta^*(u)\mu_t - \gamma^*(u)\right]$$

with functions $\alpha^*(\cdot), \beta^*(\cdot), \gamma^*(\cdot)$ *defined by Proposition 3.*

The conditional risk neutral distribution of the return given the volatility factor is defined by the decomposition (from the Law of Iterated Expectations, see (3.3) for the historical analog):

$$\mathcal{L}_t^*(u,v) = E^*[\exp\left\{-v\mu_{t+1}\right\}\mathcal{L}_{(r|\sigma),t}^*(u)\,|I(t)]\,. \tag{A.2}$$

Moreover (see e.g., argument in Bierens, 1982), this decomposition is unique. Hence, the result of the lemma allows us to conclude that

$$\mathcal{L}_{(r|\sigma),t}^*(u) = g_t(u\left|\mu_{t+1}\right.) = \exp\left[-\alpha^*(u)\,\mu_{t+1} - \beta^*(u)\,\mu_t - \gamma^*(u)\right].$$

■

Proof of Lemma A.1. We know from (2.3) that

$$\mathcal{L}_t^*(u,v) = \frac{\mathcal{L}_t(u+\zeta_r, v+\zeta_\sigma)}{\mathcal{L}_t(\zeta_r,\zeta_\sigma)} = \frac{N_t}{D_t}.$$

The numerator can be written as

$$\begin{aligned}
N_t &= E[\exp\left\{-(u+\zeta_r)r_{t+1} - (v+\zeta_\sigma)\mu_{t+1}\right\}|I(t)] \\
&= E[\exp\left\{-(v+\zeta_\sigma)\mu_{t+1}\right\}E[\exp\{-(u+\zeta_r)r_{t+1}\}\,|I^\sigma(t)]\,|I(t)] \\
&= \exp\{-\gamma(u+\zeta_r) - \beta(u+\zeta_r)\mu_t\}\,E[\exp\left\{-\left[v+\zeta_\sigma+\alpha\left(u+\zeta_r\right)\right]\mu_{t+1}\right\}|I(t)]\,.
\end{aligned}$$

Similarly, the denominator can be written as

$$D_t = \exp\{-\gamma(\zeta_r) - \beta(\zeta_r)\mu_t\}\,E[\exp\left\{-\left[\zeta_\sigma+\alpha\left(\zeta_r\right)\right]\mu_{t+1}\right\}|I(t)]\,.$$

By computing the ratio N_t/D_t, we get

$$\mathcal{L}_t^*(u,v) = \exp\{-\gamma^*(u) - \beta^*(u)\mu_t\}\,\frac{B_t}{C_t}$$

with

$$\begin{aligned}
B_t &= E[\exp\left\{-\left[v+\zeta_\sigma+\alpha\left(u+\zeta_r\right)\right]\mu_{t+1}\right\}|I(t)]\,, \\
C_t &= E[\exp\left\{-\left[\zeta_\sigma+\alpha\left(\zeta_r\right)\right]\mu_{t+1}\right\}|I(t)]\,.
\end{aligned}$$

Using the definition of $\alpha^*(\cdot)$, we can write

$$
\begin{aligned}
B_t &= E[\exp\left\{-\left[v+\zeta_\sigma+\alpha^*(u)+\alpha(\zeta_r)\right]\mu_{t+1}\right\}|I(t)] \\
&= \exp\left\{-a\left[v+\zeta_\sigma+\alpha^*(u)+\alpha(\zeta_r)\right]\mu_t\right\}\exp\left\{-b\left[v+\zeta_\sigma+\alpha^*(u)+\alpha(\zeta_r)\right]\right\}.
\end{aligned}
$$

By definition of functions $a^*(\cdot)$ and $b^*(\cdot)$,

$$
\begin{aligned}
a\left[v+\zeta_\sigma+\alpha^*(u)+\alpha(\zeta_r)\right] &= a^*\left[v+\alpha^*(u)\right]+a\left[\zeta_\sigma+\alpha(\zeta_r)\right], \\
b\left[v+\zeta_\sigma+\alpha^*(u)+\alpha(\zeta_r)\right] &= b^*\left[v+\alpha^*(u)\right]+b\left[\zeta_\sigma+\alpha(\zeta_r)\right].
\end{aligned}
$$

Hence,

$$
\begin{aligned}
B_t = &\exp\left\{-a^*\left[v+\alpha^*(u)\right]\mu_t-a\left[\zeta_\sigma+\alpha(\zeta_r)\right]\mu_t\right\} \\
&\exp\left\{-b^*\left[v+\alpha^*(u)\right]-b\left[\zeta_\sigma+\alpha(\zeta_r)\right]\right\}.
\end{aligned}
$$

For $u=v=0$, we get

$$
C_t=\exp\left\{-a\left[\zeta_\sigma+\alpha(\zeta_r)\right]\mu_t-b\left[\zeta_\sigma+\alpha(\zeta_r)\right]\right\}.
$$

By computing the ratio, we obtain

$$
\frac{B_t}{C_t}=\exp\left\{-a^*\left[v+\alpha^*(u)\right]\mu_t\right\}\exp\left\{-b^*\left[v+\alpha^*(u)\right]\right\}.
$$

Therefore, we have shown that

$$
\begin{aligned}
\mathcal{L}_t^*(u,v) &= \exp\left\{-\gamma^*(u)-\beta^*(u)\mu_t\right\}\exp\left\{-a^*\left[v+\alpha^*(u)\right]\mu_t\right\}\exp\left\{-b^*\left[v+\alpha^*(u)\right]\right\} \\
&= E^*[\exp\left(-v\mu_{t+1}\right)g_t(u\left|\mu_{t+1}\right.)\,|I(t)],
\end{aligned}
$$

where

$$
g_t(u\left|\mu_{t+1}\right.)=\exp\left[-\alpha^*(u)\,\mu_{t+1}-\beta^*(u)\,\mu_t-\gamma^*(u)\right].
$$

This proves the desired result in the lemma. ■

Proof of Proposition 5. From (3.2) and (3.3), we have

$$
\begin{aligned}
\mathcal{L}_t(u,v) &= E[\exp\left\{-v\mu_{t+1}\right\}\exp\left\{-\alpha(u)\,\mu_{t+1}-\beta(u)\,\mu_t-\gamma(u)\right\}|I(t)] \\
&= \exp\left\{-\beta(u)\,\mu_t-\gamma(u)\right\}\exp\left\{-a\left[v+\alpha(u)\right]\mu_t\right\}\exp\left\{-b\left[v+\alpha(u)\right]\right\}.
\end{aligned}
$$

Therefore, the condition

$$
\mathcal{L}_t(\zeta_r,\zeta_\sigma)=\mathcal{L}_t(\zeta_r-1,\zeta_\sigma)
$$

is equivalent to the conjunction of two conditions

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

■

References

Ai, C., and X. Chen (2012): "The Semiparametric Efficiency Bound for Models of Sequential Moment Restrictions Containing Unknown Functions," *Journal of Econometrics*, 170(2), 442–457, Thirtieth Anniversary of Generalized Method of Moments.

Aït-Sahalia, Y., J. Fan, and Y. Li (2013): "The Leverage Effect Puzzle: Disentangling Sources of Bias at High Frequency," *Journal of Financial Economics*, 109(1), 224–249.

Aït-Sahalia, Y., J. Jacod, and J. Li (2012): "Testing for Jumps in Noisy High Frequency Data," *Journal of Econometrics*, 168(2), 207–222.

Ait-Sahalia, Y., and A. Lo (2000): "Nonparametric Risk Management and Implied Risk Aversion," *Journal of Econometrics*, 94(1-2), 9–51.

Andrews, D. W. K., and X. Cheng (2012): "Estimation and Inference With Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80(5), 2153–2211.

Andrews, D. W. K., and P. Guggenberger (2010): "Asymptotic Size and a Problem with Subsampling and with the M out of N Bootstrap," *Econometric Theory*, 26(2), 426–468.

Andrews, I., and A. Mikusheva (2016): "Conditional Inference With a Functional Nuisance Parameter," *Econometrica*, 84(4), 1571–1612.

Antoine, B., and E. Renault (2012): "Efficient Minimum Distance Estimation with Multiple Rates of Convergence," *Journal of Econometrics*, 170(2), 350–367, Thirtieth Anniversary of Generalized Method of Moments.

Ané, T., and H. Geman (2000): "Order Flow, Transaction Clock, and Normality of Asset Returns," *The Journal of Finance*, 55(5), 2259–2284.

Augustyniak, M., L. Bauwens, and A. Dufays (2019): "A New Approach to Volatility Modeling: The Factorial Hidden Markov Volatility Model," *Journal of Business & Economic Statistics*, 37(4), 696–709.

Bandi, F. M., and R. Renò (2012): "Time-varying Leverage Effects," *Journal of Econometrics*, 169(1), 94–113.

——— (2016): "Price and Volatility Co-jumps," *Journal of Financial Economics*, 119(1), 107–146.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *The Journal of Finance*, 59(4), 1481–1509.

Bierens, H. J. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20(1), 105–134.

Black, F. (1976): "Studies of Stock Price Volatility Changes.," *In: Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association,Washington DC*, pp. 177–181.

Bollerslev, T., J. Litvinova, and G. Tauchen (2006): "Leverage and Volatility Feedback Effects in High-Frequency Data," *The Journal of Financial Econometrics*, 4(3), 353–384.

Bollerslev, T., G. Tauchen, and H. Zhou (2009): "Expected Stock Returns and Variance Risk Premia," *The Review of Financial Studies*, 22(11), 4463–4492.

Britten-Jones, M., and A. Neuberger (2000): "Option Prices, Implied Price Processes, and Stochastic Volatility," *The Journal of Finance*, 55(2), 839–866.

Brownlees, C., and G. Gallo (2010): "Comparison of Volatility Measures: a Risk Management Perspective," *The Journal of Financial Econometrics*, 8(1), 29–56.

Chabi-Yo, F., R. Garcia, and E. Renault (2008): "State Dependence Can Explain the Risk Aversion Puzzle," *The Review of Financial Studies*, 21(2), 973–1011.

Chernov, M., A. Gallant, E. Ghysels, and G. Tauchen (2003): "Alternative models for stock price dynamics," *Journal of Econometrics*, 116(1-2), 225–257.

Christoffersen, P., S. Heston, and K. Jacobs (2009): "The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well," *Management Science*, 55(12), 1914–1932.

——— (2013): “Capturing Option Anomalies with a Variance-Dependent Pricing Kernel,” *The Review of Financial Studies*, 26(8), 1963–2006.

Clark, P. (1973): “A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices,” *Econometrica*, 41(1), 135–55.

Comte, F., and E. Renault (1998): “Long Memory in Continuous-time Stochastic Volatility Models,” *Mathematical Finance*, 8(4), 291–323.

Corsi, F., N. Fusari, and D. La Vecchia (2013): “Realizing Smiles: Options Pricing with Realized Volatility,” *Journal of Financial Economics*, 107(2), 284–304.

Darolles, S., C. Gouriéroux, and J. Jasiak (2006): “Structural Laplace Transform and Compound Autoregressive Models,” *Journal of Time Series Analysis*, 27(4), 477–503.

Drechsler, I., and A. Yaron (2011): “What’s Vol Got to Do with It,” *The Review of Financial Studies*, 24(1), 1–45.

Duffie, D., J. Pan, and K. Singleton (2000): “Transform Analysis and Asset Pricing for Affine Jump-Diffusions,” *Econometrica*, 68(6), 1343–1376.

Dufour, J.-M. (1997): “Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models,” *Econometrica*, 65(6), 1365–1388.

Feller, W. (1951): “Two Singular Diffusion Problems,” *Annals of Mathematics*, 54(1), 173–182.

Garcia, R., and E. Renault (1998): “A Note on Hedging in ARCH and Stochastic Volatility Option Pricing Models,” *Mathematical Finance*, 8(2), 153–161.

Gouriéroux, C., and J. Jasiak (2006): “Autoregressive Gamma Processes,” *Journal of Forecasting*, 25(2), 129–152.

Han, H., S. Khrapov, and E. Renault (2020): “The Leverage Effect Puzzle Revisited: Identification in Discrete Time,” *Journal of Econometrics*, 217(2), 230–258, Nonlinear Financial Econometrics.

Heston, S. L. (1993): “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options,” *The Review of Financial Studies*, 6(2), 327–343.

Heston, S. L., and S. Nandi (2000): “A Closed-Form GARCH Option Valuation Model,” *The Review of Financial Studies*, 13(3), 585–625.

Jackwerth, J. C. (2015): “Recovering Risk Aversion from Option Prices and Realized Returns,” *The Review of Financial Studies*, 13(2), 433–451.

Magnusson, L. M. (2010): “Inference in Limited Dependent Variable Models Robust to Weak Identification,” *The Econometrics Journal*, 13(3), S56–S79.

Magnusson, L. M., and S. Mavroeidis (2010): “Identification-Robust Minimum Distance Estimation of the New Keynesian Phillips Curve,” *Journal of Money, Credit and Banking*, 42(2-3), 465–481.

Meddahi, N., and E. Renault (2004): “Temporal Aggregation of Volatility Models,” *Journal of Econometrics*, 119(2), 355–379.

Melino, A., and A. X. Yang (2003): “State-dependent preferences can explain the equity premium puzzle,” *Review of Economic Dynamics*, 6(4), 806–830, Finance and the Macroeconomy.

Mikusheva, A. (2007): “Uniform Inference in Autoregressive Models,” *Econometrica*, 75(5), 1411–1452.

Pan, J. (2002): “The Jump-risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study,” *Journal of Financial Economics*, 63(1), 3–50.

Pierce, D. A., and L. D. Haugh (1977): “Causality in Temporal Systems: Characterization and a Survey,” *Journal of Econometrics*, 5(3), 265–293.

Podolskij, M., and M. Vetter (2009): “Bipower-type Estimation in a Noisy Diffusion Setting,” *Stochastic Processes and Their Applications*, 119(9), 2803–2831.

Renault, E. (1997): “Econometric Models of Option Pricing Errors,” *Econometric Society Monographs*, 28, 223–278.

Sangrey, P. (2019): “Uncertainty and Learning in Dynamic Financial Econometrics,” Ph.D. thesis, University of Pennsylvania.

Shephard, N., and K. Sheppard (2010): “Realising the Future: Forecasting with High-frequency-based Volatility (HEAVY) models,” *Journal of Applied Econometrics*, 25(2), 197–231.

Stock, J. H., and J. H. Wright (2000): “GMM with Weak Identification,” *Econometrica*, 68(5), 1055–1096.

Veronesi, P. (2001): "Belief Dependent Utilities, Aversion to State-Uncertainty and Asset Prices," CEPR Discussion Papers 2965, C.E.P.R. Discussion Papers.

Ziegler, A. (2007): "Why Does Implied Risk Aversion Smile?," *The Review of Financial Studies*, 20(3), 859–904.