HEC MONTRÉAL
École affiliée à l'Université de Montréal

On Time-Varying Volatility and Financial Derivatives

par
Hugo Lamarre

Thèse présentée en vue de l'obtention du grade de Ph. D. en administration
(option Ingénierie financière)

Décembre 2018

HEC MONTRÉAL
École affiliée à l'Université de Montréal

Cette thèse intitulée :

On Time-Varying Volatility and Financial Derivatives

Présentée par :

Hugo Lamarre

a été évaluée par un jury composé des personnes suivantes :

Pascal François
HEC Montréal
Président-rapporteur

Debbie J. Dupuis
HEC Montréal
Directrice de recherche

Bruno Rémillard
HEC Montréal
Codirecteur de recherche

Nicolas A. Papageorgiou
HEC Montréal
Membre du jury

Lars Stentoft
Western Ontario
Examinateur externe

Gilles Caporossi
HEC Montréal
Représentant du directeur de HEC Montréal

Résumé

Nous considérons le risque de variance en temps discret dans le contexte de produits dérivés sur indices d'actions. Nous effectuons deux études dont l'une est de nature empirique et l'autre est de nature numérique. La première est motivée par les préférences en termes de risque de variance observées sur le marché des options. Nous proposons une approche de couverture utilisant les rendements *hautes fréquences* et démontrons qu'elle améliore la performance ajustée pour le risque d'une stratégie d'investissement. Deuxièmement, nous proposons une approche de *quantisation conditionnelle* qui approxime les dynamiques de marché sous risque de variance. La quantisation est un sous-champ des mathématiques probabilistes permettant la discrétisation des états d'un processus stochastique. L'algorithme proposé est numériquement efficace, particulièrement lorsque la dépendance entre la variance et le prix est forte. Nous démontrons le fort potentiel numérique de l'approche pour des applications à grand déploiement et/ou en temps réel.

Les deux premiers chapitres étudient la valeur économique pouvant être extraite de prévisions de volatilité dans le contexte d'une exposition prolongée aux primes de risque contenues dans les options sur l'indice S&P 500. Nous considérons un investisseur qui vend une option, s'engage à la laisser expirer et réajuste de manière journalière des positions dans l'indice sous-jacent dans le but de minimiser sa variance terminale. Nous estimons la valeur incrémentale d'utiliser des modèles de prévision basés sur des données journalières versus hautes fréquences. Dans ce dernier cas, la stratégie proposée est une contribution novatrice à la littérature portant sur la couverture en temps discret. Le test empirique proposé est robuste aux effets de maturité et de prix d'exercice, à un test hors-échantillon et à différents types de préférences. Entre 2002 et 2014, nous trouvons une valeur incrémentale *positive* provenant de l'utilisation de données hautes fréquences. Tous les protocoles de couverture proposés performent mieux qu'une approche non-paramétrique de type couverture-delta couramment utilisée en pratique.

Le troisième chapitre examine l'application de la quantisation aux modèles GARCH. Notre objectif est une approximation des dynamiques de marché à l'aide de chaînes de Markov à temps inhomogène et à états discrets. Certaines particularités des lois de probabilité GARCH peuvent s'avérer problématiques dans ce contexte. Nous considérons des algorithmes stochastiques et déterministiques. Dans le premier cas, nous standardisons la variance et le prix à cause d'une différence d'échelle. Dans le cas déterministique, nous considérons des produits cartésiens de quantisations à une dimension. Cette approche donne lieu à des algorithmes d'optimisation convexe très rapides. En général, nous démontrons que l'approche déterministique est beaucoup plus stable et efficace pour un grand nombre d'états discrets. Cette efficacité nous permet d'atteindre des niveaux de précision inaccessibles (dans un temps raisonnable) sous l'approche stochastique. Étant donnée une quantisation optimale, le calcul du prix d'une option peut être effectué à très bas coût. Nos résultats suggèrent que la quantisation est appropriée pour des applications qui sont trop exigeantes pour l'approche Monte Carlo telles que le courtage d'options à haute fréquence.

Mots-clés

GARCH, Stratégies de Couverture, Marché Incomplet, Prédictions de Volatilité, Volatilité Réalisée, Prime de Risque de Volatilité, Gestion des Risques, Quantisation, Programmation Dynamique

Méthodes de recherche

Méthodes quantitatives

Abstract

We present an empirical study and a numerical investigation of time-varying volatility and discrete-time models in the context of financial derivatives written on equity indices. Firstly, acknowledging empirical facts concerning aggregate variance risk preferences for the S&P 500 index, we investigate a set of profitable investment opportunities for exchange-traded vanilla options. We propose a protocol relying on *high-frequency* returns which yields positive incremental economic value i.e. in a risk-adjusted sense. Secondly, we turn to a class of model approximations known as quantizations, which improves the numerical efficiency of many recursive algorithms related to financial derivatives. We propose a *conditional quantization* which better accommodates strong price-variance dependence effects than existing methodologies. The proposed approach is applied to S&P 500 index vanilla options in a stylized numerical experiment and displays strong numerical potential for large scale and/or real-time applications.

More precisely, we first examine the economic value of volatility timing when hedging for a profit-oriented agent who sells and hedges options on a daily basis. A held-until-maturity hypothesis allows us to depart from classical delta-hedging towards variance-optimal hedging, which offers a more viable test for ranking volatility forecasts. In particular, we estimate the incremental value of hedging under volatility models based on low-frequency (i.e. daily) versus high-frequency (i.e. five minutes) data. In the latter case, the proposed methodology is a novel non-myopic approach to hedging contingent claims in incomplete markets using realized variance. Our empirical focus is on out-of-the-money put options sold unconditionally from 2002 to 2014. We find positive incremental economic value from relying on high-frequency data. Our conclusions are robust to model specifications, moneyness-maturity effects, an out-of-sample exercise and prospect-theory preferences. All proposed variance-optimal hedges under time-varying volatility significantly outperform model-free delta-hedges often used by practitioners.

We then investigate quantization methods when applied to GARCH models towards approximating price-variance dynamics by a time-inhomogeneous discrete-state Markov chain. The special nature of GARCH probability supports may be problematic in this context. Both stochastic and deterministic methods are considered. In the former case, a standardized distortion function is minimized over $\mathbb{R}^2$ using a stochastic gradient descent. Standardization is critical due to an inherent gap in price-variance scale. In the deterministic case, we focus on Cartesian products of component-wise quantizations. This procedure —commonly known as *product* quantization— involves fast convex optimizations. Our numerical study shows deterministic methods are more reliable and efficient for large quantizers. Once proposed quantizations are obtained, option prices and variance-optimal hedge ratios are computed at very low cost, making the approach suitable for computationally challenging applications where Monte Carlo fails, such as high-frequency option trading.

Keywords

GARCH, Hedging, Market Incompleteness, Volatility Timing, Realized Volatility, Variance Risk Premium, Risk Management, Quantization, Dynamic Programming

Research methods

Quantitative methods

Contents

List of Tables

List of Figures

General Introduction

The seminal work of Black and Scholes (1973) (B&S) exposes the relationship between financial derivatives and underlying asset risk. It does so in the most elementary framework i.e. under *constant* volatility. Time-varying volatility, however, is a natural feature of financial markets. Following the dramatic fall of Lehman Brothers during the last financial crisis, most market participants likely anticipated stressed market conditions would endure for some time before returning to normal levels. Many statistical models have been proposed for formalizing this behavior, mostly stemming from either the discrete-time GARCH model of Bollerslev (1986) or the continuous-time stochastic volatility model of Heston (1993).

The most important contribution of B&S remains the axiomatic independence of financial derivatives to underlying asset return expectations, commonly known as *drifts*. This independence is unfortunately premised on a very strong set of assumptions pertaining to market *completeness* i.e. the ability to dynamically replicate financial derivatives using the underlying asset (and a bond).

This ability invariably presupposes infinitely frequent transactions in the underlying market, which are prevented by transaction costs in practice. Time-varying volatility is an additional source of market *incompleteness*, due to volatility being an untradable mathematical construct. Perfect replication is thus unavailable in reality and option market participants must account for *un-hedgeable* risk factors, e.g. discrete-time trading, time-varying volatility, jumps and model misspecification.

The redistribution of such risks amongst option market participants explains the striking success of derivatives markets; the Chicago Board Options Exchange [CBOE] has annual trading volumes above 1 billion contracts and aggregate notional values for the S&P 500 index are typically above 4 trillion USD. For example, a pension plan may participate in the option market for the protection provided by put options against jump risk, as this protection can *not* be acquired in the underlying market alone.

Assuming perfect replication is available, investors are indifferent to holding self-financing replication portfolios or actual derivatives. From absence of arbitrage, portfolio *values* thus equate derivatives *prices*. Harrison and Pliska (1981) show prices may be formulated as mathematical expectations under a market model with no propensity for either positive or negative returns, i.e. under a martingale or *risk-neutral* measure.

Several facts explain why both researchers and practitioners make do with complete market frameworks. Risk neutral prices do not explicitly depend on drifts, which are challenging to forecast over short horizons. Complete market frameworks are typically more tractable than their incomplete market counterparts, often yielding analytical pricing and hedging expressions. Under B&S, the so-called *implicit volatility* derived from market observations allows practitioners to efficiently communicate prices. The replication strategy under complete markets —commonly known as *delta-hedging*— is model-free; see e.g. Alexander and Nogueira (2007).

To circumvent markets being incomplete in reality, researchers often *first* posit preferences holding at the market equilibrium for a representative agent and *then* infer a replication protocol from resulting option prices —which we henceforth refer to as *preference-based pricing*.

The empirical success of a preference-based pricing model is unsurprisingly related to its flexibility. Most successful works allow for a parameterized family of risk-neutral measures and rely on parameter estimates from historical datasets including option prices; see e.g. Bates (2003) for a related discussion. For example, Heston (1993) allows for a price of volatility risk parameter, while Christoffersen et al. (2013) allows for a volatility aversion parameter. Too much flexibility, however, may create identifiability issues; see Branger and Schlag (2008).

While the preference-based approach is motivated by market equilibria, utility functions and stochastic discount factors in the spirit of Rubinstein (1976), hedged and risk-averse participants likely assess their ability to replicate an option *prior* to determining its value. In other words, deriving protocols from prices somewhat inverts the natural causality.

Unfortunately, replication is often viewed as a second-tier concept, which rarely merits empirical inquiries of its own. A search for "option pricing" in the Web of Science database yielded roughly 20,000 results, as opposed to 2,500 results for "option hedging" —most of which were actually focused on option pricing.

The striking success of risk-neutral pricing appears to have further prompted researchers and practitioners to overlook (or have little interest in) drifts in all derivatives matters. But dynamics of un-hedged option prices *implicitly* depend on drifts, as a direct consequence of future derivatives prices being a function of future underlying prices. The first relevant option return investigation was tardily proposed by Coval and Shumway (2001) in the context of the capital asset pricing model (CAPM); also see Broadie et al. (2009) and references therein.

For equity indices, volatility dynamics being negatively correlated to levels is a well-acknowledged empirical fact, classically attributed to a *leverage effect* in firms balance sheet; see e.g. Campbell and Hentschel (1992). Since (1) investors dislike scenarios for which the S&P 500 index is low and (2) the index is negatively related to volatility, we may reasonably expect investors to dislike scenarios for which volatility is high.

Since (1) volatility acts as a negative beta asset in the CAPM and (2) options are positively related to volatility, holding an option has a positive cost in the long-run, similarly to insurance. This cost can *not* be explained by leveraged market exposure under the CAPM and is attributed to a *volatility risk premium* by Bakshi and Kapadia (2003). The existence of such a premium is consistent with the fast and sustained growth in volatility-related investment products, such as over-the-counter variance swaps.

Alexander and Nogueira (2007) show delta-hedging protocols are not variance-optimal under leverage effects; see also Garcia and Renault (1998). Preference-based pricing under time-varying volatility hence fails to provide a truly unified pricing and hedging framework. Schweizer (1996) somewhat solves this issue in continuous-time, but fails in discrete-time due to signed pricing measures. In our opinion, Gârleanu et al. (2009) propose the first economically satisfying and empirically successful resolution by considering market segmentation.

Overall, market incompleteness as generated by time-varying volatility and discrete-time trading offers a rich and highly relevant framework. It is the focus of the present thesis. Our contribution to this framework is both empirical and numerical. The first and second chapters present an empirical study on the economic value of volatility forecasts for profit-oriented agents who sell and hedge S&P 500 index options. The third chapter investigates numerical methodologies, which may be used to solve hedging protocols more efficiently.

The first chapter is motivated by the following facts: (1) selling and hedging S&P 500 options is risky *and* profitable in the long-run, (2) hedging protocols are impacted by volatility views, and (3) volatility forecasting methodologies have varying degrees of statistical success. A large strand of the literature is indeed dedicated to improving volatility forecasts under the family of GARCH models. For example, recent advances proposed by Shephard and Sheppard (2010) show significant statistical value from considering high-frequency datasets.

A natural question arising in this context is whether better volatility forecasts (in a statistical sense) translate to better risk-adjusted returns when selling and hedging options. We build an empirical test towards answering this question and *quantifying* incremental economic values from statistical improvements in forecasts.

The proposed test isolates the impact of volatility timing abilities in a manner that is robust to model specifications. In particular, we work under a martingale constraint which prohibits market timing abilities i.e. drift forecasts. We also refrain from distributional assumptions about model innovations.

We empirically find significant incremental economic value from considering volatility estimators based on high-frequency data —commonly known as *realized variances*.

The second chapter contains methodological details pertaining to the proposed empirical test, such as data cleaning procedures and robustness checks.

The first two chapters highlight numerical challenges which arise from market incompleteness. We indeed resort to Monte Carlo for recursively solving conditional expectations emerging from the *dynamic programming principle* when building hedging protocols. Explicitly solving such expectations is particularly challenging under time-varying volatility in discrete-time as (1) time-discreteness prevents us from relying on the local behavior of the solution (e.g. using the Feynman-Kac formula) and (2) price-volatility *continuum* requires us to search in a large space of two-dimensional *functions*. Many financial problems such as portfolio allocation or American option pricing behave similarly. Attempts to guess an analytical solution are usually doomed to fail.

The third chapter is a numerical endeavor which investigates GARCH dynamics approximations given by time-inhomogeneous discrete-state Markov chains. The theoretical foundation is provided by quantization theory, which may be traced back to information theory introduced in the late 1950s by Bell Laboratories for approximating electrical signals; see e.g. Lloyd (1982). This theory was only recently applied to the field of numerical probability by Pagès (1997), allowing for many interesting financial applications, such as American option pricing by Bally et al. (2005).

While quantization has already been applied to other market models, GARCH dynamics pose specific challenges. For example, the two-dimensional Euclidean norm is flawed, due to conditional variances being several orders of magnitude smaller than log-prices. One-day ahead GARCH probability supports are also very peculiar, which could pose additional challenges not met by typical stochastic volatility models.

We propose novel *conditional* quantizations inspired by product quantizations of Fiorin et al. (2017). We show the high relative numerical efficiency of the proposed conditional approach in three stylized settings for the S&P 500 index, namely European and American option pricing and variance-optimal hedging. The proposed approach yields levels of accuracy comparable to existing pricing benchmarks. Preliminary results suggest the approach is suitable for numerically challenging applications, such as high-frequency option trading or real-time option data streaming.

References

Alexander, C. and Nogueira, L. M. (2007). Model-free hedge ratios and scale-invariant models. *Journal of Banking and Finance*, 31(6):1839–1861.

Bakshi, G. and Kapadia, N. (2003). Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies*, 16(2):527–566.

Bally, V., Pagès, G., and Printems, J. (2005). A quantization tree method for pricing and hedging multidimensional American options. *Mathematical Finance*, 15(1):119–168.

Bates, D. S. (2003). Empirical option pricing: A retrospection. *Journal of Econometrics*, 116(1):387–404.

Bates, D. S. (2005). Hedging the smirk. *Finance Research Letters*, 2(4):195–200.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Branger, N. and Schlag, C. (2008). Can tests based on option hedging errors correctly identify volatility risk premia? *Journal of Financial and Quantitative Analysis*, 43(04):1055–1090.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–651.

Britten-Jones, M. and Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, 55(2):839–866.

Broadie, M., Chernov, M., and Johannes, M. (2009). Understanding index option returns. *Review of Financial Studies*, 22(11):4493–4529.

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3):281–318.

Carr, P. and Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, 22(3):1311–1341.

Christoffersen, P., Heston, S., and Jacobs, K. (2013). Capturing option anomalies with a variance-dependent pricing kernel. *Review of Financial Studies*, 26(8):1963–2006.

Coval, J. D. and Shumway, T. (2001). Expected option returns. *The Journal of Finance*, 56(3):983–1009.

Fengler, M. R. and Hin, L. (2015). Semi-nonparametric estimation of the call-option price surface under strike and time-to-expiry no-arbitrage constraints. *Journal of Econometrics*, 184(2):242–261.

Fiorin, L., Pages, G., and Sagna, A. (2017). Product Markovian quantization of an R^d -valued Euler scheme of a diffusion process with applications to finance. Working Paper. Available online at http://arxiv.org/abs/1511.01758.

Garcia, R. and Renault, E. (1998). A note on hedging in ARCH and stochastic volatility option pricing models. *Mathematical Finance*, 8(2):153–161.

Gârleanu, N., Pedersen, L. H., and Poteshman, A. M. (2009). Demand-based option pricing. *Review of Financial Studies*, 22(10):4259–4299.

Harrison, J. M. and Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215–260.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343.

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137.

Pagès, G. (1997). A space quantization method for numerical integration. *Journal of Computational and Applied Mathematics*, 89(1):1–38.

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.

Rubinstein, M. (1976). The valuation of uncertain income streams and the pricing of options. *The Bell Journal of Economics*, 7(2):407–425.

Schweizer, M. (1996). Approximation pricing and the variance-optimal martingale measure. *Annals of Probability*, 24(1):206–236.

Shephard, N. and Sheppard, K. (2010). Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 25(2):197–231.

Chapter 1

The Economic Value of Volatility Timing using Realized Volatility for Hedged S&P 500 Index Options

HUGO LAMARRE

DEBBIE J. DUPUIS

BRUNO RÉMILLARD

Abstract

This paper examines the economic value derived from holding negative inventories of S&P 500 index options from 2002 to 2014. Our focus is on profit-oriented and risk-averse agents attempting to improve their terminal risk-reward trade-off using volatility-driven protocols in the underlying market. More precisely, we perform variance-optimal risk minimization under various GARCH-type specifications. The resulting test is well-specified, robust to market timing abilities, and relevant to practitioners. We focus on the incremental value of high- over low-frequency-based volatility forecasts. In the former case, the proposed protocol is a novel non-myopic approach to hedging contingent claims

using realized variance. For out-of-the-money put options with one to three months-to-maturity, we find positive incremental value ranging from 5 to 95 basis points per annum on a risk-adjusted basis. The positiveness is robust to an out-of-sample exercise, moneyness-maturity effects, short-run option premia dynamics, and prospect-theory preferences.

1.1 Introduction

Volatility models are believed to have practical relevance for agents in equity markets. Beyond mean-variance portfolio allocation, few investment opportunities have actually been studied and empirically shown to benefit from volatility forecasts. We provide novel measurements for the S&P 500 index in an opportunity set characterized by static option positions and dynamic index protocols. In other words, we look at *dynamic hedging* as a means for profit-oriented and risk-averse agents —e.g. proprietary traders— to convert volatility knowledge into economic value.

Our work is closely related to Fleming et al. (2001, 2003) from whom we borrow the nomenclature *economic value of volatility timing*. We use it here in its broadest sense, i.e. as it refers to risk-averse agents implementing volatility-driven protocols towards improving the risk-reward trade-off of some already profitable investment opportunity. Whereas they consider minimum-variance portfolios à la Markowitz (1952), we consider variance-optimal hedging of a contingent claim à la Schweizer (1995). Under incomplete markets, it makes sense for agents to minimize terminal risk without fully reconciling market dynamics with option prices through arbitrage theory —effectively disentangling risk and reward concerns.

Our setting is economically motivated by recent empirical pricing advances of Gârleanu et al. (2009) who expose the role of buying demand pressures in the S&P 500 index option market. They propose a satisfying explanation for portfolios of interest having *positive* long-run expectations, with positiveness empirically robust to underlying dynamics. A profit-oriented agent hence has ongoing motivation for selling options on a daily basis, regardless of his or her adopted forecasting methodology for subsequently managing risk. Investors having access to statistically superior forecasts presumably implement protocols that reduce risk more efficiently, translating into higher economic value e.g. by allowing agents to reap more profits (in absolute) for a given budget of risk. Quantifying this economic value is the purpose of this paper.

We present empirical results under a comprehensive range of GARCH-type specifications, including *realized variance*-based models relying on high-frequency data. Such models are known to react more quickly to changes in prevailing volatility —a feature which is likely beneficial to protocols. We derive variance-optimal protocols relying on realized variance under the multiplicative-error-model (MEM) class of Engle (2002). To our knowledge, this is a novel methodological contribution to the discrete-time hedging literature.

To further illustrate our empirical endeavor, we preliminarily present a numerical experiment under a highly stylized economy, where agents are constrained to the model of Black and Scholes (1973) (B&S) and where market incompleteness arises from discrete-time portfolio revisions only. A *daily* revision setting is selected as a practical compromise between achievable risk diminution and transaction costs throughout. The equity market index m_t indexed by trading days $t = 0, \ldots, \tau$ follows a discretized geometric Brownian motion with an initial value of 1 i.e. $m_0 = 1$, a drift parameter of 5% per annum and a (constant) volatility parameter of 18% per annum.

For a given put contract with strike price k_0 and τ trading-days-to-maturity (TDM), our interest lies in the opportunity set,

$$\pi = \underbrace{c_0}_{\text{Initial option proceeds}} - \underbrace{(k_0 - m_\tau)^+}_{\text{Contingent liability}} + \underbrace{\sum_{t=1}^{\tau} \phi_t (m_t - m_{t-1})}_{\text{Volatility timing profits \& losses}} ,$$

spanned by volatility timing protocols ϕ_t, where $(x)^+ = \max(x, 0)$ and c_0 is the initial value of the contract. In line with the *sign* of the premium documented by e.g. Bakshi and Kapadia (2003), we assume c_0 corresponds to a 5% implicit volatility markup, i.e. a B&S price with a volatility parameter of 23% per annum. We further assume investors commit one unit of capital per option sold, such that π is readily interpreted as the overall rate of return of the strategy —presented in basis points per annum (bps/yr) throughout.

We consider two investors performing volatility timing via a classical B&S delta-hedging strategy: an *uniformed investor* who relies on the long-run volatility of 18% and a *perfect volatility foresight investor* who unrealistically relies on an *ex-post* volatility estimator. Table 1.1 presents simulation results when selling one-month-to-maturity out-of-the-money (OTM) put options.

Table 1.1: Risk-reward analysis of preliminary Monte Carlo experiment using 10,000,000 simulations for a put option with $\tau = 21$ and $k_0 = 0.95$. The market index is given by $m_t = \exp\left(\sum_{i=1}^{t} r_i\right)$ with $r_i = u - \delta/2 + \sqrt{\delta}\eta_i$, where η_i are uncorrelated standard normal random variates. We assume $u = 0.05/252$ and $\delta = 0.18^2/252$. Volatility timing protocols are given by $\phi_t(\sigma_{t,\tau}) = -\Phi\left(-(\log(m_{t-1}/k_0) + \sigma^2_{t,\tau}/2)/\sigma_{t,\tau}\right)$ where $\sigma_{t,\tau}$ is a total volatility forecast over $[t,\tau]$ and Φ is the standard normal cumulative density function. Numbers are converted to bps/yr according to $12000 \times \pi$. $\mathrm{E}[\pi]$ corresponds to a sample mean, $\mathrm{std}[\pi]$ to a standard deviation and Sharpe ratio to the first row divided by the second row.

Investor Type	Uninformed	Perfect Foresight
$\sigma^2_{t,\tau}$	$(\tau-(t-1))\delta$	$\sum_{i=t}^{\tau} r_i^2$
$\mathrm{E}[\pi]$ (bps/yr)	460.32	456.53
$\mathrm{std}[\pi]$ (bps/yr)	316.56	267.68
Sharpe Ratio	1.45	1.71

Under this toy economy, our working hypothesis takes the following form: agents with more(less) accurate volatility forecasts have Sharpe ratios closer to the upper(lower) bound of 1.71(1.45). This conjecture is consistent with absence of arbitrage due to market incompleteness. As long as market imbalance conditions described by Gârleanu et al. (2009) persist, agents are indeed being fairly compensated (by a 5% volatility markup) for bearing risk generated by discrete-time revisions. How this risk is managed in reality is a source of heterogeneity amongst profit-oriented agents, with some likely having a persisting economic advantage over others.

While the previous market model allows us to remain in the realm of B&S, it is obviously at odds with volatility timing. In a similar numerical experiment with market simulations performed under the GARCH specification of Heston and Nandi (2000)[1], we find Sharpe ratios ranging from 0.59 to 1.15. An agent with perfect volatility foresight now nearly doubles the Sharpe ratio of a uniformed agent. Given that such a GARCH specification is known to capture key features of equity markets including non-normality and a leverage effect, this preliminary numerical result should convince readers of the merit of the present endeavor and its practical relevance.

B&S delta-hedges are highly misspecified under GARCH dynamics as evidenced by the strong non-normality of previous π simulations (not shown). We henceforth depart from B&S towards risk-minimizing (i.e. variance-optimal) protocols. Roughly speaking, we assume agents minimize $\mathrm{E}[\pi^2]$ under their respective expectation operator capturing future volatility views. This approach uses volatility knowledge more efficiently than B&S by overcoming its well-known shortcomings related to continuous-time revisions, constant volatility and normality.

While empirical tests could be designed using other criteria (e.g. the maximization of a given utility function), our framework is inherently robust to uncertainty in expected S&P 500 returns through a carefully selected change of measure. In order words, we isolate the value of *volatility* timing by preemptively prohibiting agents from performing *market* timing. The resulting approximation which we refer to as *martingale hedging* may be viewed as an intermediary step between delta-hedging and variance-optimal hedging of Schweizer (1995). In contrast, Fleming et al. (2001, 2003) dedicate considerable efforts to demonstrating robustness to expected returns —a feature which is here achieved by design.

Regarding practical benefits, martingale hedging is highly tractable and reminiscent of a recursive regression. Typical variance-optimal protocols likely violate liquidity and risk constraints e.g. by suggesting S&P 500 exposures well above(below) 100%(-100%) or doubling-down behaviors during market crises. The proposed test instead reflects a realistic set of market opportunities, yielding relevant implications for option market par-

[1] Model parameters are estimated using S&P 500 index returns from 2000 to 2014. The initial conditional variance is set to its long-run expectation corresponding to a volatility of roughly 18% per annum. Option prices are still assumed to be given by B&S with a volatility of 23% per annum.

ticipants.

Our empirical focus is on the incremental value of using models based on high-frequency (i.e. realized variance) data over models based on low-frequency (i.e. daily) data. The terms low-frequency and high-frequency refer to the filtration used only; decisions are always taken on a daily (i.e. low-frequency) basis. Risk-minimizing protocols are *non-myopic* and rely on multi-period dynamics, e.g. as opposed to Fleming et al. (2001, 2003) who rely on one day ahead forecasts only. Model structure hence plays a critical role due to our proposed test effectively being a joint test of model specification and economic value. To maximize the scope of results, we first consider *naive* exponential smoothing specifications and incrementally introduce two structural requirements, namely *mean reversion* (i.e. the GARCH effect) and *leverage effect*.

Martingale hedging protocols are solved semi-parametrically using *filtered historical simulation* in the spirit of Barone-Adesi et al. (2008). We hence avoid normality assumptions (both in returns and model innovations) and also overcome an endogeneity challenge in the MEM framework, i.e. the need to specify a family of copulas for model innovations. We presumably improve efficiency by enforcing the martingale condition on a sample-to-sample basis, similarly to *empirical martingale simulations* of Duan and Simonato (1998).

To avoid unfair advantage to either low- or high-frequency-based forecasts, models are constrained to two state variables (namely asset price and conditional variance) and to the same long-run return and volatility expectations. We measure economic value using the performance metric of Ingersoll et al. (2007), which is reminiscent of an ex-post utility expectation. S&P 500 options are sold on a fixed lattice of moneyness and maturity and performance metrics are calibrated over the resulting lattice to empirically control for option contract specificities.

Our empirical results show incremental risk-adjusted gains from considering high- over low-frequency data are systematically positive, ranging roughly from 5 to 95 bps/yr in the case of put options. Models allowing for both GARCH and leverage effects deliver more stable economic value for longer-term options and higher value when shortfalls are more heavily penalized. In particular, loss averse agents should be as concerned by improving one day ahead volatility forecasts as correctly specifying future volatility expectations. All risk minimizing protocols outperform model-free delta-hedging of Bates (2005). The positiveness of the incremental value is further robust to an out-of-sample exercise and volatility timing in the *option* market —with all robustness checks presented in the supplementary material.

Our contribution differs from the existing related literature. Option papers using realized variance focus mainly on pricing, including Bandi et al. (2008) and Corsi et al. (2013). In particular, Christoffersen et al. (2014) study the economic value of using realized variance when *pricing* options.

This paper proceeds as follows. Section 1.2 motivates the proposed test of economic value by situating it in the option pricing and hedging literature. Section 1.3 presents martingale hedging protocols using high-frequency data. Section 1.4 formalizes the test, with empirical considerations further discussed in Section 1.5 and details sent to the supplementary material. Empirical results are presented and discussed in Section 1.6, with robustness checks also sent to the supplementary material. Section 1.7 concludes.

1.2 Volatility timing and hedging

1.2.1 Setting

We first make some mild option pricing assumptions which allow the aggregation of empirical results in time, while guaranteeing from the onset that volatility timing profits & losses (P&Ls) account for the net cost of financing underlying positions, without the need for subsequent cumbersome adjustments.

Let S_n be the official close value of the S&P 500 index on trading day $n \in \{0, \ldots, \tau\}$, where τ is the number of trading days to the expiration of an option —a put or call— with price C_0 and strike price K_0. The forward contract value at n maturing at τ is given by,

$$F_n^\tau = (S_n - D_n^\tau) B_n^\tau, \tag{1.1}$$

where B_n^τ is a bank account value at τ from one dollar deposited at n and D_n^τ is the present value of the sum of all dividends vested over $(n,\tau]$, with both interest rates and cash dividends assumed deterministic. We introduce normalized forward contract prices $m_n^\tau = F_n^\tau/F_0^\tau$ for $n = 0,\dots,\tau$ and normalized option prices,

$$c_0(k_0,\tau) = C_0 B_0^\tau / F_0^\tau, \tag{1.2}$$

with forward-moneyness (or simply *moneyness*) $k_n = K_0/F_n^\tau$ for $n = 0,\dots,\tau$ such that $k_n > 1$ for an OTM call option and $k_n < 1$ for an in-the-money (ITM) call option, and vice-versa for put options. In the supplementary material, we interpret this setting as a change of numéraire and c_0 as the price of an option written on m —under no interest nor dividends— with strike price k_0 and τ TDM.

We further introduce a self-financing portfolio value $\{v_n\}_{n=0}^\tau$. Letting $\{\phi_n\}_{n=1}^\tau$ be a volatility timing protocol, we have for $n = 0,\dots,\tau$,

$$v_n = c_0 + \sum_{i=1}^{n} \phi_i (m_i - m_{i-1}), \tag{1.3}$$

where the sum is zero when $n = 0$ by convention and ϕ_n is the fraction of forward contracts held during $[n-1,n)$.

The *hedged option return* over $[0,\tau]$ is defined as,

$$\pi = v_\tau - c_\tau,$$

where c_τ is the option payoff, namely $c_\tau = (m_\tau - k_0)^+$ for a call option and $c_\tau = (k_0 - m_\tau)^+$ for a put option. Under our numéraire interpretation, π is an *excess* rate of return when the strategy is fully funded with one unit of numéraire, namely with one zero-coupon bond of notional value F_0^τ. Our empirical focus is on the economic value derived from strategy returns π.

1.2.2 Disentangling pricing and hedging

Finding and motivating the presence of risk premia embedded in option prices has been an ongoing research topic since the seminal work of Coval and Shumway (2001). By testing implications of the CAPM, they found other risk factors beyond leveraged market exposure are heavily priced in equity index options. Buraschi and Jackwerth (2001) argue options are non-redundant securities in that they expand one's investment opportunity set with additional un-hedgeable sources of priced risk such as stochastic volatility, jumps, hedge portfolio misspecification (i.e. model risk) and/or discrete-time hedge portfolio revisions. Empirical evidence for the S&P 500 overwhelmingly suggests un-hedgeable risk factors are positively rewarded in equilibrium such that selling and hedging options is not only risky, but profitable in the long-run.

As a possible explanation, Gârleanu et al. (2009) consider risk-averse market-makers who are compensated for absorbing buying pressures from end-users. Their theoretical results suggest market-makers increase option prices proportionally to the variance of un-hedgeable risk factors, which translates to $E[\pi] \propto \text{Var}[\pi]$ in our setting. We refer to $E[\pi]$ as the *hedged option premium* to emphasize the transfer of wealth from un-hedged option buying end-users to hedged participants. This term is also coined as a comprehensive alternative to model-specific premia such as the volatility risk premium, the presence of which is still debated in the literature most notably with respect to the competing jump risk premium. We work under the assumption that the hedged option premium is positive for reasonable choices of volatility timing protocols.

Interestingly, results from Gârleanu et al. (2009) improve when proprietary traders are included with market-makers as hedged intermediaries. Our focus is on such profit-oriented agents who have no liquidity-providing requisite and remain sufficiently small relative to end-users not to disturb the market equilibrium. In practice, profit-oriented agents likely have heterogeneous market views, hedging objectives and hedge revision frequencies leading to ex-ante disagreements about $\text{Var}[\pi]$ and incidentally $\text{E}[\pi]$. While working out the details of the resulting equilibrium falls outside the scope of this paper, it is plausible that some agents with access to more sophisticated risk management tools have a persisting economic advantage over others *without violating absence of arbitrage* due to market incompleteness. For example, heterogeneity could arise from transaction costs, leading to more or less frequent revisions or technology constraints such as limited access to high-frequency data.

While this conjecture provides a sound basis for a test of economic value, we make an additional assumption which ensures its viability by further disentangling pricing and hedging concerns: our agent pre-commits to holding option positions until maturity. This guarantees the price of un-hedgeable risk factors is never paid out by buying options back. In the same spirit, we prohibit buying options of other strike prices and/or maturities for hedging purposes e.g. towards performing delta-gamma hedging under B&S.

As a direct consequence, our agent is indifferent to intermediary option prices, steering concerns from local to global and towards the ability to cover a terminal liability. Under global concerns, delta-hedges calibrated to option prices are likely suboptimal when prices are significantly impacted by non-arbitrage considerations such as market segmentation. The quality of delta-hedges is further shown by Alexander and Nogueira (2007) to be only impacted by the fit of underlying dynamics to option prices. Delta-hedging is hence driven by aggregated volatility expectations as conveyed by option prices, not investors' individual volatility forecasting abilities. We next propose risk-minimizing protocols which are better suited to measuring the economic value of volatility timing under the current opportunity set.

1.2.3 Martingale hedging

With replication arguments failing under incomplete markets, an investor is faced with the somewhat arbitrary task of selecting a criterion to optimize, e.g. variance, value-at-risk, utility, etc. This choice is critical as it characterizes the impact of expected (underlying) returns on protocols. Volatility timing strategies should also be sufficiently tractable and well-behaved to mirror how profit-oriented agents would presumably go about extracting the hedged option premium in practice.

We focus on a quadratic criterion for risk as resulting hedges have been shown to converge to delta-hedges under complete markets and offer a natural extension of replication arguments to incomplete markets. Two quadratic risk-minimizing approaches have been proposed: local risk minimization and variance-optimal hedging; see e.g. Föllmer and Schied (2011) and references therein for a review in discrete-time.

The latter approach minimizes the stochastic distance between the liability due at expiration (c_τ) and the terminal value of a portfolio (v_τ) under the strictly self-financing formulation of Eq. (1.3). It does so by optimally controlling for the *accumulation* of hedging errors over time by solving for $\arg\min_{\{\phi\}} \mathrm{E}\left[\pi^2\right]$. This last expectation is usually taken under a probability measure representing an agent's best guess about market dynamics —including expectations about market returns and conditional variances.

The resulting mean-variance demand leads to time inconsistency and practical concerns e.g. by inciting investors to take increasingly more risk as large P&Ls accumulate. Resulting protocols could fall well-outside typical B&S delta-hedge bounds of $\pm[0,1]$, possibly provoking internal or external risk-overseeing entities to impose arbitrary constraints on trading or increase margin requirements. Brandt (2003) expresses similar concerns and relaxes the self-financing property to achieve time consistency. We instead preemptively fix the role of expected returns prior to optimizing protocols.

We assume profit-oriented agents cast their market views in terms of a *subjective martingale measure* $\mathbb{Q}$ constrained by the condition,

$$\mathrm{E}^{\mathbb{Q}}_{n-1}\left[m_n - m_{n-1}\right] = 0, \quad \text{for } n = 1, \ldots, \tau, \tag{1.4}$$

where $\mathrm{E}^{\mathbb{Q}}_{n}[.]$ denotes the expectation under measure $\mathbb{Q}$ conditional on information up to time n. Basak and Chabakauri (2012) find Eq. (1.4) to be a necessary condition for the time consistency of the variance-optimal hedging criterion, alleviating aforementioned practical concerns and making local risk minimization optimal in the variance-optimal sense. In other words, if an investor has no educated guess about market returns, a unique risk-minimizing approach arises that consolidates the two main strands of the quadratic hedging literature under incomplete markets. More importantly, the martingale condition guarantees investors are *only* set apart by their volatility forecasting skills, not by their market timing abilities.

Following Schweizer (1995),

$$\arg\min_{\{\phi\}} \mathrm{E}^{\mathbb{Q}}\left[\pi^2\right],$$

is solved recursively for $n = \tau, \ldots, 1$ by,

$$\phi_n = \frac{\mathrm{E}^{\mathbb{Q}}_{n-1}\left[(m_n - m_{n-1})p_n\right]}{\mathrm{E}^{\mathbb{Q}}_{n-1}\left[(m_n - m_{n-1})^2\right]}, \tag{1.5}$$

where $p_{n-1} = \mathrm{E}^{\mathbb{Q}}_{n-1}[p_n]$ and $p_\tau = c_\tau$. We refer to the resulting protocol as the *martingale hedging* strategy. Since our focus is on Markovian processes, protocols may be solved as daily functions of m and possibly other latent variables (such as conditional variance) by means of dynamic programming.

It might appear as if we trade one arbitrary criterion for another, namely a risk-minimizing criterion for a martingale measure. Elliott and Madan (1998) propose a specific change of measure for deriving $\mathbb{Q}$ under which investors nullify their expected return prior to minimizing risk[2]. This is formalized by the extended Girsanov principle (EGP) which fits naturally in our setting as it yields the change of measure minimally disrupting the information content about volatility.

Even though the resulting setting is reminiscent of option pricing, we emphasize that the motivation for introducing $\mathbb{Q}$ is not related to the classical risk-neutral valuation principle of Harrison and Pliska (1981). In particular, p_0 is *never* interpreted as an option price, since it fails to account for the price of un-hedgeable risk factors —the presence of which motivates the current paper. The apparent discrepancy in preferences used for hedging versus pricing is here motivated by market segmentation. We now turn to the main methodological contribution of this paper, namely realized variance-based protocols.

[2] More precisely, investors “minimize the variance of a risk adjusted discounted cost of hedging that uses risk adjusted asset prices in calculating hedging returns”.

1.3 Hedging using realized variance

1.3.1 Realized variance-based volatility forecasting

We introduce a realized variance process $\{v_n\}_{n=0}^{\tau}$, where v_n is an estimator of the quadratic variation of a continuous-time Itô representation of markets over trading day n. Lagged realized variance v_{n-1} is used to predict the conditional variance of *excess* log-returns $r_n = \log(m_n/m_{n-1})$ in the spirit of Engle and Gallo (2006) or Shephard and Sheppard (2010). We work under a probability space $(\Omega, \mathbb{F}, \mathbb{Q})$, where $\mathbb{F} = \{\mathscr{F}_n, n = 0, \ldots, \tau\}$ is the filtration generated by the sigma-algebra $\sigma(\{m_n, v_n\})$.

We consider a return generating function $r_n = \mathscr{R}(h_n, \eta_n)$ where $h_n = \mathrm{Var}_{n-1}[r_n]$ is the conditional variance of log-returns. Model innovations η_n are mean zero and variance one independent and identically distributed (iid) random variables under $\mathbb{Q}$ with $\eta_n \perp\!\!\!\perp \mathscr{F}_{n-1}$, where $\perp\!\!\!\perp$ denotes independence. To ensure the martingale condition is verified for $m_n = m_{n-1}e^{r_n}$, we let,

$$\mathscr{R}(h, \eta) = \eta h^{1/2} - \kappa(h^{1/2}), \tag{1.6}$$

where $\kappa(z) = \log(\mathrm{E}[e^{\eta_1 z}])$ is the cumulant-generating function of η_1.

We focus on a single realized variance indicator and one lag auto-regressive specifications[3],

$$\begin{cases} r_n &= \mathscr{R}(h_n, \eta_n), \\ h_{n+1} &= \mathscr{H}(h_n, \mu_n, r_n, v_n), \end{cases} \tag{1.7}$$

$$\begin{cases} v_n &= \mu_n \varepsilon_n, \\ \mu_{n+1} &= \mathscr{U}(h_n, \mu_n, r_n, v_n), \end{cases} \tag{1.8}$$

where $\mu_n = E_{n-1}[v_n]$ is the conditional expectation of realized variance and ε_n are positive innovations with mean one under $\mathbb{Q}$. Joint innovations are further assumed iid with $(\eta_n, \varepsilon_n) \perp\!\!\!\perp \mathscr{F}_{n-1}$.

Similarly to GARCH models, $\mathscr{H}$ and $\mathscr{U}$ represent auto-regressive functions of lagged $\{r, v\}$ such that $\{h, \mu\}$ is a predictable process, i.e. (h_n, μ_n) is $\mathscr{F}_{n-1}$-measurable. For

[3] This choice is motivated by computational constraints pertaining to dynamic programming.

example, the HEAVY model of Shephard and Sheppard (2010) is given by,

$$\begin{aligned} \mathscr{H}(h,\mu,r,\nu) &= \omega_r + \alpha_r \nu + \beta_r h, \\ \mathscr{U}(h,\mu,r,\nu) &= \omega_\nu + \alpha_\nu \nu + \beta_\nu \mu, \end{aligned} \tag{1.9}$$

with parameters $\{\omega_r, \alpha_r, \beta_r, \omega_\nu, \alpha_\nu, \beta_\nu\}$. For the S&P 500 index, typical estimates for α_r are greater than their GARCH counterparts, such that the HEAVY model puts more weight on recent observations and reacts more quickly to changes in prevailing volatility. Our purpose is to translate such a desirable feature into a protocol.

In this context, specifying realized variance dynamics by Eq. (1.8) is necessary, even though option payoffs do not explicitly depend on realized variance. Future daily return variances indeed depend on future realized variances for typical choices of $\mathscr{H}$. The dependence structure of (η,ε) must hence be specified, which poses the additional challenge of selecting a family of copulas. For empirical tests, we overcome this issue by sampling from historical estimates.

1.3.2 Variance-optimal protocols

We may now derive protocols sourcing volatility information from high-frequency data under the martingale condition given by Eq. (1.4). Since $\{m_{n-1}, h_n, \mu_n\}$ is a three-dimensional Markovian process under Eq. (1.7)-(1.8), martingale hedging protocols ϕ_n are computed as functions of (m_{n-1}, h_n, μ_n). We solve for p_n on a three-dimensional grid by backward induction according to,

$$p_{n-1}(m,h,\mu) = \int \check{p}_n\left(me^{\mathscr{R}}, \mathscr{H}(h,\mu,\mathscr{R},\mu\varepsilon), \mathscr{U}(h,\mu,\mathscr{R},\mu\varepsilon)\right) dF,$$

for $n = \tau-1,\dots,0$ starting from $p_\tau = c_\tau$, where F denotes the joint distribution of (η_1,ε_1) under $\mathbb{Q}$ and where the dependence of $\mathscr{R}$ on (h,η) is omitted for clarity. The $\check{x}(\cdot)$ operator henceforth represents an interpolation function. Here, $\check{p}_n(m,h,\mu)$ is a three-dimensional interpolation function applied to previously computed values of p_n on a given grid. Integrals are hence estimated using a combination of standard Monte-Carlo and interpolation techniques. A specific Monte-Carlo estimator is later proposed for empirical tests.

Similarly, protocols are recursively solved for $n = \tau, \ldots, 1$ according to,

$$\phi_n(m,h,\mu) = \frac{\int (e^{\mathscr{R}} - 1)\breve{p}_n\big(me^{\mathscr{R}}, \mathscr{H}(h,\mu,\mathscr{R},\mu\varepsilon), \mathscr{U}(h,\mu,\mathscr{R},\mu\varepsilon)\big)dF}{m\int (e^{\mathscr{R}} - 1)^2 dF},$$

where ϕ_n is predictable since m_{n-1}, h_n and μ_n are $\mathscr{F}_{n-1}$-measurable.

1.4 Economic value of volatility timing

1.4.1 Measuring economic value

We assume agents sell options on a daily basis. Each trading day is associated to an excess return π_t representing the overall outcome from selling a *single* option at t and hedging it for τ additional trading days until maturity is reached. Our empirical focus is on the resulting time series $\{\pi_t\}_{t=1}^{T}$, where T represents the total number of trading days in our sample minus τ to allow for the last option to be hedged until maturity. This time series is auto-correlated due to heavily overlapping hedging periods. For example, an investor selling monthly options on a daily basis faces 20 trading days of overlap between an option sold today and an option sold yesterday —assuming 21 trading days per month. We next consider unconditional economic metrics (i.e. specified as a sample mean) which statistically benefit from overlaps due to the path dependence of P&Ls. Sample size thus takes precedence over auto-correlation concerns.

Numerous performance measures have been proposed in the literature typically motivated by a utility function, a set of axioms, or both. In this last category, Ingersoll et al. (2007) propose a measure which is robust to uninformed portfolio manipulations, is reminiscent of a utility expectation under relative risk aversion ρ, and can be adapted to our setting. We use,

$$\Theta_{\mathrm{U}}(\pi;\rho) = \frac{252}{(1-\rho)\tau}\log\left(\frac{1}{T}\sum_{t=1}^{T}(1+w_t\pi_t)^{1-\rho}\right),$$

where w_t is related to the number of options sold at time t and is shared by all volatility timing protocols. We consider *uninformed* exposures given by $w_t = 1$ for simplicity.

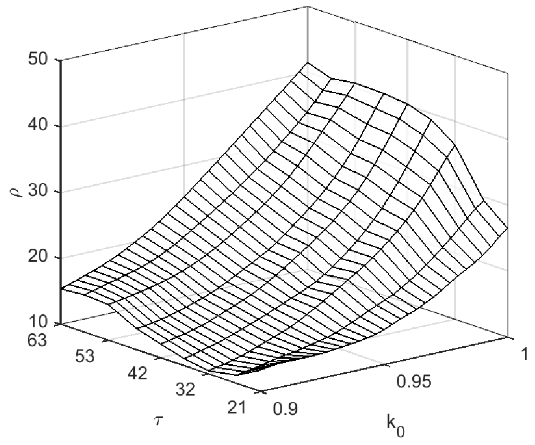
We calibrate ρ by level of moneyness and time-to-expiry,

$$\rho_{k_0,\tau} = \frac{\log(\mathrm{E}[1 + w_t \pi^0_{t;k_0,\tau}])}{\mathrm{Var}[\log(1 + w_t \pi^0_{t;k_0,\tau})]},$$

where $\pi^0_{t;k_0,\tau}$ represents benchmark returns achieved under no volatility-timing for a single option contract characterized by moneyness-maturity (k_0, τ). The variance is carefully estimated from non-overlapping subsamples and averaged to avoid biases.

Figure 1.1 displays calibrated risk aversion parameters. The calibration procedure ensures agents relying on long-run volatility views are appropriately incentivized for *all* option contracts. For example, only agents having the lowest risk aversion contemplate selling deeply OTM options, due to the corresponding ex-post risk-reward trade-off being particularly unfavorable. The resulting normalization greatly improves cross-sectional comparability along maturity and moneyness.

Figure 1.1: Cross-sectional calibration of risk aversion parameters $\rho_{k_0,\tau}$ for utility-based performance measures Θ_U.



Our risk aversion calibration, however, fails to correct for *short-run* dynamics in option premia. Under the model proposed by Bakshi and Kapadia (2003) for example, their hedged option premium increases with the sensitivity of an option price to volatility, commonly known as the *vega*. By assuming $w_t = 1$, one could hence argue that agents indirectly perform volatility timing in the option market, as the vega of a single option varies in time with prevailing risk levels. In this context, uniformed w_t exposures are arguably better defined by a fixed amount of vega, as opposed to a fixed number of option contracts. While different choices for w_t may indeed lead to different conclusions regarding incremental economic value, establishing robustness in a model-free manner is challenging. In the supplementary material, we consider vega-constant exposures under B&S using implicit volatilities.

As a fundamental feature of our test, the best hedging strategy (from a risk perspective) may not yield the best economic value if considerable profits must be forgone. For example, deeply OTM short-term options very rarely benefit from hedging. An *un-hedged* protocol (i.e. $\phi_n = 0$) may hence very well yield the best *ex-post* economic value e.g. due to extreme market events never occurring over the empirical sample —an issue commonly known as the *peso problem*. In the supplementary material, we consider a prospect theory-based performance measure which penalizes hedging shortfalls more aggressively in an attempt to realign ex-ante hedging objectives with ex-post performance assessment.

1.4.2 Historical volatility dynamics

We now introduce three classes of volatility forecasts —namely static (S0), low-frequency (LF) and high-frequency (HF)— to be tested. We follow the notation introduced in Section 1.3, but work under a probability space $(\Omega, \mathbb{F}, \mathbb{P})$ where $\mathbb{P}$ now reflects the likelihood of market events as they occurred over our entire sample, namely the *physical* measure. Distributions of η and ε under $\mathbb{P}$ are left unspecified, but still represent respectively real-valued and positive innovations.

The S0 specification serves as a benchmark and corresponds to the case when volatility timing is unavailable. The S0 model is characterized by independent and identically distributed log-returns with constant unconditional mean and variance, respectively given by $u = \mathrm{E}[r_n]$ and $\delta = \mathrm{Var}[r_n]$. More precisely, the S0 model is specified as $r_n = u + \eta_n \delta^{1/2}$ which may be viewed as the discretization of a Brownian motion with drift when η_n is Gaussian.

The incremental value from expanding one's information set should be distinguishable using naive model structures. While this goal is similar in spirit to Fleming et al. (2001, 2003), they do not provide a valid market model. Their exponential smoothing specification indeed allows volatility expectations to go to zero with time. Given the non-myopic nature of our setting, optimal protocols rely on volatility expectations until the expiration of an option. A first model requirement is hence the mean reversion of volatility expectations towards a strictly positive value, namely the GARCH effect. A second model requirement is to have agents adjust volatility forecasts more aggressively following negative market returns. This choice is motivated by the documented first order impact of leverage effects on hedging protocols; see e.g. Nandi (1998).

Finding specifications that are representative of their class (i.e. LF and HF) under such stringent structural requirements is challenging. The GJR model of Glosten et al. (1993) offers a reasonable anchor point for model selection as it introduces leverage effects using the most rudimentary mechanism, namely an indicator function —denoted here by $\mathrm{I}(.)$. To decompose the relative contribution of each requirement, we first consider the naive estimator of Fleming et al. (2001, 2003) and gradually introduce mean reversion and a leverage effect. Table 1.2 lists resulting specifications under all three classes.

The HF class follows the previously introduced MEM-type system Eq. (1.7)-(1.8). We let h_n be an affine and contemporaneous transformation of μ_n, i.e. $h_{n+1} = a + b\mu_{n+1}$, inspired by Brownlees and Gallo (2010). This choice is motivated by the fact (m_n, h_{n+1}) is now a Markovian process, hence limiting the structural superiority of HF versus LF classes and isolating the impact of expanding one's information set from modeling assumptions. As documented by Shephard and Sheppard (2010), the addition of a state variable (μ) indeed allows momentum in volatility, in contrast to strictly monotonic term structures under classical GARCH models.

We use realized variances computed from intra-daily returns during normal market hours, i.e. from *open-to-close*. Since close-to-close log-returns are also impacted by overnight gaps, b plays the critical role of increasing realized variances from an open-to-close to a close-to-close scale. In contrast, estimates for a are usually insignificant.

To fully exploit the potential of HF data, the HFA model substitutes realized variance

Table 1.2: Overview of model specifications under $\mathbb{P}$.

No Volatility Timing	
Static (S0)	$\left\{ r_n = u + \eta_n \delta^{1/2}. \right.$
Low-Frequency	
Naive (LF0)	$\begin{cases} r_n & = u + \eta_n h_n^{1/2}, \\ h_{n+1} & = \alpha e^{-\alpha}(r_n - u)^2 + e^{-\alpha} h_n. \end{cases}$
Mean Reversion (LFG)	$\begin{cases} r_n & = u + \eta_n h_n^{1/2}, \\ h_{n+1} & = \omega + \alpha (r_n - u)^2 + \beta h_n. \end{cases}$
Leverage Effect (LFA)	$\begin{cases} r_n & = u + \eta_n h_n^{1/2}, \\ h_{n+1} & = \omega + (\alpha + \mathrm{I}(r_n < u)\,\gamma)\,(r_n - u)^2 + \beta h_n. \end{cases}$
High-Frequency	
Naive (HF0)	$\begin{cases} r_n & = u + \eta_n h_n^{1/2}, \\ h_{n+1} & = a + b\mu_{n+1}, \\ \nu_n & = \varepsilon_n \mu_n, \\ \mu_{n+1} & = \alpha e^{-\alpha} \nu_n + e^{-\alpha} \mu_n. \end{cases}$
Mean Reversion (HFG)	$\begin{cases} r_n & = u + \eta_n h_n^{1/2}, \\ h_{n+1} & = a + b\mu_{n+1}, \\ \nu_n & = \varepsilon_n \mu_n, \\ \mu_{n+1} & = \omega + \alpha \nu_n + \beta \mu_n. \end{cases}$
Leverage Effect (HFA)	$\begin{cases} r_n & = u + \eta_n h_n^{1/2}, \\ h_{n+1} & = a + b\mu_{n+1}, \\ \nu_n^\star & = \varepsilon_n \mu_n, \\ \mu_{n+1} & = \omega + (\alpha + \mathrm{I}(r_n < u)\,\gamma)\,\nu_n^\star + \beta \mu_n. \end{cases}$

for realized semi-variance $\nu_n^\star$, e.g. given by the sum of squared five-minute returns over *negative* returns. While parameter γ controls for the asymmetry generated by negative *daily* returns, the HFA model further accounts for the asymmetry generated by negative *intradaily* returns. This choice introduces a novel leverage effect related to ideas sketched by Shephard and Sheppard (2010). While we limit our specification to realized semi-variance to maintain a similar structure to the LF class, economic value derived from the HFA model would likely improve when including both realized semi-variance *and* realized variance to the information set.

1.4.3 Casting subjective volatility views

This section applies the EGP to models previously calibrated to historical market returns. The EGP —roughly speaking— preserves the martingale part of market dynamics from $\mathbb{P}$ to $\mathbb{Q}$: volatility dynamics are untouched, similarly to the local risk-neutral valuation relationship of Duan (1995).

For the S0 model, the EGP modifies return dynamics in a way that respects the martingale condition Eq. (1.4) and preserves the distribution of innovations. The application of the EGP is more challenging under the LF class as volatility dynamics depend on return observations.

Subjective martingale views rely on the previously introduced martingale return-generating function given by Eq. (1.6). The martingale adaptation of the S0 model is given by $r_n = \mathscr{R}(\delta, \eta_n)$ which may be viewed as the discretization of the B&S model when η_n is Gaussian, i.e. $r_n = \eta_n\sqrt{\delta} - \delta/2$.

All applications of the EGP are presented in Table 1.3, where we rely on starred innovations defined as,

$$\eta_n^\star = \eta_n - \frac{u + \kappa({h_n}^{1/2})}{{h_n}^{1/2}}.$$

Regarding HF models, only the HFA specification actually depends on daily log-returns and therefore must be adjusted under the EGP. In particular, the martingale adaptations of both HF0 and HFG remain unchanged beyond the return-generating function.

Table 1.3: Overview of model specifications under $\mathbb{Q}$.

No Volatility Timing	
Static (S0)	$\left\{ r_n = \eta_n \delta^{1/2} - \kappa(\delta^{1/2}). \right.$
Low-Frequency	
Naive (LF0)	$\begin{cases} r_n &= \eta_n h_n^{1/2} - \kappa({h_n}^{1/2}), \\ h_{n+1} &= \alpha e^{-\alpha} \eta_n^{\star 2} h_n + e^{-\alpha} h_n. \end{cases}$
Mean Reversion (LFG)	$\begin{cases} r_n &= \eta_n h_n^{1/2} - \kappa({h_n}^{1/2}), \\ h_{n+1} &= \omega + \alpha \eta_n^{\star 2} h_n + \beta h_n. \end{cases}$
Leverage Effect (LFA)	$\begin{cases} r_n &= \eta_n h_n^{1/2} - \kappa({h_n}^{1/2}), \\ h_{n+1} &= \omega + (\alpha + \mathrm{I}(\eta_n^\star < 0)\,\gamma)\, \eta_n^{\star 2} h_n + \beta h_n. \end{cases}$
High-Frequency	
Naive (HF0)	$\begin{cases} r_n &= \eta_n h_n^{1/2} - \kappa({h_n}^{1/2}), \\ h_{n+1} &= a + b \mu_{n+1}, \\ v_n &= \varepsilon_n \mu_n, \\ \mu_{n+1} &= \alpha e^{-\alpha} v_n + e^{-\alpha} \mu_n. \end{cases}$
Mean Reversion (HFG)	$\begin{cases} r_n &= \eta_n h_n^{1/2} - \kappa({h_n}^{1/2}), \\ h_{n+1} &= a + b \mu_{n+1}, \\ v_n &= \varepsilon_n \mu_n, \\ \mu_{n+1} &= \omega + \alpha v_n + \beta \mu_n. \end{cases}$
Leverage Effect (HFA)	$\begin{cases} r_n &= \eta_n h_n^{1/2} - \kappa({h_n}^{1/2}), \\ h_{n+1} &= a + b \mu_{n+1}, \\ v_n^\star &= \varepsilon_n \mu_n, \\ \mu_{n+1} &= \omega + (\alpha + \mathrm{I}(\eta_n^\star < 0)\,\gamma)\, v_n^\star + \beta \mu_n. \end{cases}$

1.5 Empirical methodology

1.5.1 Dataset

The dataset is comprised of daily best bid and ask prices for put and call options written on the spot S&P 500 index and traded on the Chicago Board Options Exchange (CBOE) from 02-Jan-2002 to 31-Dec-2014. The selected period covers a full economic cycle, including the bull market of the mid-2000s, the financial crash of 2008-2009 and its recovery. All option contracts are European and cash-settled. Option quotes, underlying close prices and US zero-coupon yield curves are sampled from the OptionMetrics Ivy

Database. Historical cash dividends for the S&P 500 index and overnight USD LIBOR rates are extracted from Bloomberg[4]. Realized measures are extracted from Oxford-Man Institute's realized library (Heber et al., 2009). We select 5-minute estimators using 1-minute subsamples for both realized variance and realized semi-variance.

The distribution of hedged option returns most likely depends on both moneyness and maturity. Our empirical test is hence complicated by varying option contract availability over the moneyness-maturity plane. We interpolate available contracts in an arbitrage-free manner over both moneyness *and* maturity. Details of the procedure can be found in the supplementary material, together with data cleaning filters. For each trading day, the procedure yields an arbitrage-free pricing surface calibrated to mid-quotes, which may reasonably be interpreted as market clearing prices over moneyness-maturity intervals of interest.

1.5.2 Resolving forward price uncertainty

Forward contract prices matching the maturity of options are not observed on financial markets and must therefore be estimated. We resolve dividend and rate uncertainty in Eq. (1.1) using market realizations,

$$B_n^\tau = \prod_{k=n}^{\tau-1} 1 + \frac{c_k \times \mathrm{LIBOR}_k}{100 \times 360},$$

$$D_n^\tau = \sum_{k=n}^{\tau-1} \frac{\mathrm{Div}_k}{B_n^k},$$

where LIBOR_k is the overnight USD LIBOR fixed at k, c_k is the number of calendar days from k until the next business day and Div_k is the weighted sum of all index components' cash dividends with ex-dividend date k. By convention, $B_n^n = 1$ and $D_n^n = 0$. This setting ensures empirical results reflect the actual net cost of financing S&P 500 index positions on a daily basis.

[4] Dividend data are retrieved from Bloomberg by requesting the field 'LAST_DPS_GROSS' for the 'SPX Index' ticker.

The only remaining source of uncertainty in Eq. (1.1) is the spot value S_0. Option cross-sections and index spot prices have historically been sampled at their respective market close time, i.e. 15 minutes apart. In the supplementary material, we synchronize S&P 500 spot prices with option cross-sections by performing daily regressions on various interest rate term structures. Typical corrections range from 5 to 20 bps, but can reach 100 bps on particularly volatile days. Corrections are especially significant prior to 29-Sept-2008, the day on which OptionMetrics implemented significant improvements to its sampling methodology.

1.5.3 Model estimation

Models are calibrated on a 2000-2014 time series of appended *quarterly* forward contracts computed according to Eq. (1.1) using rates and dividends actually vested over the period and a synchronized spot underlying value. A two-year burn-in period prior to 2002 is added to allow for the impact of initial state estimates, i.e. $\hat{h}_1$ and $\hat{\mu}_1$, to decay before selling the first option.

We impose the same long-run log-return expectation across all models for comparability, estimated at $\hat{u} = 7.1434\text{e}{-5}$. With the exception of naive specifications which have zero long-run volatility expectations, $\text{Var}[r_n]$ is similarly imposed to $\hat{\delta} = 1.6415\text{e}{-4}$ for all models. The validity of this approach, known as *variance targeting*, has been studied in both the LF and HF case. Regarding the LF class, we target the long-run variance in the LFG model by setting ω to $\hat{\delta}(1-\alpha-\beta)$ and ω to $\hat{\delta}(1-\alpha-0.5\gamma-\beta)$ in the LFA model.

For the HF class, variance targeting is achieved by constraining parameters a and b. Let us first introduce $v = \text{E}[\mathcal{V}]$ and $v^\star = \text{E}[\mathcal{V}^\star]$, estimated as respectively $\hat{v} = 1.0653\text{e}{-4}$ and $\widehat{v^\star} = 5.3216\text{e}{-5}$. Letting $(\hat{a} = 0, \hat{b} = \hat{\delta}/\hat{v})$ for the HFG model and $(\hat{a} = 0, \hat{b} = \hat{\delta}/\widehat{v^\star})$ for the HFA model, the long-run variance of daily returns matches $\hat{\delta}$. As anticipated, $\hat{b} = 3.0847$ for the HFA case is significantly larger than $\hat{b} = 1.5409$ for the HFG model. Realized semi-variances are indeed bounded above by realized variances and must hence be scaled more aggressively. Table 1.4 shows quasi-maximum likelihood (QML) estimates for remaining parameters.

Table 1.4: QML parameter estimates. Standard errors, shown between parentheses, are computed according to Bollerslev and Wooldridge (1992).

	Static	**Low-Frequency**			**High-Frequency**		
	S0	LF0	LFG	LFA	HF0	HFG	HFA
$\hat{\alpha}$	-	0.06(0.01)	0.09(0.01)	0.00(0.01)	0.23(0.01)	0.49(0.02)	0.34(0.02)
$\hat{\beta}$	-	-	0.90(0.01)	0.89(0.01)	-	0.49(0.02)	0.63(0.02)
$\hat{\gamma}$	-	-	-	0.19(0.02)	-	-	0.02(0.03)
LogL($\hat{\eta}$)	14439	15268	15317	15400	15446	15472	15501
E[$\hat{\eta}$]	-0.000	0.003	0.002	-0.003	0.004	0.001	0.000
Var[$\hat{\eta}$]	1.00	1.12	0.99	0.97	1.13	0.98	0.95
$\rho(\hat{\eta}_t^2, \hat{\eta}_{t-1}^2)$	0.19	-0.02	-0.05	-0.07	-0.04	-0.07	-0.07
LogL($\hat{\varepsilon}$)	-	-	-	-	16187	16237	17518
E[$\hat{\varepsilon}$]	-	-	-	-	1.16	1.01	1.00
Var[$\hat{\varepsilon}$]	-	-	-	-	0.75	0.45	0.65
$\rho(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1})$	-	-	-	-	0.16	-0.01	0.01

The resulting constrained estimation procedure is rather crude. For example, the variance targeting specification for the HFA model implicitly presupposes $(r_n - \hat{u} < 0) \perp\!\!\!\perp v_n^\star$ and $\mathrm{E}_n[r_n - \hat{u} < 0] = 0.5$. Both assumptions are likely violated in practice: the correlation between $(r_n - \hat{u} < 0)$ and $v_n^\star$ is 0.15 and the sample mean of $(r_n - \hat{u} < 0)$ is 0.46. This approach is nonetheless consistent with our focus on expanding one's information set, as opposed to fine tuning modeling assumptions. In particular, empirical results likely understate economic values that could be derived from more sophisticated specifications and/or estimation procedures.

The statistical value of using HF data and allowing for structural requirements (namely, mean reversion and leverage effect) is evidenced by log-likelihood of return innovations increasing steadily from left to right in Table 1.4. As already noted by others, the statistical superiority of the HF class stems from the lower signal-to-noise ratio of realized variances versus squared daily returns. This in turns leads to higher $\hat{\alpha}$ and lower $\hat{\beta}$ and allows high-frequency models to assimilate lagged information more quickly than their low-frequency counterpart. For example, α goes from 0.09 for LFG to 0.49 for HFG, with yesterday's realized-variance having roughly five times the impact of yesterday's squared daily return on one day ahead volatility forecasts. In the HFA case, γ is insignificant, suggesting the leverage effect driven by intradaily returns overwhelms the one driven by daily returns.

1.5.4 Protocol estimation

We easily show that risk-minimizing protocols Eq. (1.5) are homogeneous functions of degree zero with respect to (m_{n-1}, k_0), i.e. $\phi_n(\alpha m_{n-1}, \alpha k_0) = \phi_n(m_{n-1}, k_0)$ for $\alpha \in \mathbb{R}$. Normalizing the underlying by the initial moneyness and introducing $\xi_n = m_n/k_0$, protocols may be written as,

$$\phi_n = \frac{\mathrm{E}^{\mathbb{Q}}_{n-1}\left[(\xi_n - \xi_{n-1})p'_n\right]}{\mathrm{E}^{\mathbb{Q}}_{n-1}\left[(\xi_n - \xi_{n-1})^2\right]} = \frac{\mathrm{E}^{\mathbb{Q}}_{n-1}\left[(e^{r_n} - 1)p'_n\right]}{\xi_{n-1}\mathrm{E}^{\mathbb{Q}}_{n-1}\left[(e^{r_n} - 1)^2\right]}, \tag{1.10}$$

where $p'_{n-1} = \mathrm{E}^{\mathbb{Q}}_{n-1}[p'_n]$ and $p'_\tau = c_\tau/k_0$. For example, $p'(\xi) = (\xi - 1)^+$ for a call option and $p'(\xi) = (1 - \xi)^+$ for a put option. Since Eq. (1.10) has no explicit dependence on initial moneyness, entire cross-sections are solved under a *single* application of dynamic programming, greatly decreasing the overall computational burden of empirical tests.

Until now, we purposely left distributions unspecified. In particular, we must still specify how expectations in Eq. (1.10) are estimated. It is a widely accepted empirical fact that market returns and GARCH-type innovations have non-zero higher order cumulants; see e.g. Engle (2002). While various non-Gaussian distributions have been studied, we adopt a semi-parametric approach by using estimated model innovations, referred to as *residuals*.

Test features are first highlighted under the most tractable S0 setting. Residuals under the static volatility model are first retrieved according to,

$$\hat{\eta}_t = \frac{r_t - \hat{u}}{\sqrt{\hat{\delta}}},$$

where (as previously) t is an index over our *entire* sample; i.e. $\{r_t\}$ represents market returns from 2000 to 2014. We then solve static protocols according to,

$$
\begin{aligned}
p'_{n-1}(\xi) &= \frac{1}{T}\sum_{t=1}^{T} \breve{p}'_n(\xi e^{\hat{\mathscr{R}}(\hat{\delta},\hat{\eta}_t)}), \\
\phi_n(\xi) &= \frac{\sum_{t=1}^{T}\left(e^{\hat{\mathscr{R}}(\hat{\delta},\hat{\eta}_t)}-1\right)\breve{p}'_n(\xi e^{\hat{\mathscr{R}}(\hat{\delta},\hat{\eta}_t)})}{\xi\sum_{t=1}^{T}\left(e^{\hat{\mathscr{R}}(\hat{\delta},\hat{\eta}_t)}-1\right)^2},
\end{aligned}
\tag{1.11}
$$

for $n = \tau, \dots, 1$, where $\hat{\mathscr{R}}$ is the empirical counterpart of $\mathscr{R}$ given by,

$$
\hat{\mathscr{R}}(h,\eta) = \eta h^{1/2} - \log\left(\frac{1}{T}\sum_{t=1}^{T} e^{\hat{\eta}_t h^{1/2}}\right), \tag{1.12}
$$

and depends on the entire sample of residuals.

To meet requirements of the EGP, the S0 specification under $\mathbb{Q}$ must respect two conditions. First, $\mathbb{Q}$ innovations must have the same law as under $\mathbb{P}$. This is met by summing over the empirical distribution $\{\hat{\eta}_t\}_{t=1}^{T}$. Second, the model must satisfy the martingale condition Eq. (1.4), which may be written in terms of log-returns as $\mathrm{E}_{n-1}^{\mathbb{Q}}\left[e^{r_n}\right] = 1$ for $n = 1, \dots, \tau$. This property is guaranteed on a sample-by-sample basis by construction, i.e. $\frac{1}{T}\sum_{t=1}^{T} e^{\hat{\mathscr{R}}(\hat{\delta},\hat{\eta}_t)} = 1$, in a manner reminiscent of the *empirical martingale simulation* of Duan and Simonato (1998). Efficiency gains when pricing arguably translate to our hedging setting, but a rigorous demonstration of this statement falls outside the scope of this paper.

Here, we follow an "in-sample" approach in the sense that a protocol is estimated using parameters and residuals calibrated on data from 2000 to 2014 and used throughout. We do so in a spirit similar to Bates (2003), i.e. to ensure each model is taken seriously "as a genuine data generating process". This choice is once again consistent with our focus on expanding one's information set, this time by avoiding frequent recalibrations and time-varying model parameters.

Our semi-parametric approach arguably raises in-sample bias concerns. When hedging monthly options however, preliminary tests suggest excluding any 21-day residual block, i.e. $\{\hat{\eta}_t\}_{t=t_0}^{t_0+20}$, from sums in Eq. (1.11) does not have a significant impact on protocols. For robustness, we present out-of-sample results (by re-estimating models on a yearly basis) in the supplementary material.

Under the LF class, we similarly estimate residuals according to

$$\hat{\eta}_t = \frac{r_t - \hat{u}}{\sqrt{\hat{h}_t}}$$

where $\hat{h}_t$ are the filtered states from any of the three LF specifications, namely LF0, LFG or LFA. We solve LF protocols as functions of two variables since $\{\xi_n, h_{n+1}\}$ is a Markovian process,

$$\begin{aligned} p'_{n-1}(\xi,h) &= \frac{1}{T}\sum_{t=1}^{T} \breve{p}'_n(\xi e^{\hat{\mathscr{R}}(h,\hat{\eta}_t)}, \hat{\mathscr{H}}(h,\hat{\eta}_t)), \\ \phi_n(\xi,h) &= \frac{\sum_{t=1}^{T}\left(e^{\hat{\mathscr{R}}(h,\hat{\eta}_t)}-1\right)\breve{p}'_n(\xi e^{\hat{\mathscr{R}}(h,\hat{\eta}_t)}, \hat{\mathscr{H}}(h,\hat{\eta}_t))}{\xi\sum_{t=1}^{T}\left(e^{\hat{\mathscr{R}}(h,\hat{\eta}_t)}-1\right)^2}, \end{aligned} \tag{1.13}$$

for $n = \tau, \dots, 1$, where $\hat{\mathscr{H}}$ can be found in Table 1.5, with,

$$\hat{\eta}^\star(h,\eta) = \eta - \frac{\hat{u}}{\sqrt{h}} - \frac{1}{\sqrt{h}}\log\left(\frac{1}{T}\sum_{t=1}^{T} e^{\hat{\eta}_t h^{1/2}}\right), \tag{1.14}$$

the empirical counterpart of previously introduced starred innovations $\eta^\star$. As already mentioned, this innovation adjustment ensures the conditional variance process is minimally disrupted by the change of measure —as per the EGP. Both $\hat{\mathscr{R}}(h,\eta)$ and $\hat{\eta}^\star(h,\eta)$ implicitly rely on the entire sample through $\{\hat{\eta}_t\}_{t=1}^{T}$, which in this case represents residuals filtered using either LF0, LFG or LFA.

Under the HF class, we first retrieve bivariate residuals as

$$\left(\hat{\eta}_t = \frac{r_t - \hat{u}}{\sqrt{\hat{h}_t}}, \quad \hat{\varepsilon}_t = \frac{v_t}{\hat{\mu}_t}\right),$$

for the HF0 and HFG models, with $\hat{\varepsilon}_t = v_t^\star/\hat{\mu}_t$ computed from realized semi-variances in the HFA case. We then solve HF protocols,

$$\begin{aligned} p'_{n-1}(\xi,h) &= \frac{1}{T}\sum_{t=1}^{T} \breve{p}'_n(\xi e^{\hat{\mathscr{R}}(h,\hat{\eta}_t)}, \hat{\mathscr{H}}(h,\hat{\eta}_t,\hat{\varepsilon}_t)), \\ \phi_n(\xi,h) &= \frac{\sum_{t=1}^{T}\left(e^{\hat{\mathscr{R}}(h,\hat{\eta}_t)}-1\right)\breve{p}'_n(\xi e^{\hat{\mathscr{R}}(h,\hat{\eta}_t)}, \hat{\mathscr{H}}(h,\hat{\eta}_t,\hat{\varepsilon}_t))}{\xi\sum_{t=1}^{T}\left(e^{\hat{\mathscr{R}}(h,\hat{\eta}_t)}-1\right)^2}, \end{aligned} \tag{1.15}$$

for $n = \tau, \dots, 1$, where $\hat{\mathscr{H}}$ is also given by Table 1.5. $\hat{\mathscr{R}}$ and $\hat{\eta}^\star$ are still respectively given by Eq. (1.12) and (1.14), but defined according to their respective HF residuals —either HF0, HFG or HFA.

Using the direct proportionality between contemporaneous states under the HF class, we simplified specifications such that $\hat{\mathscr{U}}$ does not explicitly appears in Eq. (1.15). For example, the HF0 specification may be written as,

$$\begin{cases} r_n = \eta_n h_n^{1/2} - \kappa(h_n{}^{1/2}), \\ h_{n+1} = \alpha e^{-\alpha} \varepsilon_n h_n + e^{-\alpha} h_n, \end{cases}$$

when $a = 0$ and $b > 0$. For general choices of $\mathscr{U}$ however, solutions would rely on μ_n, consistently with expressions derived in Section 1.3.

By summing over empirical pairs $(\hat{\eta}_t, \hat{\varepsilon}_t)$, the dependence between η and ε is implicitly characterized by the empirical copula, hence overcoming the challenge of appropriately specifying a parametrized family of copulas.

Table 1.5: Overview of $\hat{\mathscr{H}}$. These expressions are used in conjunction with Eq. (1.13) and (1.15) to respectively compute LF and HF protocols; see also Eq. (1.14) for the definition of $\hat{\eta}^\star$.

Requirement	Specification
Low-Frequency	
Naive (LF0)	$\hat{\mathscr{H}}(h,\eta) = \hat{\alpha} e^{-\hat{\alpha}} \hat{\eta}^\star(h,\eta)^2 h + e^{-\hat{\alpha}} h.$
Mean Reversion (LFG)	$\hat{\mathscr{H}}(h,\eta) = \hat{\delta}(1-\hat{\alpha}-\hat{\beta}) + \hat{\alpha}\hat{\eta}^\star(h,\eta)^2 h + \hat{\beta} h.$
Leverage Effect (LFA)	$\hat{\mathscr{H}}(h,\eta) = \hat{\delta}(1-\hat{\alpha}-0.5\hat{\gamma}-\hat{\beta})$
	$+ (\hat{\alpha} + \mathrm{I}(\hat{\eta}^\star(h,\eta) < 0)\hat{\gamma})\hat{\eta}^\star(h,\eta)^2 h + \hat{\beta} h.$
High-Frequency	
Naive (HF0)	$\hat{\mathscr{H}}(h,\eta,\varepsilon) = \hat{\alpha} e^{-\hat{\alpha}} \varepsilon h + e^{-\hat{\alpha}} h.$
Mean Reversion (HFG)	$\hat{\mathscr{H}}(h,\eta,\varepsilon) = \hat{\delta}(1-\hat{\alpha}-\hat{\beta}) + \hat{\alpha}\varepsilon h + \hat{\beta} h.$
Leverage Effect (HFA)	$\hat{\mathscr{H}}(h,\eta,\varepsilon) = \hat{\delta}(1-\hat{\alpha}-0.5\hat{\gamma}-\hat{\beta})$
	$+ (\hat{\alpha} + \mathrm{I}(\hat{\eta}^\star(h,\eta) < 0)\hat{\gamma})\varepsilon h + \hat{\beta} h.$

Table 1.5 shows the HF class essentially replaces squared adjusted return innovations $\eta^{\star 2}$ with realized variance innovations ε. Our framework hence succeeds in expanding

one's information set *and* maintaining strong comparability under all structural requirements i.e. from naive to leverage effect. We also emphasize that proposed protocols are derived non-myopically. In particular, volatility timing protocols are distinguished by both differences in variance forecasts *and hedging surfaces* i.e. ϕ_n as a function of underlying price and conditional variance. This approach contrasts with naively feeding variance estimators into the B&S delta-hedging formula; e.g. as performed by Bakshi and Kapadia (2003).

Regarding implementation concerns related to the evaluation of $\breve{p}'(\cdot)$, we maximize knot density in areas of high option price curvature and interpolate using splines. More precisely, since m carries most of the curvature around k_0, we concentrate interpolation knots around the strike price using a quadratic transformation. In contrast, p is almost linear in variance, especially at higher levels. We hence consider less variance knots distributed exponentially over h. We extrapolate linearly for simulations falling outside the grid.

For put options, we find the S0 protocol is very close to a B&S delta-hedging strategy calibrated to the long-run variance $\hat{\delta}$. The S0 protocol sells slightly more underlying contracts than B&S for deeply OTM options. This is consistent with a fatter than normal left tail for the empirical distribution of log-returns. In line with unfavorable volatility spikes being empirically related to negative log-returns, protocols allowing for a leverage effect (i.e. LFA and HFA) systematically sell roughly 0 to 5% more underlying contracts than their strictly mean reverting counterparts (i.e. respectively LFG and HFG). Also, LF0 and HF0 hedging surfaces are almost indistinguishable, most likely due to their similar variance term structures. Naive cases hence allow for a rough assessment of incremental economic value derived from differing one day ahead variance forecasts *only*, i.e. excluding the non-myopic impact of variance term structures on hedging surfaces.

1.6 Results and discussion

1.6.1 Preliminary analysis

We now present empirical results for OTM put options with one to three months-to-maturity and with levels of moneyness ranging from 0.9 to 1. These contracts have well-documented premia and are highly liquid. Corresponding results for OTM call options can be found in the supplementary material. The model-free delta-hedging methodology (DH) of Bates (2005) is reported as a volatility timing benchmark relevant to both practitioners and researchers. As pointed out by e.g. Alexander and Nogueira (2007) however, DH is risk-minimizing if and only if the correlation between the variance and the underlying is non-zero. A priori, risk minimizing protocols under a leverage effect (i.e. LFA and HFA) hence over-perform DH in terms of risk reduction. Un-hedged (UH) results are also presented for completeness.

Table 1.6 presents descriptive statistics for π. The hedged option premium is always above 100 bps/yr. Under DH and S0, it is maximized for at-the-money(ATM) options, in line with the volatility risk premium interpretation of Bakshi and Kapadia (2003). Under LF and HF cases however, the premium is maximized for OTM options i.e. $k_0 = 0.95$. Correlations with market returns are significantly lower under LFA and HFA, consistent with protocols under a leverage effect systematically selling more underlying contracts.

Sample means for π^2 (not shown) decrease with increasing model likelihoods, as is expected from our risk minimizing objective. In particular, allowing for a leverage effect leads to the most significant risk reduction. This observation confirms the strong linkage between forecasting abilities and hedging effectiveness and is consistent with the preliminary numerical experiment presented in Section 1.1.

Sharpe ratios in Table 1.6 provide a preliminary assessment of economic value. They increase steadily from S0 to LF and from LF to HF. This effect appears to be mostly driven by a decrease in risk as captured by standard deviations —once again in line with the preliminary numerical experiment. But OTM Sharpe ratios *decrease* from 1.15 for

HF0 to 1.09 when allowing for a leverage effect.

Hedged option returns have significantly greater Sharpe ratios than the market in all cases, with a market and hedged option correlation typically below 0.4. These last statistics provide strong ex-post motivation for selling and hedging options, e.g. towards diversifying an investor's portfolio using premia embedded in options. Still, the Sharpe ratio may not be appropriate here due to the large kurtosis. We next turn to proposed utility-based metrics and main results of this paper.

Table 1.6: Descriptive statistics for hedged put returns with 21 TDM. The first row(M) shows statistics for corresponding market returns. The mean(Mean), standard deviation(Std) and 95% value-at-risk(VaR) are presented as bps/yr. The Sharpe ratio(Shrp) is the first column (i.e. Mean) divided by the second column (Std). All statistics are computed from strictly non-overlapping subsamples which are then averaged, including one lag autocorrelation(Auto) and correlation with market returns(Corr).

		Mean	Std	Shrp	Skew	Kurt	Auto	VaR	Corr
	M	560.44	1,679.71	0.33	-1.04	7.28	0.00	-2,182.68	1.00
At-the-Money ($k_0 = 1$)									
	UH	622.07	1,095.09	0.57	-2.68	14.43	-0.02	-1,830.56	0.89
	DH	231.48	288.02	0.80	-1.59	9.86	0.12	-408.50	0.58
	S0	173.41	252.96	0.69	-2.73	21.58	0.12	-484.34	0.32
LF	LF0	176.63	209.73	0.84	-0.95	11.93	0.13	-298.70	0.32
	LFG	177.62	204.87	0.87	-1.21	11.74	0.13	-298.70	0.33
	LFA	164.36	188.74	0.87	-0.55	9.57	0.08	-238.82	0.08
HF	HF0	182.41	202.24	0.90	-1.17	11.32	0.12	-291.98	0.36
	HFG	173.48	188.63	0.92	-0.99	9.50	0.13	-247.62	0.30
	HFA	167.95	180.91	0.93	-0.65	8.60	0.10	-225.22	0.12
Out-of-the-Money ($k_0 = 0.95$)									
	UH	415.00	667.02	0.62	-4.79	36.88	-0.02	-1,501.34	0.69
	DH	205.09	252.84	0.81	-2.94	18.16	0.24	-424.59	0.49
	S0	138.75	296.47	0.47	-3.78	30.49	0.20	-626.35	0.33
LF	LF0	197.95	200.42	0.99	-2.17	17.40	0.20	-337.32	0.28
	LFG	190.19	199.98	0.95	-2.44	18.11	0.21	-344.38	0.35
	LFA	171.69	176.05	0.98	-1.24	14.10	0.17	-274.73	0.04
HF	HF0	215.17	186.48	1.15	-2.08	17.32	0.18	-309.84	0.36
	HFG	202.86	185.06	1.10	-1.85	16.56	0.20	-300.81	0.34
	HFA	190.23	175.17	1.09	-1.49	16.00	0.19	-280.18	0.18
Deeply Out-of-the-Money ($k_0 = 0.9$)									
	UH	253.01	408.81	0.62	-6.42	65.09	-0.05	-1,092.05	0.51
	DH	131.06	206.03	0.64	-4.95	39.35	0.22	-454.17	0.39
	S0	106.96	278.26	0.38	-4.55	38.94	0.12	-619.05	0.42
LF	LF0	142.02	164.02	0.87	-3.48	28.65	0.21	-317.06	0.25
	LFG	138.13	163.97	0.84	-3.77	28.88	0.23	-317.83	0.34
	LFA	120.11	137.66	0.87	-1.42	18.14	0.21	-220.98	0.00
HF	HF0	148.89	146.75	1.01	-3.07	26.10	0.22	-270.13	0.31
	HFG	139.74	149.43	0.94	-2.84	23.86	0.25	-270.37	0.32
	HFA	130.01	139.98	0.93	-2.26	21.60	0.25	-244.81	0.18

1.6.2 Economic value measurements

Table 1.7 presents cross-sections of *absolute* economic value metrics Θ_U under previously calibrated risk aversion parameters. Figure 1.2 displays corresponding cross-sections of *incremental* economic value and provides a continuous view over the moneyness-maturity half-plane.

We readily observe strictly positive gains from performing volatility timing (i.e. from S0 to LF) and from relying on realized variance-based protocols (i.e. from LF to HF) under all structural requirements and over the entire moneyness-maturity half-plane. Volatility timing under LF indeed generates economic value over S0 ranging from roughly 22 to 132 bps/yr under the naive case (i.e. LF0-S0), 50 to 153 bps/yr under the mean reversion case (i.e. LFG-S0) and 53 to 190 bps/yr under the leverage effect case (i.e. LFA-S0). More interestingly, volatility timing under HF generates gains over the LF class ranging from 10 to 67 bps under the naive case (i.e. HF0-LF0), 5 to 95 bps under the mean reversion case (i.e. HFG-LFG) and 7 to 27 bps under the leverage effect case (i.e. HFA-LFA).

On an absolute basis, naive specifications —i.e. LF0 and HF0— surprisingly perform best for short-term OTM options. This suggests accuracy of one day ahead variance forecasts is the main driver of economic value, as opposed to differences in hedging surfaces. The over-performance, however, quickly dissipates for closer-to-the-money options and longer-term maturities, The leverage effect appears to be most relevant for ATM options under LF. Interestingly, we do not observe similar gains under HF, suggesting realized variances already capture most of the information related to the leverage effect.

Regarding DH, risk minimizing protocols over-perform under all metrics and specifications. Market segmentation or other non-arbitrage effects (e.g. a volatility risk premium) significantly impacting option prices presumably yields DH protocols that are suboptimal under global concerns. In the leverage effect case, the strong over-performance of risk minimizing protocols is further consistent with DH omitting the partial derivative of option prices with respect to variance; e.g. as documented by Garcia and Renault (1998).

Table 1.7: Cross-section of economic value Θ_U from selling and hedging put options.

		Trading-Days-to-Maturity (τ)				
		21	32	42	53	63
		At-the-Money ($k_0 = 1$)				
	UH	-6,738.76	-11,328.20	-11,856.53	-12,257.27	-10,732.84
	DH	102.30	61.80	45.16	34.32	5.27
	S0	49.49	-11.13	-80.76	-73.16	-64.19
LF	LF0	113.92	87.85	51.47	8.22	-37.37
	LFG	117.18	96.96	72.43	46.80	20.20
	LFA	115.26	112.10	109.35	100.32	89.41
HF	HF0	124.01	102.35	78.72	56.32	30.32
	HFG	123.59	123.48	126.99	124.86	115.52
	HFA	122.94	124.21	126.48	123.61	114.79
		Out-of-the-Money ($k_0 = 0.95$)				
	UH	-395.92	-1,235.17	-2,454.53	-3,909.42	-4,887.83
	DH	149.65	139.24	134.64	113.15	85.74
	S0	51.39	44.02	33.34	39.26	36.58
LF	LF0	164.86	155.72	146.09	119.17	90.66
	LFG	156.97	147.76	139.90	118.63	95.60
	LFA	147.16	140.41	135.44	124.02	113.76
HF	HF0	186.82	179.05	174.29	156.21	134.79
	HFG	175.17	172.01	171.85	159.55	146.15
	HFA	165.78	162.27	160.90	150.46	140.34
		Deeply Out-of-the-Money ($k_0 = 0.9$)				
	UH	40.76	46.41	-161.93	-595.05	-659.07
	DH	97.96	99.67	103.17	102.31	87.44
	S0	40.75	39.68	38.46	38.11	39.32
LF	LF0	122.60	127.85	126.46	119.31	108.12
	LFG	118.60	120.69	117.36	108.65	97.18
	LFA	107.17	106.01	103.55	100.64	92.38
HF	HF0	133.60	143.86	145.47	143.06	135.56
	HFG	124.00	130.77	131.73	131.44	125.17
	HFA	116.41	120.84	121.26	122.64	116.94

Figure 1.2: Cross-section of *incremental* economic value Θ_U from selling and hedging put options.

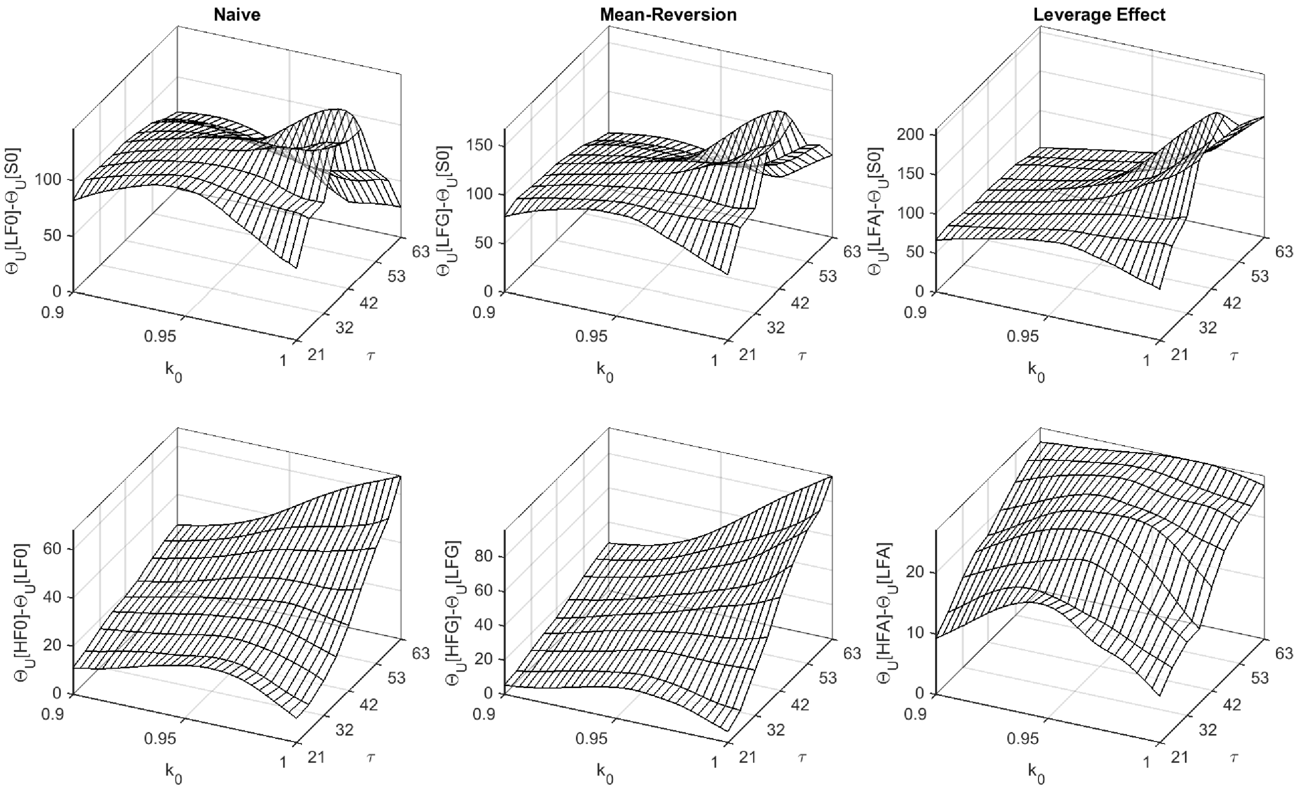


Table 1.8: Subsample analysis of Θ_U for 63 TDM OTM ($k_0 = 0.95$) put options. The last column sums the subsample-to-subsample absolute value differences, with higher values indicative of model misspecification.

		Subsample				
		2002-2004	2005-2007	2008-2011	2012-2014	$\sum$|Diff|
	UH	-1,843.24	152.75	-6,705.78	568.58	16,128.88
	DH	258.36	145.23	-181.04	275.39	895.82
	S0	202.22	101.94	-185.41	144.49	717.53
	LF0	275.70	101.66	-99.46	184.52	659.13
LF	LFG	255.19	104.03	-76.76	188.21	596.92
	LFA	237.52	95.07	62.88	80.83	192.59
	HF0	256.87	120.72	16.58	202.20	425.92
HF	HFG	248.94	109.52	80.56	176.28	264.10
	HFA	242.08	102.01	101.80	132.71	171.20

Table 1.8 presents subsample results for 63 TDM OTM put options. Unsurprisingly, we find that volatility timing is most beneficial during the highly turbulent 2008-2011 period, with incremental gains from LF to HF ranging from roughly 39 to 157 bps. Results for other periods are less definitive. In particular, positive incremental economic value appears to be robust in the leverage effect case *only* —i.e. from LFA to HFA. The last column shows the sum of absolute value differences from subsample-to-subsample. High values suggest misspecification; i.e. under-hedged protocols benefit during calm periods, but fail when markets are stressed. For example, S0 outperforms all volatility timing protocols during 2012-2014 at the cost of a *negative* 185 bps performance during 2008-2011. We find that HF offers the most stability for longer-term options. Allowing for a leverage effect also appears beneficial, e.g. with a subsample-to-subsample variation decreasing from roughly 597 bps for the LFG model to 193 bps for the LFA model.

The strong decline in performance for naive protocols across maturities, together with the wild corresponding subsample variations, suggest naive protocols may be misspecified and under-hedged. In particular, agents concerned with hedging shortfalls likely select either LFA or HFA, as these protocols generally offer a much more stable performance with mild drawdowns. Under the LF class, LFA even delivers a *positive* performance during 2008-2011. Another concern is the strong performance of the un-hedged protocol during 2005-2007 and 2012-2014. Over such short subsamples, we face a Peso problem; i.e. managing risk for put options is never beneficial during strong market rallies. Given the extreme negative performance of roughly -67% during 2008-2011, risk-averse agents are unlikely to select the un-hedged protocol on an ex-ante basis.

We confirm these intuitions in the supplementary material using a prospect theory metric under loss aversion. We calibrate this metric towards deterring agents from selling un-hedged options. The resulting metric heavily focuses on hedging shortfalls and is arguably more consistent with our hedging objective. Protocols with a leverage effect now systematically outperform exponential smoothing specifications, in line with misspecification suspicions. In absolute, the HFA model performs best.

In the supplementary material, we test robustness to calibration procedures, vega-constant selling strategies in the option market, and out-of-sample protocols. While some variations are observed, the positiveness of the incremental value from LF to HF is robust in all cases.

1.7 Conclusion

This paper embraces market incompleteness in order to provide a novel investment setting for testing the economic value of volatility timing in the S&P 500 option market. We set aside option pricing concerns and instead provide empirical insights on risk minimization. This setting is motivated by the empirical fact that hedged net short index option inventories are profitable in the long-run for a wide range of risk management protocols.

Our empirical results show the statistical value of using realized variance does translate to significant economic value. This suggests *statistical* arbitrage opportunities persist in the option market: agents selling and hedging options under better variance forecasts have better risk-adjusted returns.

Since our focus is on the economic impact of expanding one's information set, we prioritize comparability between low- and high-frequency models in terms of model structure, long-run expectations and the dimensionality of Markovian processes. This puts stringent constraints on high-frequency models which could be relaxed in practice in order to further improve economic value.

Under the model of Gârleanu et al. (2009), time variations in ex-ante π expectations are likely driven by other factors than underlying variance, such as the prevailing risk aversion of a representative agent and the put buying demand of end-users e.g. pension plans. This observation motivates our focus on volatility timing for underlying positions, as opposed to option positions. In particular, empirical inquiries related to *when* to sell options, *which* option contracts to sell, and *how many* short option positions to hold, likely benefit from inventory data —a conjecture left for future research.

While we do not provide pricing implications, our endeavor is still consistent with arbitrage-free option prices resulting from complex interactions between option buying end-users, profit-oriented proprietary traders and liquidity providing market-makers — each having their own individual views about future volatility. In particular, our results could likely be interpreted in the context of a three-agent equilibrium under which market-makers redistribute risk from end-users to proprietary traders. Under our held-until-maturity assumption, proprietary traders presumably handle excess inventories more efficiently by disregarding shocks on intermediary option prices (e.g. shocks in option demand) which are costly to hedge. A quantity of interest would be the overall welfare gains from disentangling the role of market-makers and proprietary traders in the model of Gârleanu et al. (2009), i.e. respectively (1) scalping and hedging option demand risk and (2) volatility timing and hedging underlying market risk.

References

Alexander, C. and Nogueira, L. M. (2007). Model-free hedge ratios and scale-invariant models. *Journal of Banking and Finance*, 31(6):1839–1861.

Bakshi, G. and Kapadia, N. (2003). Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies*, 16(2):527–566.

Bandi, F. M., Russell, J. R., and Yang, C. (2008). Realized volatility forecasting and option pricing. *Journal of Econometrics*, 147(1):34–46.

Barone-Adesi, G., Engle, R. F., and Mancini, L. (2008). A GARCH option pricing model with filtered historical simulation. *Review of Financial Studies*, 21(3):1223–1258.

Basak, S. and Chabakauri, G. (2012). Dynamic hedging in incomplete markets: A simple solution. *Review of Financial Studies*, 25(6):1845–1896.

Bates, D. S. (2003). Empirical option pricing: A retrospection. *Journal of Econometrics*, 116(1):387–404.

Bates, D. S. (2005). Hedging the smirk. *Finance Research Letters*, 2(4):195–200.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

Bollerslev, T. and Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariance. *Econometric Reviews*, 11(1):143–172.

Brandt, M. W. (2003). Hedging demands in hedging contingent claims. *Review of Economics and Statistics*, 85(1):119–140.

Brownlees, C. T. and Gallo, G. M. (2010). Comparison of volatility measures: A risk management perspective. *Journal of Financial Econometrics*, 8(1):29–56.

Buraschi, A. and Jackwerth, J. (2001). The price of a smile: Hedging and spanning in option markets. *The Review of Financial Studies*, 14(2):495–527.

Christoffersen, P., Feunou, B., Jacobs, K., and Meddahi, N. (2014). The economic value of realized volatility: Using high-frequency returns for option valuation. *Journal of Financial and Quantitative Analysis*, 49(3):663–697.

Corsi, F., Fusari, N., and Vecchia, D. L. (2013). Realizing smiles: Options pricing with realized volatility. *Journal of Financial Economics*, 107(2):284–304.

Coval, J. D. and Shumway, T. (2001). Expected option returns. *The Journal of Finance*, 56(3):983–1009.

Duan, J.-C. (1995). The GARCH option pricing model. *Mathematical Finance*, 5(1):13–32.

Duan, J.-C. and Simonato, J.-G. (1998). Empirical martingale simulation for asset prices. *Management Science*, 44(9):1218–1233.

Elliott, R. J. and Madan, D. B. (1998). A discrete time equivalent martingale measure. *Mathematical Finance*, 8(2):127–152.

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, 17(5):425–446.

Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1):3–27.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, 56(1):329–352.

Fleming, J., Kirby, C., and Ostdiek, B. (2003). The economic value of volatility timing using “realized” volatility. *Journal of Financial Economics*, 67(3):473–509.

Föllmer, H. and Schied, A. (2011). *Stochastic Finance. An Introduction in Discrete Time*. Graduate Textbook Series. Walter De Gruyter, Berlin, 3rd edition.

Garcia, R. and Renault, E. (1998). A note on hedging in ARCH and stochastic volatility option pricing models. *Mathematical Finance*, 8(2):153–161.

Gârleanu, N., Pedersen, L. H., and Poteshman, A. M. (2009). Demand-based option pricing. *Review of Financial Studies*, 22(10):4259–4299.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.

Harrison, J. M. and Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215–260.

Heber, G., Lunde, A., Shephard, N., and Sheppard, K. (2009). Oxford-Man Institute's realized library. Oxford-Man Institute, University of Oxford, Version 0.2.

Heston, S. L. and Nandi, S. (2000). A closed-form GARCH option valuation model. *Review of Financial Studies*, 13(3):585–625.

Ingersoll, J., Spiegel, M., Goetzmann, W., and Welch, I. (2007). Portfolio performance manipulation and manipulation-proof performance measures. *The Review of Financial Studies*, 20(5):1503–1546.

Markowitz, H. M. (1952). Portfolio Selection. *Journal of Finance*, 7(1):77–91.

Nandi, S. (1998). How important is the correlation between returns and volatility in a stochastic volatility model? Empirical evidence from pricing and hedging in the S&P 500 index options market. *Journal of Banking and Finance*, 22(5):589–610.

Schweizer, M. (1995). Variance-optimal hedging in discrete time. *Mathematics of Operations Research*, 20(1):1–32.

Shephard, N. and Sheppard, K. (2010). Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 25(2):197–231.

Chapter 2

Supplementary Material for “The Economic Value of Volatility Timing using Realized Volatility for Hedged S&P 500 Index Options”

HUGO LAMARRE

DEBBIE J. DUPUIS

BRUNO RÉMILLARD

Abstract

The supplementary material proceeds as follows. Section 2.1 interprets the normalization of option prices as a change of numéraire. Section 2.2 synchronizes underlying spot prices with option cross-sections by performing rate term structure regressions on a daily basis. Section 2.3 presents the option data reduction procedure, including an overview of the raw dataset and a data cleaning methodology. Section 2.4 performs robustness checks for empirical results and Section 2.5 concludes with tests of economic value for call options.

2.1 Change of numéraire

We consider a synthetic forward contract denoted by m_n^τ. Such a contract may be viewed as a non-dividend-paying asset, namely $(S_n - D_n^\tau)$, relative to the numéraire F_0^τ/B_n^τ, i.e. a zero-coupon bond maturing in τ trading days with notional value F_0^τ. Option prices relative to the numéraire given by Eq. (1.2) are interpreted as stationary prices for option contracts written on asset m_n^τ under no dividends nor interests. This interpretation relies on three assumptions.

Firstly, the underlying process is scale-invariant, i.e. the distribution of index returns does not depend on the level measurement scale. Alexander and Nogueira (2007) show the homogeneity of a payoff is preserved by the pricing operator under scale-invariance. Secondly, an option is priced as a function of a forward contract matching its maturity. This last assumption is reasonable given a futures contract is the most common hedging instrument in practice.

Under these first two assumptions, we have,

$$C_0(S_0, K_0, \tau, B_0^\tau, D_0^\tau) = \frac{f(F_0^\tau, K_0, \tau)}{B_0^\tau},$$

$$f(\alpha F_0^\tau, \alpha K_0, \tau) = \alpha f(F_0^\tau, K_0, \tau), \quad \forall \alpha > 0.$$

We obtain c_0 by setting α to $1/F_0^\tau$,

$$C_0(S_0, K_0, \tau, B_0^\tau, D_0^\tau) B_0^\tau / F_0^\tau = f(F_0^\tau/F_0^\tau, K_0/F_0^\tau, \tau) \equiv c_0(k_0, \tau).$$

Thirdly, we assume options are free of static arbitrage opportunities, a sufficient condition for the existence of an *arbitrary* risk-neutral measure $\tilde{\mathbb{Q}}$ under which C_0 is a discounted expectation; see e.g. Carr and Madan (2005). For call options,

$$\begin{aligned} c_0(k_0, \tau) &= \left(\frac{1}{B_0^\tau} \mathrm{E}^{\tilde{\mathbb{Q}}}[(S_\tau - K_0)^+] \right) B_0^\tau / F_0^\tau, \\ &= \mathrm{E}^{\tilde{\mathbb{Q}}}[(F_\tau^\tau / F_0^\tau - K_0/F_0^\tau)^+], \\ &= \mathrm{E}^{\tilde{\mathbb{Q}}}[(m_\tau^\tau - k_0)^+]. \end{aligned}$$

The aforementioned interpretation follows.

In practice, volatility timing protocols are likely implemented using futures contracts and a margin account. The committed capital covers initial margin requirements and subsequent margin calls. Hedged option returns hence act as excess rates of return if one unit of numéraire is initially committed. In this case, the value of a margin account has a lower absorbing bound, namely $\pi = -1$, at which point positions are forcedly unwound by the brokerage firm and an agent is ruined.

An alternative numéraire commonly used is the initial option price. This choice assumes significantly more leverage and generates larger absolute rates of return. Unscheduled cash infusions are most likely needed on a regular basis to keep the strategy afloat, more so when markets are stressed and availability of capital is low. Using option prices as numéraire hence understates the effect of forced liquidation; see Santa-Clara and Saretto (2009). In contrast, the probability of margin requirements depleting all committed capital is low under our numéraire —at least under reasonable protocols. In this sense, our empirical results are robust to margin requirements.

The change of numéraire could presumably impact how agents cast market views in term of a martingale measure. If an equivalent martingale measure exists under which the *discounted* asset $S_n - D_n^\tau$ is a martingale, a τ-forward measure also exists under which $S_n - D_n^\tau$ relative to the proposed numéraire is a martingale, and vice-versa; see e.g. Example 2 of Geman et al. (1995). Under deterministic interest rates and dividends, we can show the two measures coincide. Investors are thus indifferent between applying the EGP to S_n or to m_n^τ.

The change of numéraire greatly streamlines classical option pricing results. For example, usual arbitrage-free bounds are expressed as $c_0 \in ((k_0-1)^+, k_0)(c_0 \in ((1-k_0)^+, 1))$ for put(call) options. Put-call parity is expressed as $1 - k_0 = c_0^c - c_0^p$, where $c_0^p(c_0^c)$ is a normalized put(call) price. The B&S pricing model is expressed as $c_0 = k_0 N(-d_-) - N(-d_+)(c_0 = N(d_+) - k_0 N(d_-))$ for put(call) options where $d_\pm = \big(-\log(k_0) \pm \sigma^2\tau/2\big)/\sigma\sqrt{\tau}$ and σ is the single model parameter.

2.2 Underlying spot price synchronization

Our test relies on an non-observable forward contract matching the maturity of an option. With interest rate and dividend realizations assumed known, the only remaining source of uncertainty in forward contract price is the underlying *spot* price. Option cross-sections are likely sampled at a slightly different time than official S&P 500 close prices. Our purpose is to estimate $S_0^\star$, the underlying spot price used by market participants to price options, and use it in lieu of the official close price during tests.

We first estimate option-implicit bank account values using unsynchronized prices S_0 and put-call parity $F_0^\tau - K_0 = B_0^\tau(C_0^c - C_0^p)$, where C_0^p and C_0^c are respectively a put and call *mid-quote*,

$$B_0^\tau(K_0) = \frac{K_0}{(S_0 - D_0^\tau) - (C_0^c - C_0^p)}. \tag{2.1}$$

This procedure is common in the literature; see e.g. Jackwerth and Rubinstein (1996), Ait-Sahalia and Lo (1998), or Buraschi and Jackwerth (2001). For all maturities, we estimate a *single* option-implicit bank account $\hat{B}_0^\tau$ as the median of all near-the-money (NTM) evaluations of Eq. (2.1)[1]. We also introduce $\widetilde{B}_0^\tau$ the value at time τ of one dollar invested at time 0 in U.S. Treasury instruments[2].

Table 2.1 presents descriptive statistics for total interest rates defined as logarithms of bank account values. More precisely, implicit rates (IR) and risk-free rates (RFR) are respectively given by $\log(\hat{B}_0^\tau)$ and $\log(\widetilde{B}_0^\tau)$.

IRs generally follow RFRs over the period, with a sustained hike from the mid-2000s to the banking crisis and a strong subsequent drop. Biases between IR and RFR, computed as the median of $\mathrm{IR} - \mathrm{RFR}$, are typically within ±4 bps, with the exception of longer-term

[1] More precisely, $\hat{B}_0^\tau$ is computed as a median over all options such that $K/((S_0 - D_0^\tau)\tilde{B}_0^\tau) \in [0.95, 1.05]$, where $\widetilde{B}_0^\tau$ acts as a preliminary bank account estimate and is defined hereinafter. Quotes with no bid price or with a bid price greater than its ask price are excluded from computations. If $\hat{B}_0^\tau$ can not be computed for a given maturity, we linearly interpolate or horizontally extrapolate *annual* implicit rates $\log(\hat{B}_0^\tau) \times 252/\tau$. We interpolate for 212 estimates over a total of 45,721; 108 out of these were linear extrapolations.

[2] $\widetilde{B}_0^\tau$ is inferred from the zero-coupon yield curve provided by OptionMetrics. We linearly interpolate or horizontally extrapolate the curve and use an "actual over 365" day-count convention with continuous compounding, i.e. $\widetilde{B}_0^\tau = \exp((r(\tilde{\tau}) \times \tilde{\tau})/(100 \times 365))$ where $r(\tilde{\tau})$ is the extrapolated interest rate at $\tilde{\tau}$ and $\tilde{\tau}$ is the number of *calendar* days corresponding to τ *trading* days.

Table 2.1: Descriptive statistics of option-implicit and risk-free total rates by subsample and TDM. The first two columns show the median (Med) and median absolute deviation (MAD) of IR$-$RFR where IR and RFR are respectively given by $\log(\hat{B}_0^\tau)$ and $\log(\widetilde{B}_0^\tau)$. The last column (X-MAD) is the MAD of $\log(B_0^\tau(K_0)) - \log(\hat{B}_0^\tau)$ across strike prices. All values are presented in bps.

		IR − RFR		IR
TDM (τ)	Subsample	Med	MAD	X-MAD
$[0.0, 10.0)$	2002 – 2004	2.83	9.47	1.65
	2005 – 2007	1.93	5.93	0.95
	2008 – 2011	-0.44	3.85	1.33
	2012 – 2014	0.00	3.26	0.94
$[10.0, 31.0)$	2002 – 2004	0.95	9.29	1.52
	2005 – 2007	0.39	6.36	0.99
	2008 – 2011	-0.91	4.16	1.64
	2012 – 2014	0.94	3.42	1.04
$[31.0, 63.0)$	2002 – 2004	0.54	9.34	1.58
	2005 – 2007	-0.17	6.50	0.81
	2008 – 2011	-2.07	5.20	1.68
	2012 – 2014	3.34	4.02	1.36
$[63.0, 252.5)$	2002 – 2004	2.77	10.15	1.91
	2005 – 2007	0.31	7.46	1.19
	2008 – 2011	-3.88	7.61	2.35
	2012 – 2014	13.19	8.80	2.46

options having a bias above 10 bps towards the end of the sample. Dispersions in IR and RFR differences, computed as the median absolute deviation (MAD) of IR − RFR, range from roughly 3 to 10 bps.

The last column (X-MAD) shows cross-sectional dispersions in option-implicit bank account estimates i.e. across strikes. Corresponding values are typically well below 3 bps, suggesting put-call parity holds tightly for any given maturity.

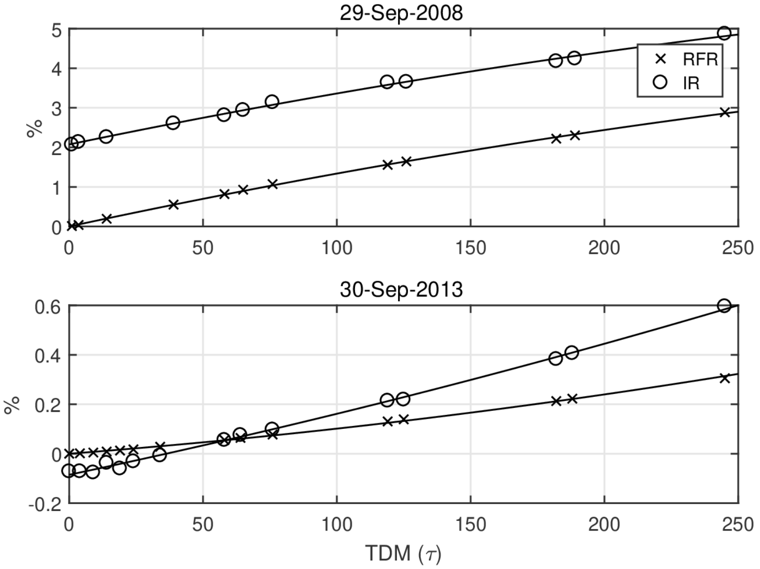
The partial derivative of IR with respect to $\log(S_0)$ converges to one for ATM options approaching maturity[3]. The underlying spot price is thus the overwhelming source of IR errors for short-term NTM options, with a 1% error in S_0 roughly resulting in a 1% error in option-implicit bank account value. We hence expect *intercepts* in IR term structure to

[3] The partial derivative is given by $S_0/((S_0 - D_0^\tau) - (C_0^c - C_0^p))$, where, by definition, $C_0^c - C_0^p = 0$ for ATM options and $D_0^\tau = 0$ when $\tau = 0$.

be informative of synchronization issues.

For example, the top panel of Figure 2.1 displays an extreme manifestation of this effect on 29-Sept-2008, a highly volatile day following the bankruptcy of Lehman Brothers. From 4:00PM to 4:15PM, the index increased by roughly 2%, consistent with the observed gap in term structure between IRs and RFRs. We inspected several other days in late 2008 with large market variations in the last minutes of trading and observed consistent gaps. The bottom panel of Figure 2.1 shows a large difference in slope between IR and RFR term structures. This observation is typical for the last two years of our sample and consistent with previously reported biases for long-term maturities.

Figure 2.1: Total implicit (IR) and risk-free (RFR) rates for all available maturities below 252 TDM on 29-Sept-2008 and 30-Sept-2013. Solid lines represent the estimated regression equation (2.2).



We perform the following regression *on a daily basis* for expirations with less than 252 TDM,

$$\log(b) = \beta_{10}\tau + \beta_{01}x + \beta_{11}x\tau + \beta_{20}\tau^2 + \varepsilon, \tag{2.2}$$

where x is 1 or 0 whether b is respectively given by the option-implicit $\hat{B}_0^\tau$ or the U.S. Treasury $\tilde{B}_0^\tau$ bank account and ε is a white noise. On any given day, $\hat{\beta}_{01}$ is interpreted as the synchronization error and $\hat{\beta}_{11}$ as the difference in IR and RFR slopes.

Resulting R^2 values are always above 0.9 and have a sample mean of 0.997. Table 2.2 presents yearly statistics for $\hat{\beta}_{01}$ and $\hat{\beta}_{11}$. Average IR intercept $\hat{\beta}_{01}$ is typically below 2 bps in absolute value. Slope differences $\hat{\beta}_{11}$ display significantly more persistence than $\hat{\beta}_{01}$, with auto-correlations often above 0.60. These results are consistent with intercepts capturing transient and unbiased measurement errors in S_0, but slope differences capturing persistent effects which could be economically significant, e.g. related to the relative ability of an agent to finance short- versus long-term option strategies. Overall, our two-factor characterization of option-implicit and U.S. Treasury rates appears well specified.

Table 2.2: Descriptive statistics of estimated parameters β_{01} and β_{11} for the regression equation (2.2). Avg(Std) is a sample mean(standard deviation) over daily observations. AC is the corresponding one-lag autocorrelation. We convert $\hat{\beta}_{01}$ to bps and $\hat{\beta}_{11}$ to *annualized* bps according to respectively $10000 \times \hat{\beta}_{01}$ and $252 \times 10000 \times \hat{\beta}_{11}$.

		$\hat{\beta}_{01}$			$\hat{\beta}_{11}$	
Year	Avg	Std	AC	Avg	Std	AC
2002	0.86	20.70	0.08	6.68	8.88	0.56
2003	1.54	14.12	0.13	7.04	8.52	0.85
2004	-0.92	11.82	-0.00	1.69	4.76	0.71
2005	-1.46	10.05	0.20	-1.82	4.41	0.40
2006	-0.36	10.26	0.11	-2.73	6.80	0.72
2007	-0.25	13.87	0.10	6.64	6.91	0.70
2008	2.60	22.28	0.24	-7.28	23.00	0.93
2009	-3.57	8.76	0.08	-6.39	13.49	0.91
2010	-0.52	5.49	0.09	-5.12	9.97	0.73
2011	-1.08	9.13	0.15	3.97	6.26	0.62
2012	-0.50	5.75	0.02	8.51	8.63	0.58
2013	-1.70	5.64	0.14	36.73	7.34	0.78
2014	-0.86	5.49	0.34	33.36	6.13	0.74

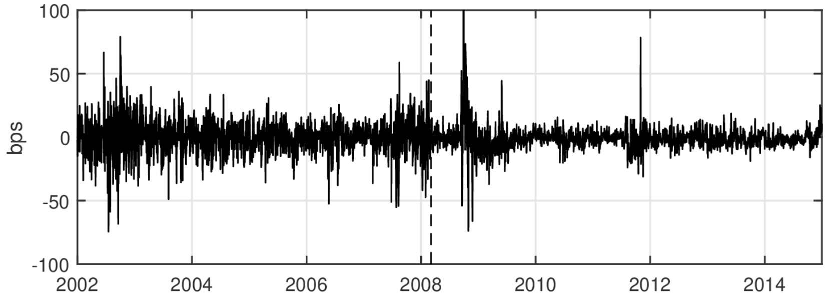
Finally, the synchronized underlying value is given by,

$$\hat{S}_0^{\star} = S_0 e^{\hat{\beta}_{01}}. \tag{2.3}$$

This last expression arises when evaluating Eq. (1.1) at time 0, i.e. $F_0^0 = (S_0 - D_0^0)B_0^0$, with B_0^0 estimated according to Eq. (2.2) for the option-implicit bank account i.e. in the case $x = 1$.

Following manual inspections, the proposed regression successfully identifies days with large index variations in the last minutes of trading throughout our sample. In particular, the procedure appears most relevant prior to 2008, as OptionMetrics improved its sampling methodology in following years. Figure 2.2 still shows large *occasional* corrections after 2008 e.g. during the U.S. debt crisis in late 2011. This last observation is consistent with synchronization issues being more problematic during volatile periods. Overall, observations suggest persisting benefits to relaxing market integration assumptions and inferring underlying spot prices from option cross-sections.

Figure 2.2: Daily time series of $\hat{\beta}_{01}$ in bps from regression equation (2.2). We truncated 29-Sept-2008 (207bps). The dashed vertical line marks 5-Mar-2008, the day on which OptionMetrics implemented significant improvements to its sampling methodology.



2.3 Option data reduction

Distributional properties of hedged option returns most likely depend on moneyness and maturity. For example, Bakshi and Kapadia (2003) show the hedged option premium is maximized for ATM options under stochastic volatility models. Branger and Schlag (2008) extend the characterization of hedged option returns to models with jumps and to discrete-time portfolio revisions. Controlling for moneyness and maturity factors during tests is complicated by daily changes in contract availability, e.g. as markets move, as CBOE adds new strike prices or expiration types, or simply as time passes.

Some papers focus on a single option contract closest to a moneyness-maturity target; see e.g. Buraschi and Jackwerth (2001), Coval and Shumway (2001), or Driessen and Maenhout (2007). To mitigate data rejections, other papers aggregate options falling within a given moneyness-maturity interval; see e.g. Bakshi et al. (1997). Finally, some authors advocate interpolating option prices. For example, Broadie et al. (2007) argue interpolation is a “pragmatic compromise, as it uses nearly all of the information in the cross section of option prices without, in [their] opinion, introducing any substantive biases”. We stand by this opinion. Using best bid and ask prices provided by OptionMetrics, our goal is to build a daily set of arbitrage-free pricing *surfaces*.

While our test relies on interest rate *realizations*, a similar assumption here would create spurious arbitrage opportunities, leading to unwarranted arbitrage-based data rejections and awkward bends in calibrated pricing surfaces. These effects are most problematic for long-term deeply ITM options. During the option data reduction *only*, forward contracts are evaluated using an unsynchronized underlying S_0 and a put-call parity-based bank account $\hat{B}_0^\tau$ defined in Section 2.2. This approach corrects for possible synchronization issues, while reflecting interest rate *expectations* as conveyed by option market prices[4].

Next, we present the raw dataset, extract mid-quotes deemed as sufficiently efficient estimators of market-clearing prices, mitigate induced sparsity by applying put-call parity and finally calibrate surfaces to processed mid-quotes.

2.3.1 Dataset

We register an option observation at (τ, k_0) on any trading day if at least a put or a call quote appears in the raw dataset provided by OptionMetrics. We reject 17 official trading days due to missing data entries in the Oxford-Man Institute’s realized library (Heber et al., 2009), presumably due to the poor quality of corresponding intradaily data. We retrieve roughly 2.7M observations distributed unevenly over $3,256$ trading days from 2002 to 2014, with cross-section presented in Table 2.3. We refer to this set as the set of daily option quotes. This is not to be confused with the set of distinct option *contracts*,

[4] We experimented with various discounting schemes for dividends in Eq. (1.1), including no discounting, and found no material impact on results.

containing roughly $40,674$ put and $40,674$ call contracts, each with a life ranging from a few days to a few years.

Table 2.3: Number of daily option quotes by TDM (τ), subsample, and moneyness (k_0).

				k_0			
TDM (τ)	Subsample	< 0.925	$[0.925, 0.975)$	$[0.975, 1.025)$	$[1.025, 1.075)$	≥ 1.075	Total
$[0.0, 10.0)$	2002 – 2004	5,627	2,361	2,784	2,031	4,243	324,613
	2005 – 2007	10,485	4,976	6,161	3,820	2,482	
	2008 – 2011	41,964	10,261	10,915	9,173	21,830	
	2012 – 2014	100,543	25,853	25,826	21,051	12,227	
$[10.0, 31.0)$	2002 – 2004	11,584	4,280	5,561	4,202	8,862	614,939
	2005 – 2007	23,236	9,661	10,123	8,700	5,535	
	2008 – 2011	81,297	12,711	12,996	12,933	44,762	
	2012 – 2014	197,659	46,044	46,174	40,662	27,957	
$[31.0, 63.0)$	2002 – 2004	14,375	3,426	4,520	3,745	11,624	591,927
	2005 – 2007	22,652	8,166	11,077	8,058	5,663	
	2008 – 2011	94,111	15,252	15,893	15,582	54,081	
	2012 – 2014	178,038	31,949	32,573	30,102	31,040	
$[63.0, 252.5)$	2002 – 2004	27,478	5,333	5,558	5,859	29,064	789,566
	2005 – 2007	43,441	9,948	9,513	7,324	12,823	
	2008 – 2011	123,313	16,827	17,480	16,959	102,784	
	2012 – 2014	220,409	24,091	25,171	24,628	61,563	
≥ 252.5	2002 – 2004	13,851	2,843	2,845	2,290	13,346	380,879
	2005 – 2007	27,701	4,823	3,913	3,158	8,859	
	2008 – 2011	60,778	5,906	6,071	5,979	51,219	
	2012 – 2014	100,835	8,608	8,919	8,583	40,352	
Total		1,399,377	253,319	264,073	234,839	550,316	2,701,924

Table 2.4 shows daily cross-sections have greatly evolved over the years, with the daily number of (τ, k_0) pairs increasing from 250 in 2002 to 3000 in late 2014. In fact, more than 50% of all quotes are observed during the last three years of the sample.

Table 2.4: Yearly sum (Total) and median (Daily) of number of daily option quotes.

Year	Total	Daily
2002	64,861	253
2003	64,412	259
2004	68,419	275
2005	74,795	296
2006	85,300	337
2007	112,203	439
2008	154,819	580
2009	214,588	848
2010	235,818	939
2011	255,852	1,015
2012	324,930	1,298
2013	411,492	1,644
2014	634,435	2,345
Total	2,701,924	-

We now give a brief overview of extensions made to the listing by CBOE. When appropriate, we refer to a CBOE Regulatory or Information Circular identification number as respectively (RCXX-XXX) or (ICXX-XXX).

Table 2.5 shows the bulk of the dataset is composed of *traditional* expirations which settle on the third Friday of each month. On any given trading day, there is a minimum of three consecutive near-term traditional expirations and a total of six concurrent traditional expirations. By the end of 2007, the number of consecutive near-term traditional expirations increased by one, for a total of up to seven (IC07-204).

CBOE gradually introduced new expiration types to complement traditional expirations. *Quarterly* and *end-of-month* expirations were respectively introduced on 21-Feb-2007 (IC07-013) and 07-Jul-2014 (RG14-081), with four concurrent expirations each.

Long Term Equity Anticipation Securities (LEAPs) offer longer-term maturities, typically ranging from one to three years. They are eventually converted back into traditional

Table 2.5: Daily number of expirations and strike prices by expiration type. The typical number of expirations (nExp) and number of strike prices per expiration (nStr) are computed as respectively a mode and median. Since end-of-month expirations only appear at the very end of our sample, they were omitted.

	Traditional		Weekly		Quarterly		LEAP	
Year	nExp	nStr	nExp	nStr	nExp	nStr	nExp	nStr
2002	6	34	0	-	0	-	2	25
2003	6	33	0	-	0	-	2	25
2004	6	34	0	-	0	-	2	23
2005	6	36	0	5	0	-	2	25
2006	6	43	1	5	0	-	3	20
2007	6	44	1	7	4	12	3	23
2008	7	55	1	7	4	12	3	29
2009	7	78	1	15	4	28	3	48
2010	7	81	1	15	4	26	3	50
2011	7	85	1	45	4	31	3	51
2012	7	85	4	67	4	44	3	53
2013	7	92	4	85	4	51	3	63
2014	7	84	6	123	4	54	4	75

expirations as they come to maturity. At least two LEAP expirations maturing in two years or less are available on any given trading day. A third LEAP expiration with a maturity of at least two years was added in late 2005.

The Short-Term Option Series Program introduced *weekly* expirations on 28-Oct-2005. It was initially limited to a single concurrent expiration with five strike prices and a contract duration of one week (IC05-138). This program was later replaced by the End-of-Week Expirations Program on 02-Dec-2010 (IC10-174, RG10-112), which increased weekly contract duration. On 31-May-2012 (RG12-066) and 30-Jan-2014 (RG14-010), CBOE added respectively five and eight consecutive near-term expiration weeks.

For traditional expirations, the settlement value is computed from an underlying *opening* price. All other expirations, with the exception of weekly expirations before 02-Dec-2010, are settled at the close price. For options settling on market open, we adjust τ by subtracting 0.5, i.e. with $\tau \in \{0, 0.5, 1, 1.5, \ldots\}$.

Table 2.5 also presents the typical number of strike prices by expiration type. From

2002 to 2014, CBOE increased by a little less than three times the number of strike prices for traditional expirations. Corresponding changes for weeklys are even more dramatic, from five strike prices to over a hundred. CBOE further adds strike prices as a given expiration approaches maturity; e.g. we observe more than 150 strike prices for the front-month expiration after 2012 (not shown), nearly double the corresponding median over all traditional expirations[5].

The first two columns of Table 2.6 show typical moneyness ranges for front-month expirations. Upward jumps in maximum moneyness during the banking crisis are explained by a decrease in index levels. In contrast, CBOE lowered the minimum strike price from around 700 index points before the crisis to 200 by late 2008. To this day, CBOE maintains extremely low strike prices with moneyness as low as 5%. The last column shows the number of NTM options. Since 2005, all NTM options have a constant strike price interval of five index points, the smallest allowed by CBOE. Reported variations are hence driven by fluctuations in index levels.

Overall, the typical number of available expirations sustained a three-fold increase from 8 in 2002 to 25 in 2014. The quality of strike price listings has also improved over the years in terms of moneyness, with more NTM contracts and lower levels of moneyness, as a consequence of both the direct involvement of CBOE and index levels increasing.

[5] The front-month expiration is defined as the first traditional expiration coming to maturity.

Table 2.6: Strike price coverage for front-month expirations, i.e. the first traditional expiration coming to maturity. The typical available moneyness range (Min & Max) and the daily number of NTM options (N) are both computed as the median of daily data. NTM options are characterized by $k_0 \in [0.95, 1.05]$.

	Moneyness		NTM
Year	Min	Max	N
2002	0.67	1.49	12
2003	0.60	1.33	15
2004	0.46	1.30	21
2005	0.43	1.21	24
2006	0.56	1.16	26
2007	0.54	1.14	29
2008	0.52	1.46	26
2009	0.22	1.60	19
2010	0.19	1.38	23
2011	0.12	1.50	26
2012	0.11	1.33	28
2013	0.06	1.28	33
2014	0.27	1.24	39

2.3.2 Preliminary processing

As opposed to transaction data, quote data allow for synchronized cross-sectional observations at the cost of uncertainty in clearing prices. A small percentage of option transactions even occurs *outside* bid and ask prices. Mid-quotes acting as estimators of market-clearing prices are subject to cross-sectional heteroskedasticity due to variation in market making costs; see e.g. George and Longstaff (1993). Possibly worse, mid-quotes may be biased over moneyness-maturity areas with extremely low trading activity, leading to cross-sectional correlation. For example, deeply OTM options often have no bid price or the same ask price for multiple contracts of a given expiration.

The calibration of pricing surfaces (to come) —while robust to sparsity— is not robust to cross-sectional heteroskedasticity nor correlation. We hence calibrate pricing surfaces using exclusively mid-quotes deemed as sufficiently efficient estimators of clearing prices, a property which we loosely refer to as *liquidity*.

To do so, we filter the dataset according to how traders *perceive* liquidity in practice, namely by their ability to convert put-call quote pairs into a *single* B&S implicit volatility parameter (IV), defined as the σ parameter solving the B&S pricing equation for a given market price. Our goal is to attain a reasonable trade-off between biases of retained mid-quotes and moneyness-maturity sparsity induced by data filters.

We apply three filters sequentially:

I *Routine Filter;* We discard all expirations with $\tau \notin [1.5, 252.5)$ and reject quotes with a bid or ask price lower than or equal to zero and/or with a bid price greater than its ask price.

II *Bid-Ask IV Spread Filter;* First, we reject all options for which either the bid or ask price may not be converted to IV. Second, we discard options for which IV_0^b and/or IV_0^a do not fall into $[5\%, 100\%]$, where IV_0^b and IV_0^a are respectively the annualized IV computed from a bid and ask option price. Third, we reject quotes for which the *relative* bid-ask IV spread $(\mathrm{IV}_0^a - \mathrm{IV}_0^b)/\delta_0$ is greater than 2, where δ_0 is the 21 TDM NTM bid-ask IV spread[6].

III *Put-Call IV Difference Filter;* We first fit a piecewise quadratic polynomial on OTM IVs (from both put and call options) for each expiration; as described in Appendix B of Broadie et al. (2007). We then compute the historical median and MAD of *relative* calibration errors for different moneyness-maturity intervals. We use the intervals shown in Table 2.3. For each interval, we reject put and call options — *including* ITM options— which have a relative IV error falling outside the median plus or minus *five* MADs.

The preliminary maturity upper bound filter is motivated by the focus of empirical tests on one to three months-to-maturity options. The lower bound is motivated by issues

[6] We first compute the median of $(\mathrm{IV}_0^a - \mathrm{IV}_0^b)$ over NTM options for the front- and back-month expirations using both put and call options, denoted respectively by δ_{FM} and δ_{BM}. The back-month expiration is defined as the *second* traditional expiration coming to maturity. We then let $\delta_0 = (\tau_{\mathrm{BM}} - 21)/(\tau_{\mathrm{BM}} - \tau_{\mathrm{FM}})\delta_{\mathrm{FM}} + (21 - \tau_{\mathrm{FM}})/(\tau_{\mathrm{BM}} - \tau_{\mathrm{FM}})\delta_{\mathrm{BM}}$, where $\tau_{\mathrm{FM}}(\tau_{\mathrm{BM}})$ is the TDM for the front-(back-)month.

in computing IVs for options coming to maturity.

By relying on bid and ask IVs, the second filter implicitly censors all quotes for which either the bid and/or ask price do not respect usual arbitrage bounds; see e.g. Manaster and Koehler (1982). We then discard options with large bid-ask IV spreads $(\mathrm{IV}_0^a - \mathrm{IV}_0^b)$. This spread is approximated to the first order by $(c_0^a - c_0^b)/\mathrm{vega}$, where $c_0^a(c_0^b)$ is a normalized ask(bid) option price and vega is the partial derivative of B&S prices with respect to the volatility parameter; see e.g. Hentschel (2003).

Over moneyness, vega is bell-shaped, centered close to one, and more peaked for short-term maturities. As we move away-from-the-money and closer-to-maturity, the cumulative effect of larger observed bid-ask spreads and lower vega generates extreme IV uncertainty. Resulting rejection patterns are hence cone-shaped over moneyness-maturity, with tighter *retained* moneyness ranges for shorter-term maturities. For example, this is in line with the intuition that an OTM put option with $k_0 = 0.95$ is more liquid for a three months-to-maturity expiration than for a one week-to-maturity expiration, as the probability of ending ITM is almost zero for the latter.

Cross-sections of bid-ask IV spreads are normalized by δ_0 to avoid rejection trends coming from long-run liquidity levels. In fact, rejection rates slightly *decreased* during the last banking crisis due to volatility levels increasing and vega being less peaked, even though liquidity decreased and bid-ask spreads increased (not shown). This is consistent with far-away-from-the-money options containing more valuable IV information when prevailing volatility levels are high.

The third filter is intended to control for two types of faulty observations. First, large calibration errors are treated as outliers —e.g. due to data entry errors— and are rejected similarly to Constantinides et al. (2013). Put and call options with the same strike price and maturity theoretically share the same IV[7]. The ability to communicate the prices of both a put and a call option using a *single* IV is presumably an important factor of perceived liquidity. Second, we hence discard ITM call(put) options for which IVs sig-

[7]. Under B&S, put-call parity can be written as $(1-k) = (N(d_+^c) + N(-d_+^p)) - k(N(d_-^c) + N(-d_-^p))$, where values with superscript $p(c)$ are computed from the put(call) volatility parameter.

nificantly diverge from their OTM put(call) counterparts. This is implicitly achieved by quadratic polynomials being fitted to OTM options *only*, but being used to reject both OTM *and* ITM options.

Figure 2.3 shows a typical application on call options of proposed filters for two expirations. We readily see bid-ask IV spreads (delimited by light-gray areas) widening as we move away-from-the-money. This effect is much more aggressive for the short-term maturity in the left tail, resulting in systematic rejections for $k_0 < 0.96$ by the second filter; see the left side of the top panel. The third filter rejects five nonconsecutive quotes for the short-term maturity and a NTM quote for the mid-term maturity, due to sampling anomalies. The third filter further rejects three consecutive deeply ITM quotes for the mid-term maturity, due to biased mid-quotes being exposed by strong put and call IV differences.

Figure 2.3: Typical pattern for filter II and III in the IV space for call options (22-Apr-2005). Solid lines are the quadratic polynomials fitted on OTM options *only*. Light-gray areas represent IVs bid-ask spreads, i.e. with upper bound IV_0^a and lower bound IV_0^b. Markers indicate IV at mid-quotes.

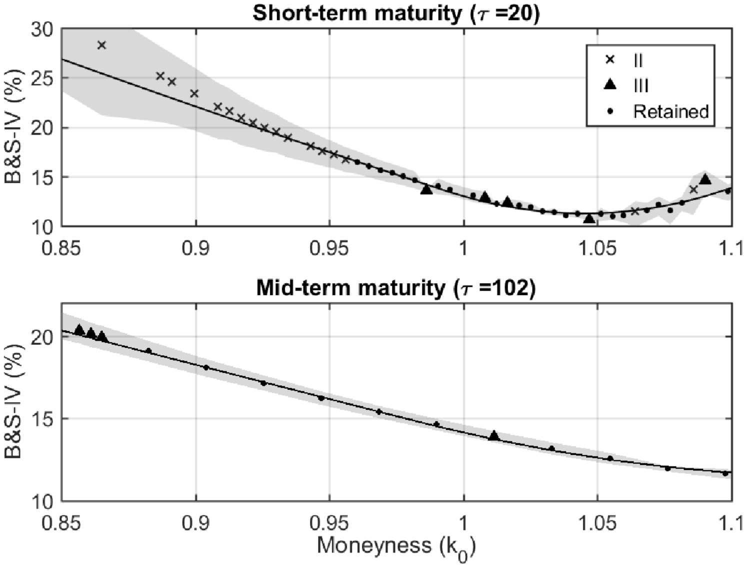


Table 2.7 shows resulting rejections throughout the period for each filter. The preliminary maturity filter censors roughly $1/6$ of all options, mostly LEAP contracts. The majority of erroneous quotes reported in the second row occurred for deeply OTM options due to missing bid prices.

The significant put-call difference for the second filter is explained by the higher number of contracts available in the left tail (i.e. $k_0 \ll 1$) and the IV transformation failing significantly more often for deeply ITM calls than for deeply OTM puts; see third row. This last effect is mostly due to lower arbitrage bound violations by bid prices.

The third filter rejects equal proportions of ITM and OTM options for shorter-term maturities, but significantly more ITM options than OTM options for longer-term maturities (not shown). This is consistent with the third filter capturing sampling anomalies that are equally likely over the moneyness-maturity plane, but also mid-quote biases that are particularly strong for long-term ITM options. Both issues are relatively small, as evidenced by a rejection rate of roughly 3% for both put and call options; see sixth row.

Table 2.7: Number of censored quotes (Censored) and corresponding rate of rejection (%) from sequentially applying filters (I,II and III) to put and call options. We start with 2,701,924 raw observations for both put and call options.

		Put Options		Call Options	
#	Filter	Censored	%	Censored	%
I	$\tau < 1.5$ or $\tau \geq 252.5$	455,428	16.1	455,428	16.1
	Erroneous or no bid	263,676	9.3	186,692	6.6
II	No IV_0^b or IV_0^a	360,353	12.8	705,633	25.0
	$\text{IV}_0^{a,b} \leq 5\%$ or $\geq 100\%$	2,856	0.1	5,681	0.2
	$(\text{IV}_0^a - \text{IV}_0^b)/\delta \geq 2$	427,442	15.1	441,218	15.6
III	5-MADs from OTM IV	91,485	3.2	82,880	2.9
Total		1,601,240	56.7	1,877,532	66.5

Table 2.8 displays cross-sectional rejection rates by subsample. We readily confirm anticipated cone-shaped rejection patterns and further note a slight asymmetry, i.e. with ITM options being rejected more often than OTM options. Rejection rates for OTM put options with $k_0 \in [0.925, 0.975)$ and more than 10 TDM are always below 7%.

As a final step before the calibration of surfaces, we mitigate induced sparsity by enforcing put-call parity and replacing rejected put mid-quotes by $c_0^{p\star} = c_0^c - (1 - k_0)$, where c_0^c is a retained call mid-quote. Data imputation is performed only if a put-call parity-based price $c_0^\star$ falls within initial bid and ask prices, i.e. $c_0^{p\star} \in [c_0^{pb}, c_0^{pa}]$ where $c_0^{pb}(c_0^{pa})$ are bid(ask) put prices. We proceed similarly for rejected call mid-quotes according to $c_0^{c\star} = c_0^p + (1 - k_0)$.

Over a total of 240,741(515,744) possible data imputations for put(call) options, only 4,628(6,387) were discarded due to put-call parity-based prices falling outside initial bid and ask prices. Such a high imputation rate suggests data filters capture biases in mid-quotes, as opposed to genuine arbitrage opportunities. For example, differences in put-call IVs captured by the third filter can not be arbitraged using put-call parity in practice.

Overall, the final number of processed mid-quotes is very close for put and call options, with a total retention of roughly 51.8% for both. The data imputation procedure hence succeeds at mitigating asymmetric rejection patterns; surface estimators for put *and* call options should behave similarly. While the final retention rate may seem low, the procedure retains more than 95% of *put* contracts for moneyness-maturity intervals considered in tests, namely $\tau \in [21, 63]$ and $k_0 \in [0.9, 1]$), in more than 90% of trading days.

When one is concerned with liquid option quotes, the proposed filtering methodology formalizes ad hoc rejection schemes classically found in the literature. First, *static* rectangular filters (e.g. as used by Dumas et al. (1998)) are often motivated by liquidity considerations, but fail to adapt to time variations in cross-sectional liquidity patterns. Second, our procedure acts as a data-driven alternative to the *systematic* imputation of all ITM options e.g. as performed by Ait-Sahalia and Lo (1998). Third, data imputation mitigates arbitrage-based rejections in the spirit of Bakshi et al. (1997). Such rejections are unwarranted when surface estimators already non-trivially correct for arbitrage opportunities present in the dataset e.g. as noted by Ait-Sahalia and Duarte (2003). Finally, we abstain from filtering negative option-implicit interest rates e.g. as performed by Constantinides et al. (2013). Corresponding options indeed contain valuable information once synchronization issues have been mitigated.

Table 2.8: Percentage rejection rates by TDM (τ), subsample, and moneyness (k_0). Panels for $\tau \geq 252.5$ (not shown) are systematically equal to 100% due to the routine filter (I).

		k_0				
TDM (τ)	Subsample	< 0.925	$[0.925, 0.975)$	$[0.975, 1.025)$	$[1.025, 1.075)$	≥ 1.075
				Put Options		
$[0.0, 10.0)$	2002 – 2004	88.0	39.7	38.1	97.8	100.0
	2005 – 2007	94.4	54.7	51.0	99.7	100.0
	2008 – 2011	82.2	34.6	46.8	96.6	99.9
	2012 – 2014	86.3	34.5	47.4	100.0	100.0
$[10.0, 31.0)$	2002 – 2004	56.0	3.1	4.5	57.8	96.9
	2005 – 2007	68.4	6.8	11.3	82.4	99.5
	2008 – 2011	62.8	4.0	4.9	48.6	95.7
	2012 – 2014	60.6	2.7	10.0	89.6	100.0
$[31.0, 63.0)$	2002 – 2004	39.9	2.4	3.7	11.3	86.6
	2005 – 2007	47.7	3.1	3.8	40.2	94.0
	2008 – 2011	44.9	2.5	3.0	12.4	82.9
	2012 – 2014	51.7	3.2	4.5	53.1	98.9
$[63.0, 252.5)$	2002 – 2004	13.9	2.2	3.3	2.9	66.0
	2005 – 2007	27.3	4.5	3.8	8.1	55.4
	2008 – 2011	31.3	5.8	4.0	4.4	65.8
	2012 – 2014	43.4	4.4	3.7	14.9	75.9
				Call Options		
$[0.0, 10.0)$	2002 – 2004	100.0	97.9	38.6	42.0	93.7
	2005 – 2007	100.0	99.4	51.4	58.9	96.5
	2008 – 2011	99.9	95.4	44.0	37.9	87.6
	2012 – 2014	100.0	99.5	46.0	49.2	95.2
$[10.0, 31.0)$	2002 – 2004	96.6	51.9	3.4	5.3	74.3
	2005 – 2007	97.9	68.8	9.7	20.7	75.8
	2008 – 2011	95.3	33.5	3.5	4.7	64.0
	2012 – 2014	99.9	67.9	8.3	16.8	75.9
$[31.0, 63.0)$	2002 – 2004	82.0	4.2	2.7	1.7	58.5
	2005 – 2007	89.4	13.8	3.5	8.3	58.7
	2008 – 2011	81.7	4.9	2.6	2.2	50.0
	2012 – 2014	96.7	20.7	3.8	7.2	61.3
$[63.0, 252.5)$	2002 – 2004	50.7	3.1	1.9	1.3	33.9
	2005 – 2007	58.9	3.3	3.4	5.1	21.2
	2008 – 2011	61.6	6.2	3.7	3.0	37.5
	2012 – 2014	83.8	7.9	3.2	3.9	29.4

2.3.3 Surface calibration

Bates (1991, 2000) fits *unidimensional* splines on moneyness slices during tests. Our purpose is to extend the empirical methodology to *bivariate* splines in order to further control for the maturity factor. Maturity extrapolation dramatically increases sample size. For example, one-month-to-maturity options are estimated on a daily basis, rather than observed once a month (for traditional expirations).

We do so under *arbitrage constraints* to ensure calibrated surfaces may reasonably be interpreted as simultaneous and rational market-clearing prices over the entire moneyness-maturity plane. More precisely, we estimate arbitrage-free bivariate tensor-product B-splines on processed mid-quotes. We follow Fengler and Hin (2015), who provide relevant asymptotic results and numerical simulations. We slightly adapt their estimator by fixing pricing surfaces to the payoff at $\tau = 0$[8]. The resulting estimator yields prices for options coming to maturity, required for the implementation of model-free delta-hedging strategies *until maturity*.

Regarding spline knot selection, we find daily guided knot searches generate highly volatile knot sets, ultimately leading to jagged time series. Consistently, Fengler and Hin (2015) find guided searches in the spirit of Zhou and Shen (2001) are unstable and often break down early. When delta-hedging, daily price instability leads to unwarranted portfolio revisions and excessive transaction costs. We thus use a single fixed set of moneyness and maturity knots, with knot density matching the expected convexity of the option pricing operator[9].

Finally, we retrieve surfaces under the original pricing scale,

$$\hat{C}_0(F_0^\tau, B_0^\tau, K_0, \tau) = \hat{c}_0(k_0, \tau) F_0^\tau / B_0^\tau,$$

[8] Following the notation of Fengler and Hin (2015), we let $v_0 = \ldots = v_{p_2} = 0$ and perform the quadratic optimization under the additional constraint $\theta_{j_1,0} = (\xi_{j_1}^* - 1, 0)^+$ for puts, and similarly for calls $\theta_{j_1,0} = (1 - \xi_{j_1}^*, 0)^+$, for $j_1 = \{0, \ldots, q_1\}$. To accommodate the discontinuity in first derivative, we consider moneyness knot sequences of the form $\underline{x} = \xi_0 = \ldots = \xi_{p_1} < \xi_{p_1+1} < \ldots < \xi_{p_1+k} = \ldots = \xi_{2p_1+k-1} < \ldots < \xi_{q_1} < \xi_{q_1+1} = \ldots = \xi_{q_1+p_1+1} = \bar{x}$ with $k > 0$, $\xi_{p_1+k} = 1$ and $q_1 > 2p_1$.

[9] We consider 20 moneyness knots distributed quadratically and concentrated at $k_0 = 1$. TDM knots are fixed at critical horizons, namely $\{0, 10, 21, 63, 126, 252\}$. Average of daily AICs is -15.08(-15.37) for put(call) options, only slightly worse than the optimal AIC of -16.72 on 1-Dec-2010 reported by Fengler and Hin (2015).

where $\hat{c}_0(k_0,\tau)$ is a calibrated surface, $K_0 = k_0 F_0^\tau$, and (F_0^τ, B_0^τ) is computed for all relevant TDMs by linearly interpolating or horizontally extrapolating annualized option-implicit interest rates, i.e. $\log(\hat{B}_0^\tau) \times 252/\tau$.

Overall, the data reduction procedure summarizes 2.7M scattered raw quote observations from 2002 to 2014 into roughly 0.8M spline coefficients which are neatly organized over a fixed moneyness-maturity grid. Figure 2.4 shows a typical pricing surface. The observed non-monotonicity along τ for deeply ITM put options is a consequence of the de-normalization procedure from $\hat{c}_0$ to $\hat{C}_0$; *normalized* $\hat{c}_0$ surfaces always respect static arbitrage constraints, including monotonicity along τ. Interpolated price time series (not shown) reflect genuine —not overly jagged nor biased— changes in option market conditions. For example, the correlation between one-month-to-maturity ATM $\hat{c}_0$ prices and the VIX index is 0.994.

Figure 2.4: Typical put pricing surface (22-Nov-2006). Markers represent available S&P 500 option mid-quotes constrained over $K_0 \in [1200, 1500]$ and $\tau \in [10, 207]$ for clarity. The official S&P 500 close value was 1406.09.

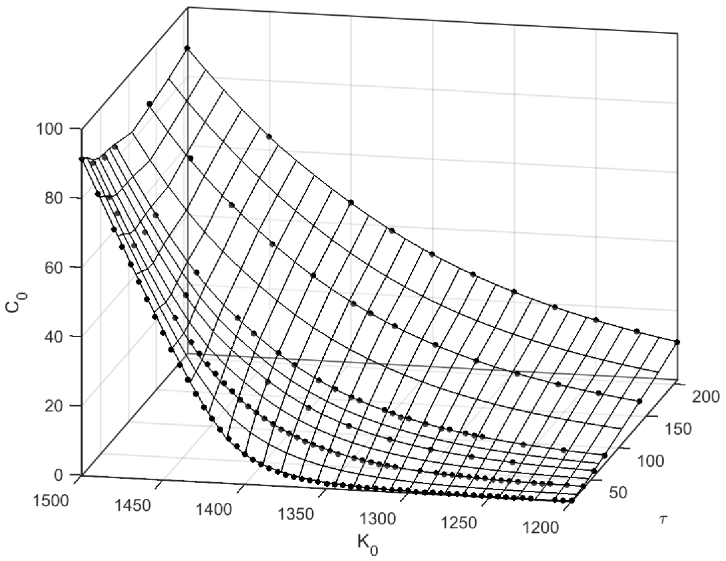


Table 2.9 presents medians and MADs of *bid-ask deviations* defined as $-1+2\frac{\hat{C}_0 - C_0^b}{C_0^a - C_0^b}$, where (C_0^b, C_0^a) are respectively bid and ask prices. A bid-ask deviation is -100%, 0% or

100% when $\hat{C}_0$ corresponds to respectively the bid, mid or ask price. Since corresponding statistics are computed over the raw dataset (as opposed to the processed dataset), deviations allow us to assess mid-quote biases in areas where minimum liquidity requirements were not met. We find mid-quotes for OTM options usually underestimate clearing prices, as evidenced by positive bid-ask deviations. This is consistent with corresponding quotes often having no bid (i.e. a bid price of zero), in which case transactions are more likely to occur closer to the ask price. Interestingly, mid-quotes for deeply OTM call options appear to *overestimate* clearing prices.

For put options, the mean bid-ask deviation is 1.63% and the standard deviation is 34.25% over moneyness-maturity intervals considered during tests (not shown). Calibrated put surfaces hence act as economically relevant representations of prices over intervals of interest. Regarding call options, surface fit quickly degrades as we approach $k_0 = 1.1$, as evidenced by MADs above 50%. In intervals of interest, roughly 21% of calibrated call prices fall outside bid-ask prices, as opposed to only 2% for put options. The corresponding mean bid-ask deviation for call options is 36.73% and the standard deviation is 127.97%. Overall, surface fits appear consistent with most common presumptions regarding S&P 500 index option liquidity, such as deeply OTM call options being less liquid than deeply OTM put options.

The semi-parametric approach of Fengler and Hin (2015) provides flexibility in capturing cross-sectional features. Strong shape restrictions from arbitrage constraints improve efficiency and mitigate over-fitting concerns often associated to nonparametric approaches. Competing maturity extrapolation approaches mostly apply the forward equation of Dupire (1994) under daily recalibrations of local variance functions. While semiparametric in nature, these approaches react badly to the presence of arbitrage opportunities in the dataset. For example, Kahalé (2004) presupposes cross-sections are already free of arbitrage opportunities. Non-trivial corrections proposed by Andreasen and Huge (2011) are more or less severe on short-term options depending on the choice of a forward or backward resolution algorithm. Bachem et al. (2013) propose a global correction scheme that overcomes the presence of arbitrage opportunities in a non-trivial fashion.

Other approaches interpolate IVs instead of prices, motivated by the fact IVs are typically less convex and have a more regular scale than prices, a feature first exploited by Shimko (1993). The relevant dataset is even provided by OptionMetrics. The IV transformation, however, introduces non-linearity in the formulation of arbitrage constraints. While Gatheral and Jacquier (2014) propose a valid arbitrage-free parameterization of IVs, arbitrage-free *and* semi-parametric IV surface estimators have yet to be proposed in the literature. To our knowledge, Fengler and Hin (2015) are the first and only to show efficiency gains from imposing arbitrage constraints to option prices in both moneyness *and* maturity in a semi-parametric setting.

Table 2.9: Medians of bid-ask deviations, i.e. $-1+2\frac{\hat{C}_0-C_0^b}{C_0^a-C_0^b}$, for calibrated S&P 500 option pricing surfaces by TDM (τ), moneyness (k_0) and subsample. Corresponding MADs are given in parentheses. All numbers are presented in percent.

		Moneyness (k_0)				
TDM(τ)	Subsample	< 0.925	$[0.925, 0.975)$	$[0.975, 1.025)$	$[1.025, 1.075)$	≥ 1.075
				Put Options (%)		
$[0.0, 10.0)$	2002 – 2004	34(69)	6(40)	8(28)	9(11)	-2(19)
	2005 – 2007	14(72)	1(55)	25(39)	5(9)	-1(11)
	2008 – 2011	21(75)	21(57)	8(38)	15(14)	2(11)
	2012 – 2014	46(75)	15(87)	21(53)	6(10)	-1(9)
$[10.0, 31.0)$	2002 – 2004	27(50)	0(16)	-1(11)	10(13)	3(21)
	2005 – 2007	25(45)	-1(16)	-2(18)	28(18)	8(17)
	2008 – 2011	16(35)	-1(10)	-1(10)	9(16)	6(14)
	2012 – 2014	19(45)	-1(32)	-3(38)	20(17)	1(11)
$[31.0, 63.0)$	2002 – 2004	4.7(22)	0(8)	1(6)	-2(10)	4(23)
	2005 – 2007	8(22)	0(8)	0(11)	20(27)	23(21)
	2008 – 2011	3(13)	1(6)	2(8)	-4(12)	10(17)
	2012 – 2014	5(22)	-5(15)	-2(22)	16(22)	7(17)
$[63.0, 252.5)$	2002 – 2004	1(7)	-0(3)	-0(3)	-0(4)	-1(22)
	2005 – 2007	3(10)	0(4)	0(6)	-1(10)	6(15)
	2008 – 2011	2(9)	0(5)	0(6)	0(7)	3(16)
	2012 – 2014	4(15)	0(9)	2(12)	-3(13)	-2(16)
				Call Options (%)		
$[0.0, 10.0)$	2002 – 2004	-0(20)	3(12)	6(29)	31(71)	-78(19)
	2005 – 2007	4(16)	7(10)	32(54)	23(100)	-92(7)
	2008 – 2011	5(10)	8(11)	6(40)	106(142)	-86(12)
	2012 – 2014	2(9)	4(10)	20(69)	11(108)	-93(6)
$[10.0, 31.0)$	2002 – 2004	3(21)	-1(6)	-1(9)	24(40)	-55(34)
	2005 – 2007	6(18)	0(6)	-2(14)	74(92)	-70(27)
	2008 – 2011	5(12)	-1(5)	-0(9)	15(39)	-56(41)
	2012 – 2014	2(11)	1(11)	-4(35)	89(112)	-78(21)
$[31.0, 63.0)$	2002 – 2004	0(19)	0(4)	0(5)	-1(14)	-17(37)
	2005 – 2007	2(16)	-0(4)	2(8)	24(54)	2(73)
	2008 – 2011	3(10)	1(5)	2(7)	-4(16)	-5(49)
	2012 – 2014	-1(13)	-2(9)	-1(22)	39(64)	-15(59)
$[63.0, 252.5)$	2002 – 2004	-1(11)	-0(3)	-0(2)	0(4)	-0(13)
	2005 – 2007	0(11)	0(3)	0(4)	-0(8)	2(15)
	2008 – 2011	2(10)	0(4)	-0(5)	1(8)	-2(16)
	2012 – 2014	-10(18)	0(7)	1(10)	-2(14)	0(21)

2.4 Robustness checks

2.4.1 Volatility timing in the option market

The single contract rule, namely $w_t = 1$, is devised for simplicity to shield empirical results from volatility timing in the *option* market. One could, however, argue that such timing still *implicitly* occurs, as agents are likely aware of the dependence of $\mathrm{E}_t[\pi_t]$ on market volatility. For example under the model of Bakshi and Kapadia (2003), $\mathrm{E}_t[\pi_t]$ is proportional to the vega of an option and to the volatility risk premium (VRP), both impacted by prevailing volatility. A similar effect is observed under the model of Gârleanu et al. (2009). $\mathrm{E}_t[\pi_t]$ as a function of market volatility is unfortunately highly model-dependent.

Our goal is to show robustness to *uninformed* decision rules in the option market, i.e. decisions taken under an information set shared by all option market participants. For example, the option-implicit vega under B&S is readily available, whereas VRP predictions require additional knowledge. The vega is further the main source of residual (i.e. un-hedgeable) risk for hedged option returns under many pricing frameworks. *All* agents may thus reasonably be expected to react to changes in risk, i.e. $\mathrm{Var}_t[\pi_t]$, using vega as the relevant ex-ante proxy.

For the proposed robustness check, we let agents hold a fixed amount of *vega*, as opposed to a fixed number of *contracts*. This approach only *partly* correct for time variations in premium, i.e. $\mathrm{E}_t[\pi_t]$, under the framework of Bakshi and Kapadia (2003). Uninformed agents indeed remain unaware of changes in VRP. We recall that w_t acts as a *constant* scaling parameter from t to $t+\tau$ for options sold at time t. In particular, agents still pre-commit to holding options until maturity.

For a given moneyness and maturity, we consider a *vega-constant* decision rule,

$$w_t = \left(\frac{\text{B\&S implicit vega at } t}{\text{B\&S long-run vega}}\right)^{-1} = \frac{\varphi(\bar{d}_+)}{\varphi(d_{t,+})},$$

where

$$d_{t,+} = \frac{-\log(k_0) + \mathrm{IV}_t^2\tau/2}{\mathrm{IV}_t\sqrt{\tau}},$$

with IV_t the σ parameter solving the B&S model for an option price given by our data reduction procedure on trading day t (for moneyness k_0 and τ TDM), and,

$$\bar{d}_+ = \frac{-\log(k_0) + \hat{\delta}\tau/2}{\sqrt{\hat{\delta}\tau}},$$

with $\hat{\delta}$ the long-run variance estimate presented in the main paper.

Figure 2.5 shows resulting weights for one-month-to-maturity put options. The vega of ATM options is not significantly impacted by market volatility, such that the vega-constant rule is indistinguishable from the single contract rule, i.e. $w_t \approx 1$. In other words, any deviation from the single contract rule for ATM options is *informed* and requires additional abilities e.g. with regards to VRP predictions.

For OTM options, we observe leveraged exposures (as high as 225%) during calm markets and conservative exposures (as low as 25%) during stressed markets. This is consistent with the intuition that agents must sell less OTM option contracts during market crises to reap the same profits as during bull markets. This effect gets stronger as we move further away-from-the-money.

Figure 2.5: Daily option weights by level of moneyness under the vega-constant decision rule for one-month-to-maturity put options.

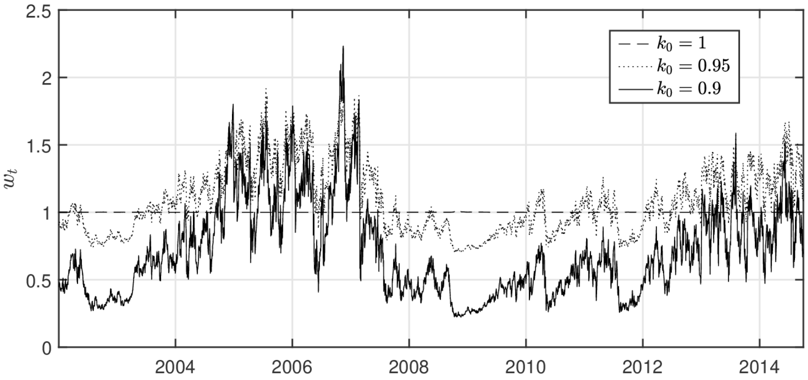
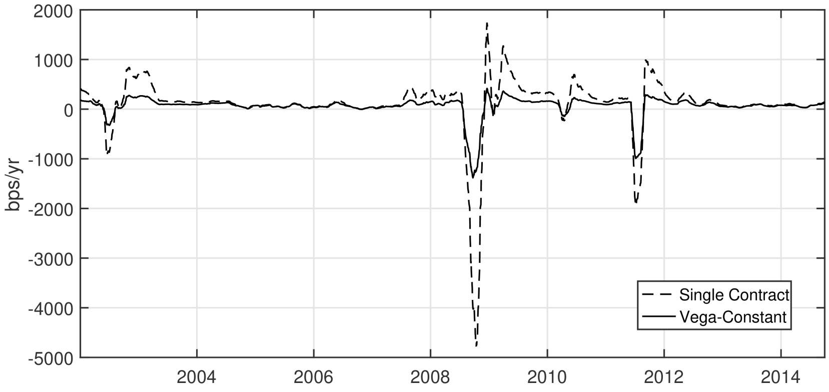


Figure 2.6 shows proposed weights succeed in stabilizing $\text{E}_t[w_t\pi_t]$ through time for deeply OTM options. The new decision rule is particularly effective for positive shocks in premium, e.g. with spikes in early 2003, early 2009, late 2010 and late 2011 being almost entirely smoothed. Drawdowns are greatly mitigated, but not fully eliminated. For

example, the minimum return during the last financial crisis goes from roughly -48% to roughly -12%.

Figure 2.6: Monthly rolling average of $w_t \pi_t$ under the static (S0) protocol for one-month-to-maturity deeply OTM options with $k_0 = 0.9$.



We calibrate risk aversion parameters in Θ_U using the same procedure as in the main paper. Resulting $\rho_{k_0,\tau}$ parameters (not shown) span a tighter range. In particular, the ex-post risk-reward trade-off under the static protocol —as captured by $\rho_{k_0,\tau}$— now appears almost as favorable for deeply OTM options as for ATM options. In other words, the additional risk management layer preemptively corrects for cross-sectional effects. The calibration procedure for ρ is thus not as essential here as in the main paper, but is still applied for consistency.

Figure 2.10 shows the resulting cross-section of economic values. We readily see the positiveness of incremental value from the LF class to the HF class is robust under all model requirements. As expected, ATM results are not materially impacted by the current exercise and were omitted for conciseness. In absolute, OTM options under the HF0 protocol still maximize economic value.

Deeply OTM options now generate significantly less economic value. In particular, all un-hedged protocols now yield *negative* economic values. Subsample results (not shown), however, show an increase in economic value for un-hedged protocols during 2005-2007 and 2012-2014 under the vega-constant rule.

Table 2.10: Cross-section of economic value Θ_U under the vega-constant decision rule.

		Trading-Days-to-Maturity (τ)				
		21	32	42	53	63
		Out-of-the-Money ($k_0 = 0.95$)				
	UH	-807.20	-2,227.68	-3,224.00	-4,632.63	-5,395.24
	DH	145.38	131.16	124.51	104.46	77.40
	S0	50.13	42.52	32.74	39.46	37.05
	LF0	156.10	145.54	135.93	110.22	81.49
LF	LFG	149.98	138.84	130.97	110.82	87.93
	LFA	139.40	132.33	128.52	118.80	109.60
	HF0	179.11	169.41	165.16	148.64	128.03
HF	HFG	168.33	164.56	165.63	154.82	142.56
	HFA	158.60	154.99	154.84	145.91	136.97
		Deeply Out-of-the-Money ($k_0 = 0.9$)				
	UH	-292.43	-241.16	-462.25	-1,166.42	-1,157.30
	DH	41.98	61.89	71.78	75.03	67.13
	S0	15.82	26.16	28.70	30.91	33.26
	LF0	57.23	83.70	92.58	90.89	85.45
LF	LFG	56.38	79.24	85.34	81.95	76.26
	LFA	51.22	67.17	73.18	75.80	73.29
	HF0	66.04	95.95	107.65	112.04	111.06
HF	HFG	62.73	87.44	97.75	103.76	103.73
	HFA	58.29	79.94	89.15	96.35	96.51

2.4.2 Peso problem and loss aversion

Similarly to most classic utility functions, the metric used in the main paper can be approximated by a quadratic utility function. This metric hence arguably leads to conclusions that are more or less consistent with a mean-variance framework, as e.g. used by Fleming et al. (2001, 2003). Θ_U could favor un-hedged (or under-hedged) protocols when adverse events are rare and corresponding return distributions are strongly *negatively skewed.* This effect is related to the Peso problem discussed by Bondarenko (2003) in a similar context. Our goal is to present ex-post results that are consistent with un-hedged protocols *never* adopted on an ex-ante basis. We do so by using a prospect-theory metric under loss aversion.

More precisely, we focus on reward-to-risk ratios derived from behavioral utility functions; see e.g. Holthausen (1981). Zakamouline (2014) proposes a generalization nesting most reward-to-risk performance measures, such as the Sortino or Omega ratio. We follow their empirical application and consider, for a given coefficient of loss aversion ($\lambda \geq 1$) and a *minimum acceptable return* (MAR),

$$\Theta_{\mathrm{PT}}(\pi;\lambda,\mathrm{MAR}) = \sqrt{\frac{252}{\tau}}\left(\frac{\mathrm{HOP}(\pi)-(\lambda-1)\mathrm{HS}_1(\pi)}{(\mathrm{HS}_2(\pi))^{1/2}}\right),$$

where

$$\mathrm{HOP}(\pi) = \frac{1}{T}\sum_{t=1}^{T} w_t\pi_t,$$

is related to the *hedged option premium* and,

$$\mathrm{HS}_\beta(\pi) = \frac{1}{T}\sum_{t=1}^{T}(\mathrm{MAR}-w_t\pi_t)^\beta \mathrm{I}(w_t\pi_t < \mathrm{MAR})$$

is related to *hedging shortfalls* below the MAR, with $\beta \in \{1,2\}$.

We focus on the vega-constant decision rule for w_t. As discussed in Section 2.4.1, this choice partly corrects for cross-sectional effects and already slightly penalizes un-hedged protocols. We assume MAR$= 0$ for simplicity. We ensure *un-hedged* agents are indifferent to extracting the hedged option premium by calibrating loss aversion parameters according to $\Theta_{\mathrm{PT}} = 0$,

$$\lambda_{k_0,\tau} = \frac{\mathrm{HOP}(c_0-c_\tau)}{\mathrm{HS}_1(c_0-c_\tau)}+1.$$

Figure 2.7 shows calibrated loss aversions are increasing in moneyness. During calm periods, the fact adverse events are less pronounced than anticipated is increasingly beneficial as we move away-from-the-money, e.g. with deeply OTM options *never* exercised. The calibration procedure hence effectively counteracts this effect by increasingly focusing on stressed periods as we move away-from-the-money. In contrast, we recall calibrated risk aversions $\rho_{k_0,\tau}$ *decrease* with k_0 under the utility-based metric. Θ_{PT} hence provides additional insights under drastically different attitudes towards risk than Θ_U, both in time due to loss aversion and over the moneyness-maturity plane due to our proposed calibration procedure.

Figure 2.7: Cross-sectional calibration of loss aversion parameters $\lambda_{k_0,\tau}$ for prospect theory-based performance measure Θ_{PT}.

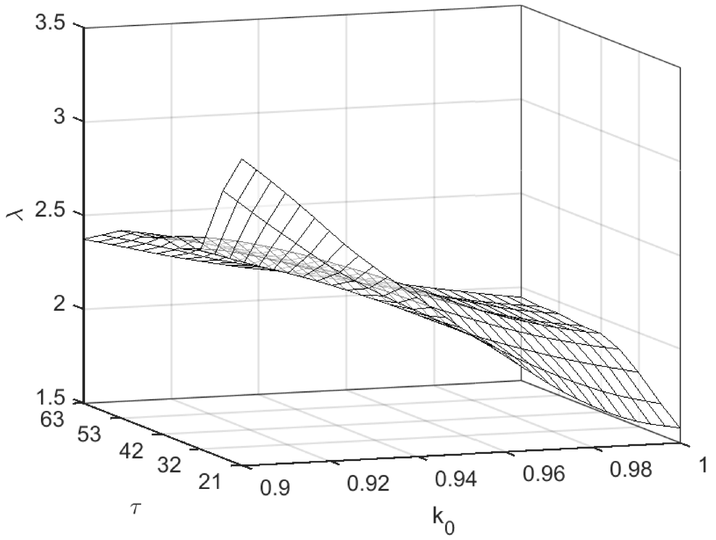


Table 2.11 presents cross-sectional results under loss aversion. The positiveness of incremental economic values is robust under all model requirements, confirming our main conclusion. In stark contrast to the utility metric, we observe a systematic increase in economic value with maturity, with the exception of LF0 for ATM options. This observation is explained by the negative skewness being more pronounced for shorter-term maturities, such that *loss averse* agents prefer to sell and hedge longer-term maturities.

Interestingly, specifications accounting for leverage effects now appear much more attractive relative to exponential-smoothing and mean-reverting models, at least for ATM and OTM options. This last observation is consistent with naive protocols being systematically under-hedged —a fact which eluded the utility metric for shorter-term maturities. In absolute, HFA yields the best performance with $\Theta_{\mathrm{PT}} = 1.33$ for three months-to-maturity OTM options, as compared to 1.17 for HFG.

The un-hedged protocol now *over*-performs LFG and LFA for short-term deeply OTM options. The proposed calibration thus falls short of mitigating the high appeal of un-hedged protocols for such contracts. As a further robustness check, we let agents relax their MARs, such that only the most extreme negative returns are perceived as losses.

Table 2.11: Cross-section of prospect theory-based metric Θ_{PT} for put options. Results for unhedged option returns are not shown as they are systematically equal to zero by construction.

		Trading-Days-to-Maturity (τ)				
		21	32	42	53	63
		At-the-Money ($k_0 = 1$)				
	DH	0.39	0.52	0.66	0.76	0.75
	S0	0.30	0.55	0.67	0.73	0.78
	LF0	0.59	0.67	0.67	0.64	0.58
LF	LFG	0.61	0.72	0.76	0.76	0.74
	LFA	0.65	0.85	0.99	1.03	1.01
	HF0	0.66	0.74	0.77	0.79	0.76
HF	HFG	0.68	0.92	1.13	1.26	1.26
	HFA	0.72	0.98	1.20	1.34	1.35
		Out-of-the-Money ($k_0 = 0.95$)				
	DH	0.14	0.32	0.32	0.40	0.48
	S0	-0.25	-0.02	-0.00	0.10	0.28
	LF0	0.39	0.54	0.54	0.55	0.60
LF	LFG	0.34	0.50	0.52	0.55	0.62
	LFA	0.46	0.68	0.72	0.81	0.90
	HF0	0.68	0.84	0.87	0.89	0.92
HF	HFG	0.67	0.91	1.03	1.13	1.17
	HFA	0.70	0.98	1.11	1.25	1.33
		Deeply Out-of-the-Money ($k_0 = 0.9$)				
	DH	-0.19	-0.07	0.01	0.15	0.12
	S0	-0.61	-0.43	-0.35	-0.22	-0.17
	LF0	0.06	0.24	0.30	0.39	0.37
LF	LFG	-0.00	0.14	0.17	0.26	0.24
	LFA	-0.05	0.21	0.30	0.43	0.42
	HF0	0.28	0.51	0.59	0.69	0.68
HF	HFG	0.21	0.38	0.44	0.60	0.61
	HFA	0.20	0.43	0.55	0.78	0.80

Table 2.12 presents corresponding results for MARs of -50, -100, -500 bps/yr. Lowering the MAR by only 50 bps/yr yields systematically positive values for volatility timing protocols. These metrics hence succeed in deterring loss-averse agents from selling unhedged options. All previous conclusions regarding relative and absolute economic values hold. In particular, deeply OTM options now reach their highest economic values under the HFA protocol, consistent with previous observations for ATM and OTM options. We

finally note that the over-performance of risk-minimizing versus delta-hedging protocols is robust to loss aversion in all cases.

Table 2.12: Prospect theory metric Θ_{PT} under different minimum acceptable returns (MARs) for deeply OTM put options. Loss aversion is calibrated as previously, such that results for un-hedged option returns are still systematically equal to zero by construction.

		Trading-Days-to-Maturity (τ)				
		21	32	42	53	63
		MAR of -50 bps/yr				
	DH	-0.14	-0.04	0.05	0.19	0.15
	S0	-0.57	-0.40	-0.32	-0.18	-0.13
	LF0	0.16	0.31	0.38	0.46	0.43
LF	LFG	0.10	0.21	0.24	0.33	0.30
	LFA	0.26	0.41	0.50	0.62	0.60
	HF0	0.43	0.61	0.70	0.78	0.77
HF	HFG	0.35	0.48	0.54	0.71	0.72
	HFA	0.38	0.57	0.70	0.94	0.96
		MAR of -100 bps/yr				
	DH	-0.09	-0.01	0.08	0.23	0.19
	S0	-0.54	-0.38	-0.29	-0.15	-0.10
	LF0	0.25	0.38	0.45	0.52	0.49
LF	LFG	0.19	0.27	0.31	0.39	0.36
	LFA	0.47	0.56	0.64	0.75	0.73
	HF0	0.56	0.71	0.80	0.88	0.86
HF	HFG	0.48	0.56	0.64	0.82	0.84
	HFA	0.53	0.69	0.83	1.08	1.11
		MAR of -500 bps/yr				
	DH	0.25	0.23	0.40	0.61	0.59
	S0	-0.31	-0.23	-0.09	0.13	0.21
	LF0	0.91	0.88	0.96	0.97	0.92
LF	LFG	0.88	0.78	0.82	0.84	0.80
	LFA	1.72	1.46	1.53	1.75	1.82
	HF0	1.54	1.48	1.59	1.67	1.69
HF	HFG	1.68	1.43	1.62	1.99	2.27
	HFA	1.82	1.67	1.99	2.61	3.15

2.4.3 Performance measure calibration

Table 2.13 shows a sensitivity analysis with respect to aversion parameters ρ and λ for one-month-to-maturity OTM put options. We readily see how proposed calibration procedures appropriately penalize the UH protocol, i.e. with Θ_U going from 310 to -807 and Θ_{PT} going from 0.82 to 0. Most importantly, conclusions regarding *incremental* economic values are robust to calibration procedures in all cases.

Table 2.13: Sensitivity analysis of performance measures for one-month-to-maturity OTM put options. Calibrated values (Cal.) correspond to $\rho_{0.95,21} = 23.90$ and $\lambda_{0.95,21} = 2.17$. We assume agents follow the vega-constant rule.

			Θ_U			Θ_{PT}	
		$\rho = 5$	$\rho = 10$	Cal.	$\lambda = 1$	$\lambda = 1.5$	Cal.
	UH	310.24	173.05	-807.20	0.82	0.47	0.00
	DH	198.95	186.16	145.38	1.21	0.75	0.14
	S0	128.97	111.57	50.13	0.72	0.31	-0.25
	LF0	190.40	181.99	156.10	1.46	1.00	0.39
LF	LFG	183.97	175.67	149.98	1.41	0.95	0.34
	LFA	163.83	157.63	139.40	1.62	1.13	0.46
	HF0	207.92	200.78	179.11	1.75	1.30	0.68
HF	HFG	195.21	188.49	168.33	1.73	1.28	0.67
	HFA	182.27	176.30	158.60	1.78	1.32	0.70

2.4.4 Out-of-sample exercise

Table 2.14 displays economic values Θ_U derived under *out-of-sample* protocols. More precisely, we re-estimate each model on a yearly basis —with data going back to 2000— and use hedging surface estimates for the following year *only*.

We find previous conclusions regarding the *incremental* value from the LF to the HF class are robust in all cases. We emphasize that DH results differ from the main paper, even tough DH is out-of-sample by construction. This is due to risk aversion now being calibrated to *out-of-sample* hedged option returns under S0. Gaps in performance between risk-minimizing and delta-hedging protocols unsurprisingly decrease when compared to in-sample results, especially for ATM and OTM options. Still, we find that the HF class over-performs DH under all model requirements.

Table 2.14: Cross-section of economic value Θ_U from selling and hedging put options under *out-of-sample* protocols. We assume agents follow the single contract decision rule.

		Trading-Days-to-Maturity (τ)				
		21	32	42	53	63
				At-the-Money ($k_0 = 1$)		
	UH	-5,321.88	-10,083.28	-10,846.10	-11,290.20	-10,101.95
	DH	115.07	83.43	75.21	65.43	34.67
	S0	47.11	-18.37	-89.48	-93.46	-110.75
	LF0	120.78	103.43	80.68	48.66	9.58
LF	LFG	123.01	106.92	89.10	65.92	37.91
	LFA	118.90	113.51	110.25	100.46	88.12
	HF0	129.79	110.20	91.00	69.87	41.43
HF	HFG	129.49	123.84	124.52	119.49	106.69
	HFA	127.94	125.52	126.39	121.79	111.07
				Out-of-the-Money ($k_0 = 0.95$)		
	UH	-70.55	-511.53	-1,136.79	-2,219.88	-3,564.66
	DH	163.67	158.92	157.34	137.54	111.46
	S0	47.60	43.20	35.91	37.67	32.74
	LF0	170.85	166.04	160.58	137.89	113.72
LF	LFG	163.08	156.45	149.91	129.78	107.98
	LFA	155.74	148.62	141.72	127.75	116.22
	HF0	191.99	186.26	183.17	165.51	143.83
HF	HFG	184.64	180.71	177.93	162.26	146.16
	HFA	172.76	167.61	163.61	150.43	138.70
				Deeply Out-of-the-Money ($k_0 = 0.9$)		
	UH	74.21	86.18	-77.28	-403.64	-381.04
	DH	101.51	104.08	108.42	108.76	97.81
	S0	39.22	37.17	35.11	32.91	34.51
	LF0	123.33	129.67	129.80	123.74	115.07
LF	LFG	119.32	122.33	120.03	111.10	101.13
	LFA	114.30	113.69	110.94	105.62	96.47
	HF0	134.55	144.67	146.23	143.42	136.96
HF	HFG	126.40	133.67	134.78	132.69	126.09
	HFA	118.54	122.51	122.14	121.16	114.56

2.5 Call option results

Table 2.15 shows descriptive statistics for hedged *call* returns. Our focus is on the single contract rule here. ATM call results are consistent with ATM put results. As we move away-from-the-money, however, the hedged call premium decreases. Lower return expectations are accompanied by lower risk measures; e.g. with a value-at-risk of roughly -10% for un-hedged ATM call options, as opposed to -18% for corresponding put options.

This last observation is consistent with the intuition that put options are vulnerable to violent market crashes, whereas negative inventories of call options under-perform during less volatile market rallies. Sharpe ratios for call options now increase slightly when allowing for a leverage effect. Overall, preliminary statistics suggest selling and hedging call options is less attractive than put options, significantly less so for deeply OTM options.

Table 2.16 shows incremental economic value from the LF class to the HF class dissipates as we move away-from-the-money and closer-to-maturity. Still, subsample and prospect theory evidences[10] suggest model misspecification leads to under-hedged call strategies, as similarly discussed for put options. Economic values derived from HFG and HFA also remain very close for longer-term maturities, in line with realized variance already capturing —at least part of— the leverage effect.

Overall, economic values extracted from put options are greater than call options, motivating our focus on put options in the main paper. We still find volatility timing benefits when hedging call options, even though call results are less definite regarding the incremental value of using realized variance-based forecasts.

[10] Results are available on request.

Table 2.15: Descriptive statistics for hedged call returns with 21 TDM. The first row(M) shows statistics for corresponding market returns. The mean(Mean), standard deviation(Std) and 95% value-at-risk(VaR) are presented as bps/yr. The Sharpe ratio(Shrp) is the first column (i.e. Mean) divided by the second column (Std). All statistics are computed from strictly non-overlapping subsamples which are then averaged, including one lag autocorrelation(Auto) and correlation with market returns(Corr).

		Mean	Std	Shrp	Skew	Kurt	Auto	VaR	Corr
	Market	560.44	1,679.71	0.33	-1.04	7.28	0.00	-2,182.68	1.00
				At-the-Money ($k_0 = 1$)					
	UH	42.61	862.03	0.05	-0.92	5.85	0.06	-1,039.57	-0.81
	DH	214.05	288.46	0.74	-1.46	9.47	0.13	-411.42	0.60
	S0	154.39	254.46	0.61	-2.41	18.76	0.14	-468.02	0.34
	LF0	157.60	212.94	0.74	-0.79	11.44	0.14	-291.48	0.34
LF	LFG	158.60	208.24	0.76	-1.03	11.18	0.14	-291.39	0.35
	LFA	145.34	191.17	0.76	-0.46	9.85	0.09	-236.46	0.11
	HF0	163.39	206.10	0.79	-0.99	10.80	0.14	-289.49	0.39
HF	HFG	154.45	191.44	0.81	-0.84	9.47	0.14	-248.70	0.32
	HFA	148.93	183.03	0.81	-0.58	8.93	0.10	-226.32	0.15
				Out-of-the-Money ($k_0 = 1.05$)					
	UH	211.22	401.36	0.53	-2.33	26.18	0.13	-758.42	-0.48
	DH	135.84	184.70	0.74	-0.57	15.15	0.19	-253.01	0.53
	S0	155.42	224.51	0.69	-1.00	11.38	0.04	-313.68	0.42
	LF0	134.31	174.17	0.77	-0.31	14.21	0.15	-252.63	0.52
LF	LFG	132.16	160.66	0.82	-0.68	14.36	0.16	-241.01	0.48
	LFA	120.75	139.37	0.87	-0.60	15.57	0.08	-190.25	0.16
	HF0	113.79	153.94	0.74	-1.10	16.11	0.17	-242.85	0.47
HF	HFG	115.12	130.94	0.88	-0.83	11.93	0.13	-175.49	0.33
	HFA	108.83	123.22	0.88	-1.04	13.29	0.06	-170.31	0.07
				Deeply Out-of-the-Money ($k_0 = 1.1$)					
	UH	82.01	203.29	0.40	-0.67	55.76	0.31	-388.72	-0.24
	DH	25.80	103.40	0.25	0.10	43.65	0.11	-175.50	0.35
	S0	26.33	124.36	0.21	-3.16	37.05	0.04	-277.23	0.12
	LF0	10.24	128.72	0.08	-0.37	33.12	0.08	-223.59	0.46
LF	LFG	14.00	111.14	0.13	-1.22	35.26	0.13	-205.68	0.40
	LFA	21.84	95.50	0.23	-1.29	39.98	0.05	-180.31	0.09
	HF0	9.33	106.34	0.09	-1.41	34.35	0.13	-201.53	0.42
HF	HFG	21.16	79.98	0.26	-1.10	33.40	0.05	-148.04	0.21
	HFA	24.18	78.77	0.31	-1.96	41.00	0.01	-165.91	-0.03

Table 2.16: Cross-section of economic value Θ_U from selling and hedging call options.

		Trading-Days-to-Maturity (τ)				
		21	32	42	53	63
		At-the-Money ($k_0 = 1$)				
	UH	-1,299.67	-2,730.87	-7,571.93	-4,580.92	-7,056.93
	DH	103.43	76.24	63.57	56.86	43.75
	S0	54.68	25.33	-6.35	8.85	20.02
	LF0	101.63	84.77	58.41	30.69	3.46
LF	LFG	104.59	92.52	75.52	59.77	43.93
	LFA	101.29	102.10	102.89	97.24	89.63
	HF0	110.83	96.47	79.35	66.35	50.97
HF	HFG	109.68	113.74	119.38	119.61	113.15
	HFA	108.57	113.34	118.38	117.93	111.81
		Out-of-the-Money ($k_0 = 1.05$)				
	UH	-524.98	-1,952.38	-6,724.84	-4,240.51	-6,900.10
	DH	80.16	86.62	73.40	49.86	47.54
	S0	48.65	45.86	60.22	60.58	60.70
	LF0	85.98	72.55	39.89	-1.81	-32.51
LF	LFG	90.68	85.63	61.90	38.40	28.78
	LFA	90.00	97.01	83.73	72.34	73.98
	HF0	75.15	68.94	42.87	14.46	2.18
HF	HFG	87.97	102.19	93.55	86.34	90.63
	HFA	84.66	98.66	89.86	82.21	86.01
		Deeply Out-of-the-Money ($k_0 = 1.1$)				
	UH	28.92	-389.51	-4,072.88	-2,271.64	-4,864.96
	DH	16.84	30.45	34.82	41.35	46.61
	S0	11.35	28.38	43.44	57.91	67.33
	LF0	-3.69	-6.04	-7.40	-12.37	-21.47
LF	LFG	3.50	9.06	14.94	21.27	22.91
	LFA	14.13	25.70	33.51	44.30	46.04
	HF0	-0.30	2.49	3.54	2.31	-2.01
HF	HFG	15.75	32.56	45.02	56.72	62.66
	HFA	18.80	32.61	42.16	51.70	55.29

References

Ait-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116(1-2):9–47.

Ait-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state price densities implicit in financial asset prices. *The Journal of Finance*, 53(2):499–547.

Alexander, C. and Nogueira, L. M. (2007). Model-free hedge ratios and scale-invariant models. *Journal of Banking and Finance*, 31(6):1839–1861.

Andreasen, J. and Huge, B. (2011). Volatility interpolation. *RISK*, (3):76–79.

Bachem, O., Drimus, G., and Farkas, W. (2013). Smooth and bid-offer compliant volatility surfaces under general dividend streams. *Quantitative Finance*, 13(11):1801–1812.

Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *The Journal of Finance*, 52(5):2003–2049.

Bakshi, G. and Kapadia, N. (2003). Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies*, 16(2):527–566.

Bates, D. S. (1991). The crash of '87: Was it expected? The evidence from options markets. *The Journal of Finance*, 46(3):1009–1044.

Bates, D. S. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics*, 94(1-2):181–238.

Bondarenko, O. (2003). Why are put options so expensive? Working Paper, University of Illinois at Chicago.

Branger, N. and Schlag, C. (2008). Can tests based on option hedging errors correctly identify volatility risk premia? *Journal of Financial and Quantitative Analysis*, 43(04):1055–1090.

Broadie, M., Cheov, M., and Johannes, M. (2007). Model specification and risk premia: Evidence from Futures Options. *The Journal of Finance*, 62(3):1453–1490.

Buraschi, A. and Jackwerth, J. (2001). The price of a smile: Hedging and spanning in option markets. *The Review of Financial Studies*, 14(2):495–527.

Carr, P. and Madan, D. B. (2005). A note on sufficient conditions for no arbitrage. *Finance Research Letters*, 2(3):125–130.

Constantinides, G. M., Jackwerth, J., and Savov, A. (2013). The puzzle of index option returns. *Review of Asset Pricing Studies*, 3(2):229–257.

Coval, J. D. and Shumway, T. (2001). Expected option returns. *The Journal of Finance*, 56(3):983–1009.

Driessen, J. and Maenhout, P. (2007). An empirical portfolio perspective on option pricing anomalies. *Review of Finance*, 11(4):561–603.

Dumas, B., Fleming, J., and Whaley, R. E. (1998). Implied volatility functions: Empirical Tests. *The Journal of Finance*, 53(6):2059–2106.

Dupire, B. (1994). Pricing with a smile. *RISK*, 7(1):18–20.

Fengler, M. R. and Hin, L. (2015). Semi-nonparametric estimation of the call-option price surface under strike and time-to-expiry no-arbitrage constraints. *Journal of Econometrics*, 184(2):242–261.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, 56(1):329–352.

Fleming, J., Kirby, C., and Ostdiek, B. (2003). The economic value of volatility timing using “realized” volatility. *Journal of Financial Economics*, 67(3):473–509.

Gârleanu, N., Pedersen, L. H., and Poteshman, A. M. (2009). Demand-based option pricing. *Review of Financial Studies*, 22(10):4259–4299.

Gatheral, J. and Jacquier, A. (2014). Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1):59–71.

Geman, H., Karoui, N. E., and Rochet, J. (1995). Changes of numéraire, changes of probability measure and option pricing. *Journal of Applied Probability*, 32(2):443–458.

George, T. J. and Longstaff, F. A. (1993). Bid-ask spreads and trading activity in the S&P 100 index options market. *Journal of Financial and Quantitative Analysis*, 28(3):381–397.

Heber, G., Lunde, A., Shephard, N., and Sheppard, K. (2009). Oxford-Man Institute's realized library. Oxford-Man Institute, University of Oxford, Version 0.2.

Hentschel, L. (2003). Errors in implied volatility estimation. *Journal of Financial and Quantitative analysis*, 38(04):779–810.

Holthausen, D. M. (1981). A risk-return model with risk and return measured as deviations from a target return. *The American Economic Review*, 71(1):182–188.

Jackwerth, J. C. and Rubinstein, M. (1996). Recovering probability distributions from option prices. *The Journal of Finance*, 51(5):1611–1631.

Kahalé, N. (2004). An arbitrage-free interpolation of volatilities. *RISK*, 17:102–106.

Manaster, S. and Koehler, G. (1982). The calculation of implied variances from the Black-Scholes model: A note. *The Journal of Finance*, 37(1):227–230.

Santa-Clara, P. and Saretto, A. (2009). Option strategies: Good deals and margin calls. *Journal of Financial Markets*, 12(3):391–417.

Shimko, D. C. (1993). Bounds of probability. *RISK*, (6):33–37.

Zakamouline, V. (2014). Portfolio performance evaluation with loss aversion. *Quantitative Finance*, 14(4):699–710.

Zhou, S. and Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *Journal of the American Statistical Association*, 96(453):247–259.

Chapter 3

On the Marginal and Recursive Quantization of GARCH Models

HUGO LAMARRE

Abstract

We investigate quantization methods when applied to GARCH models. We do so in the context of approximating price-variance dynamics by a discrete-state time-inhomogeneous Markov chain. Such an approximation is useful for solving various stochastic optimal control problems in finance. Both stochastic and deterministic methods are considered. In the former case, a standardized distortion function is minimized over $\mathbb{R}^2$ using a stochastic gradient descent. Standardization is critical due to an inherent gap in price-variance scale. In the deterministic case, we focus on Cartesian products of componentwise quantizations. This procedure —commonly known as *product* quantization— involves fast convex optimizations. Our numerical study shows deterministic methods are more reliable and efficient for large quantizers. We further propose a novel deterministic quantization of variances *conditional* on prices which better accommodates strong price-variance dependence effects. We successfully apply this conditional quantization to option pricing —both European and American— and variance-optimal hedging in discrete-time.

3.1 Introduction

GARCH models (Bollerslev, 1986) successfully capture time-varying risk in financial assets time series. They are characterized by conditional variances taking values in a *continuum* as a function of lagged variances and log-returns. While state continuity provides statistical flexibility, it may complicate some applications —notably when recursive conditional expectations need to be solved backwards in time. For example, special numerical methods have been proposed for pricing and/or hedging options[1]. We instead investigate a *general* approach to solving many financial stochastic optimal control problems —be it option pricing, risk management or portfolio allocation— using an approximation of GARCH dynamics obtained via quantization theory.

While quantization theory dates back to the late 1950s, it has only been recently applied to the field of numerical probability due to the seminal work of Pagès (1997) —allowing for many interesting (mostly financial) applications. The goal is typically a time-inhomogeneous Markov chain over discrete states for some predetermined time steps. In the daily GARCH model case for example, a quantization boils down to (1) discrete sets of conditional variances and asset prices (i.e. one for each future trading day) and (2) transition probabilities (i.e. from one future trading day to the next).

Under a given optimal quantization, many stochastic control problems may be solved as recursive weighted sums at very low computational costs. An optimal quantization can be reused for as many scenarios as needed e.g. when pricing and hedging a large portfolio of options with different strike prices or when optimizing portfolios for a large number of investors with different risk aversions. The quantization approach hence offers great scalability, but a fixed initial cost must be paid in order to obtain an optimal quantization.

We investigate different numerical optimization methods which may broadly be categorized as stochastic and deterministic. Our focus is on a *quadratic* criterion —referred to as the *distortion*— which offers nice properties under both categories.

For stochastic methods, we mostly rely on the competitive learning vector quanti-

[1] See e.g. Stentoft (2005), Ben-Ameur et al. (2009) or Rémillard and Rubenthaler (2013).

zation [CLVQ] algorithm over a two-dimensional space characterized by *log*-prices and conditional variances. The algorithm is essentially a stochastic gradient descent motivated by classical stochastic approximation convergence results[2]. Since conditional variances are typically several orders of magnitude smaller than log-prices, the two-dimensional Euclidean distance is dominated by log-prices. Standardizing variables is hence critical —a point which (to our knowledge) is not discussed in the literature.

We first apply the CLVQ algorithm to marginal daily laws. This approach provides best results when one is solely concerned with time t quantization errors. When solving financial problems, daily quantizations do not suffice as full dynamics (i.e. including transition probabilities) are required. We consider two approaches: (1) building Markov chains from marginal daily quantizations using Monte Carlo estimators for transition probabilities [Marginal] and (2) recursively applying the CLVQ algorithm to the time t law induced by the time $t-1$ optimal quantization (known by recursion) in the spirit of Pagès et al. (2004a) [Markov]. The latter approach is also known as *recursive* quantization.

For deterministic methods, we focus on componentwise and recursive quantizations. Analytical expressions are unavailable over two-dimensional spaces such that relaxed criteria over one-dimensional spaces must be considered. Resulting criteria are continuous and may be optimized using traditional tools (e.g. trust-region or BFGS algorithms) which are typically much faster than stochastic gradient descents. Deterministic methods hence potentially improve the trade-off between numerical burden and quantization quality.

We consider two recursive deterministic approaches: (1) separately optimizing one quantizer for log-prices and one for conditional variances then computing a Cartesian product in $\mathbb{R}^2$ as proposed by Fiorin et al. (2017) [Product] and (2) first optimizing one quantizer for log-prices and then optimizing one conditional variance quantizer per contemporaneous log-price quantizer element under conditional laws [Conditional]. While more demanding, the latter approach offers more flexibility and is embarrassingly parallel such that the increased burden may be mitigated by additional processing units. For both deterministic approaches, transition probabilities are computed analytically —a consider-

[2] See e.g. Kushner and Yin (2003).

able advantage over stochastic methods.

We implement quantization methods under the GARCH specification of Heston and Nandi (2000) for which the limit as the time interval shrinks is the stochastic volatility model of Heston (1993). Semi-analytical formulas may be solved via numerical integration for the pricing of European options, which allows us to demonstrate the quantization approach in a controlled environment. We further investigate applications to American option pricing and variance-optimal hedging in discrete-time. Quantization methods may easily be adapted to other popular one-lag GARCH specifications, be it a classical GARCH, a NGARCH (Engle and Ng, 1993) or a GJR-GARCH (Glosten et al., 1993).

In related works, Pagès and Sagna (2015) perform quantized option pricing under a local volatility model i.e. in a one-dimensional setting. Callegaro et al. (2016) and Fiorin et al. (2017) consider the Euler discretization of stochastic volatility models with latent variances linearly driven by two normal random variates. Our setting differs in that conditional variance is a quadratic function of a single common normal random variate and is bounded from below. One day ahead conditional laws considered under recursive GARCH quantization are hence strikingly different from their continuous-time limit counterparts surveyed in the literature. Practical applications further benefit from the parsimony of GARCH models since conditional variances are *observable* quantities — allowing for straightforward estimation of parameters typically by maximum likelihood.

This paper proceeds as follows. We first provide some background on quantization and its possible applications in Section 3.2. We then turn to specific concerns related to the quantizaton of a GARCH model in Section 3.3. We discuss stochastic and deterministic methods in respectively Sections 3.4 and 3.5. In Section 3.6, we contrast numerical results for the four quantization methodologies and provide some practical guidelines. Section 3.7 presents some applications. Proofs, explicit expressions and other implementation concerns are presented and/or discussed in the Appendix. Section 3.8 concludes.

3.2 Background & Motivation

Generally speaking, letting $x_t \in \mathbb{R}^n$ be a stochastic process, $\hat{x}_t(\Gamma_t) = \sum_{i=1}^{P_t} x_t^{(i)} \mathrm{I}(x_t \in C^{(i)}(\Gamma_t))$ is the quantization of x_t induced by the quantizer $\Gamma_t = \{x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(P_t)}\}$, where $C^{(i)}(\Gamma_t)$ is the so-called *Voronoi* tessellation satisfying

$$C^{(i)}(\Gamma_t) \subset \left\{ y \in \mathbb{R}^n \middle| |x_t^{(i)} - y| = \min_{1 \leq j \leq P_t} |y - x_t^{(j)}| \right\},$$

with $|\cdot|$ the Euclidean norm and $\mathrm{I}(\cdot)$ the indicator function. One is typically interested in a set of optimal quantizations $\{\hat{x}_t(\Gamma_t)\}_{t=1}^T$ minimizing *distortion* functions

$$D_{x_t}(\Gamma_t) = \|x_t - \hat{x}_t(\Gamma_t)\|_2^2 \tag{3.1}$$

e.g. for all time steps $t = 1, \ldots, T$. Quantization applications to finance typically focus on a Euler discretization $\tilde{x}_t$ of a *continuous*-time process x_t solving a given stochastic differential equation. The random variable $\hat{x}_t(\Gamma_t)$ taking values in a set of cardinality $P_t \in \mathbb{N}^+$ then acts as the best *discrete* approximation of $\tilde{x}_t$ in L^2 space for $t = 1, \ldots, T$. Given the critical role of the cardinality of a quantization, quantizers are often referred to as P_t-quantizer. We assume the cardinality to be constant in time (i.e. $P_t = P \in \mathbb{N}^+$) and henceforth drop the explicit dependence of $\hat{x}_t$ on Γ_t for notational convenience.

Theoretical properties of optimal quantizers have been extensively studied; see e.g. Pagès (1997) for a rigorous treatment. Under mild assumptions, the distortion function unsurprisingly goes to zero as $P \to \infty$. A more relevant result is the corresponding sharp asymptotic rate given by the so-called Zador theorem, namely $\lim_{P \to \infty} P^{n/2} D_{x_t} = G$ where G is a constant entirely specified by the law of x_t; see Graf and Luschgy (2000).

Our main motivation for finding an optimal quantization is related to stochastic control problems. When straightforward solutions are unavailable, one is often able to break down a multi-period financial problem and focus on more easily solved sub-problems, a method commonly known as *dynamic programming*.

For example when solving optimal stopping time problems, this approach involves recursively (i.e. for $t = T-1, \ldots, 1$) solving conditional expectations such as $v_t \big| \mathscr{F}_t =$

$\mathrm{E}\left[f(v_{t+1},x_{t+1})\middle|\mathscr{F}_t\right]$ where v_t is commonly known as the *value* function, $\mathbb{F} = \{\mathscr{F}_t, t = 0,\ldots,T\}$ is the usual filtration generated by x_t and both f and v_T (i.e. boundary conditions) are specified by the problem. For example when f and v_T are related to the reward from holding or exercising an American option, the value function is interpreted as an option price.

Such a recursion is unsolvable for GARCH models when the functional form of v over the *continuum* of conditional variances and asset prices is unknown[3]. One must then resort to approximating v_{t+1} using semi-parametric methods and solving conditional expectations v_t either analytically or through Monte Carlo.

When the market model is given by an optimal quantization $\{\hat{x}_t\}$, conditional expectations are solved according to

$$v_t^{(i)} = \mathrm{E}\left[f(v_{t+1},\hat{x}_{t+1})\middle|\hat{x}_t = x_t^{(i)}\right] = \sum_{j=1}^{P} \mathbb{P}(\hat{x}_{t+1} = x_{t+1}^{(j)}|\hat{x}_t = x_t^{(i)})f(v_{t+1}^{(j)},\hat{x}_{t+1}^{(j)}) \tag{3.2}$$

for $i = 1,\ldots,P$ and $t = T-1,\ldots,0$ where we assume a degenerated initial P-quantizer $\Gamma_0 = \{x_0,\ldots,x_0\}$ for convenience (since x_0 is $\mathscr{F}_0$-measurable by construction). We hence avoid semi-parametric approximations altogether.

When an optimal quantization is interpreted as an *approximation* of market dynamics however, Eq. (3.2) provides a locally constant estimator (denoted by $\hat{v}$) of the actual solution v over the price-variance continuum. Convergence of $\hat{v}$ towards v as $P \to \infty$ and other asymptotic behaviors would then rely on additional technical conditions on v and f. Under such an interpretation, optimal quantizers may be used to improve *ad hoc* grids (also called knots or meshes) commonly used in the related literature[4]. For example, Duan and Simonato (2001) argue their *ad hoc* "Markov chain reproduces the probabilistic behavior of the target GARCH process". While their statement is true asymptotically, Markov chains proposed here are *pre-asymptotically* optimal i.e. optimal for a given *finite* grid size. This pre-asymptotic optimality presumably leads to numerical benefits e.g. with a given level of precision achieved using smaller grids when pricing options.

[3] In some relatively rare cases, one may guess a functional form and find an analytical (recursive) solution; see Section 3.1 of Basak and Chabakauri (2010) for such a case under a Euler discretization.

[4] See Section 3.1. of Duan and Simonato (2001) or Section 3.6 of Ben-Ameur et al. (2009).

Furthermore, an optimal quantization is computed off-line and may be applied to multiple valuation functions and/or boundary conditions. The quantization approach is hence highly scalable and efficient once optimal quantizers (and transition probabilities) are obtained. For example, Glasserman (2003) notes “the effort might be justified if a [quantization], once constructed, could be applied to price many different American options”.

3.3 GARCH Quantization

The HN-GARCH specification of Heston and Nandi (2000) is,

$$\begin{aligned} r_t &= (\mu - 1/2)h_t + \sqrt{h_t} z_t, \\ h_{t+1} &= \omega + \beta h_t + \alpha \left(z_t - \gamma \sqrt{h_t} \right)^2, \end{aligned} \tag{3.3}$$

for $t = 1, \ldots, T$ indexing trading days, where r_t and h_t are respectively a log-return and a corresponding conditional variance, z_t is a standard normal random variable and $\{h_1, \mu, \omega, \alpha, \beta, \gamma\}$ are model parameters. We consider the *forward log-price* of some financial asset

$$s_t = \sum_{s=1}^{t} r_s$$

for $t = 0, \ldots, T$ with $\sum_{s=1}^{0} r_s = 0$ by convention. One may recover observable *spot* prices according to $S_t = S_0 \exp(s_t + (r - \delta)t)$ where S_0 is an initial asset price, r is a risk-free rate and δ is a dividend yield[5]. This setting allows us to henceforth abstract from risk-free rates and dividends.

We investigate the quantization of HN-GARCH models using two sets of parameters. First, we consider an *asymmetric* case under which parameters are representative of estimates on S&P 500 index (forward) log-returns from 1996 to 2014. Second, we consider a *symmetric* case under which γ is forced to 0 and β is set to match persistence estimates for the S&P 500 index, namely $\beta + \alpha\gamma^2 \approx 0.962$. In both cases, ω and μ are set such that

[5] This setting essentially follows from a change of numéraire under deterministic and constant interest and dividend rates. More precisely, $\exp(s_t)$ may be viewed as a total-return index (i.e. including dividends) given by $S_t \exp(-\delta(T-t))$ relative to the price of a zero-coupon bound with a notional value of $S_0 \exp^{(r-\delta)T}$.

long-run daily log-return and volatility expectations are respectively 5% and 18% per annum —once again plausible values for a forward contract on the S&P 500 index[6]. Table 3.1 shows parameter values used throughout. The symmetric HN-GARCH model is *not* a classical GARCH model. Shocks in conditional variance are indeed driven by shocks in model innovation *only*, as opposed to $z_t h_t$ under the classical GARCH model.

Table 3.1: Parameters of HN-GARCH model (3.3) in the asymmetric and symmetric case.

Parameter	Asymmetric	Symmetric
μ	2.04	2.04
ω	3.28e−7	3.28e−7
α	4.5e−6	4.5e−06
β	0.8	0.962
γ	190	0

Figure 3.1 shows daily standard deviations of log-prices and conditional variances for the next 126 trading days in the symmetric and asymmetric cases. Corresponding analytical expressions are derived in Appendix A.1. We readily observe conditional variances are several order of magnitude smaller than log-prices. The scale of log-price increases monotonically through time, whereas conditional variance converges towards its limiting unconditional distribution such that its scale stabilizes in time.

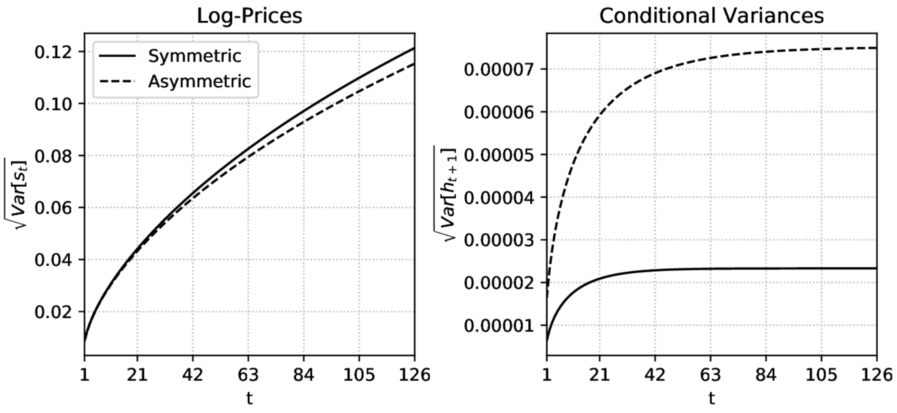
While our ultimate goal is an optimal quantization for the Markovian process $\{s_t, h_{t+1}\}_{t=1}^{T}$, the Euclidean distance is a flawed measure of distortion as log-price errors would dominate variance errors —more so for distant time steps. We hence instead consider

$$x_t = \left(\frac{s_t}{\sqrt{\mathrm{Var}\left[s_t\right]}}, \frac{h_{t+1}}{\sqrt{\mathrm{Var}[h_{t+1}]}} \right) \tag{3.4}$$

and minimize distortions (3.1) for *standardized* vectors x_t. More precisely, our preliminary goal is a set of optimal *standardized* quantizers $\Gamma_t = \{x_t^{(1)}, \ldots, x_t^{(P)}\}$ for $t = 1, \ldots, T$ with corresponding transition probabilities $p_t^{(i,j)} = \mathbb{P}(\hat{x}_t = x_t^{(j)} \big| \hat{x}_{t-1} = x_{t-1}^{(i)})$ and unconditional probabilities $p_t^{(j)} = \mathbb{P}(\hat{x}_t = x_t^{(j)})$ for $i, j = 1, \ldots, P$. Since h_1 is assumed

[6] For the HN-GARCH model, the long-run log-return variance is given by $\bar{\sigma} = (\omega + \alpha)/(1 - \beta - \alpha\gamma^2)$ and the long-run log-return expectation is given by $(\mu - 1/2)\bar{\sigma}$. Daily variances h are converted to annualized volatility percentages according to $\sqrt{252h}$ throughout.

Figure 3.1: Term structures of unconditional standard deviations for future daily log-prices and conditional variances until roughly 6 months.



known e.g. estimated from historical log-returns, we define an initial P-quantizer $\Gamma_0 = \{(0,h_1),\ldots,(0,h_1)\}$ with $p_0^{(1)} = 1$ (and $p_0^{(i)} = 0$ otherwise) and $p_1^{(i,j)} = p_1^{(1,j)}$ for $i,j = 1,\ldots,P$ for notational convenience.

For practical applications, we simply revert standardized vectors according to

$$\left(s_t^{(i)}, h_{t+1}^{(i)}\right) = \left(x_{1,t}^{(i)}\sqrt{\mathrm{Var}\left[s_t\right]}, x_{2,t}^{(i)}\sqrt{\mathrm{Var}[h_{t+1}]}\right) \tag{3.5}$$

where $x_{1,t}^{(i)}$ and $x_{2,t}^{(i)}$ are respectively the first and second components of $x_t^{(i)}$ for $t = 1,\ldots,T$ and $i = 1,\ldots,P$. We then consider the quantization

$$\begin{pmatrix} \hat{s}_t \\ \hat{h}_{t+1} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{P} s_t^{(i)} \mathrm{I}\left(x_t \in C^{(i)}(\Gamma_t)\right) \\ \sum_{i=1}^{P} h_{t+1}^{(i)} \mathrm{I}\left(x_t \in C^{(i)}(\Gamma_t)\right) \end{pmatrix}$$

for $t = 0,\ldots,T$ with $s_0^{(i)} = 0$ and $h_1^{(i)} = h_1$. Unconditional and transition probabilities are not impacted by the standardization.

For convenience, we introduce $\mathscr{S}$ and $\mathscr{H}$ such that $s_t = \mathscr{S}(s_{t-1}, h_t, z_t)$, and $h_{t+1} = \mathscr{H}(h_t, z_t)$. For example in the HN-GARCH case,

$$\begin{aligned} \mathscr{S}(s,h,z) &= s + (\mu - 1/2)h + \sqrt{h}z \\ \mathscr{H}(h,z) &= \omega + \beta h + \alpha(z - \gamma\sqrt{h})^2. \end{aligned}$$

Our focus is on two *marginal metrics*, the daily root-mean-squared [RMS] marginal quantization error,

$$\mathrm{RMS}_t = \sqrt{\frac{1}{S}\sum_{s=1}^{S} \min_{1\leq j\leq P} |\xi_{t,s} - x_t^{(j)}|^2}$$

and the total root-mean-squared [TRMS] marginal quantization error

$$\mathrm{TRMS}_T = \sqrt{\frac{1}{S}\sum_{t=1}^{T}\sum_{s=1}^{S} \min_{1\leq j\leq P} |\xi_{t,s} - x_t^{(j)}|^2}$$

where $\{\xi_{t,s}\}$ represents a typical (unconditional) GARCH simulation standardized according to (3.4) with each path indexed by s and $S = 1,000,000$. The simulation procedure is formalized in Algorithm B.5 found in Appendix A.2. We assume a starting conditional volatility of $\sqrt{252h_1} = 14\%$ throughout.

3.4 Stochastic Methods

In Section 3.4.1, we introduce three *sub*-algorithms which are specified up to a random variate sequence denoted by $\{\xi_s\}_{s=1}^{S}$ where S is a given number of simulations. Specific quantization methods differ in their choice of random variate sequence and in their estimator of transition probabilities, with Marginal and Markovian methods respectively given in Sections 3.4.2 and 3.4.3.

3.4.1 Sub-Algorithms

Unless otherwise specified, $\Gamma_0 = \{x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(P)}\}$ is an initial random quantizer i.e. a random variate sequence of size P. Here subscripts no longer represent trading days but rather iterations of a given algorithm. In particular, sub-algorithms are later used on a daily basis when quantizing full dynamics.

We first introduce competitive learning in Sub-Algorithm 1. The CLVQ algorithm relies on the distortion function (3.1) being sufficiently smooth for the convergence of a stochastic gradient descent towards a *local* minimum characterized by $\nabla D_{x_t} = 0$; see e.g. Kushner and Yin (2003) for convergence results. The choice of a *quadratic* criterion ensures optimal quantizers lie inside the convex hull of the support. Since conditional variances are bounded from below by (at least) zero, this inherently prevents quantizer elements from having negative conditional variances.

Sub-Algorithm 1. *Competitive Learning Vector Quantization*
Draw a random variates sequence $\{\xi_s\}_{s=1}^{S}$ with

$$S = \left\lceil \frac{4P^{3/2}}{\delta^\star \pi^2} \right\rceil \tag{3.6}$$

for a desired precision $\delta^\star$. Assuming δ_0 a given initial step parameter, let $s := 0$

1. *(Competitive Phase)*

$$i_{s+1} = \arg\min_{1 \le j \le P} |\xi_{s+1} - x_s^{(j)}|;$$

2. *(Learning Phase)*

$$\begin{cases} x_{s+1}^{(i_{s+1})} = x_s^{(i_{s+1})} - \delta_s (x_s^{(i_{s+1})} - \xi_{s+1}) \\ x_{s+1}^{(j)} = x_s^{(j)}, \quad j \neq i_{s+1} \end{cases} \tag{3.7}$$

   where

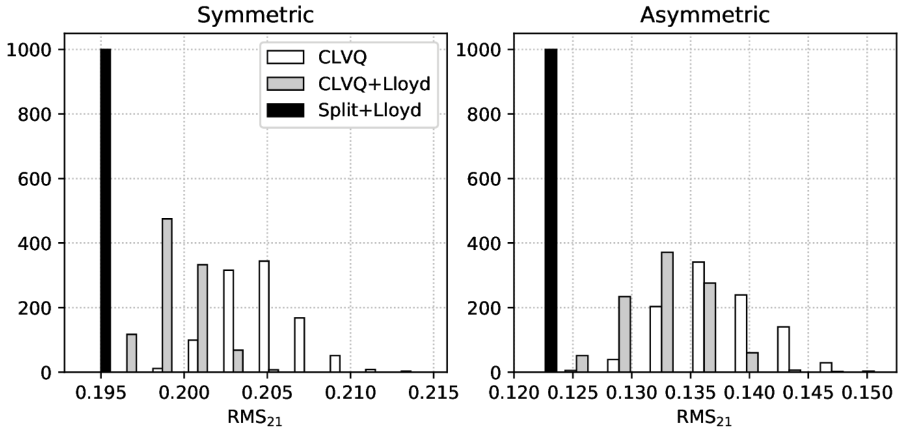
$$\delta_s = \left(\frac{1}{\delta_0} + \frac{\pi^2 s}{4P^{3/2}} \right)^{-1};$$

let $s := s+1$ and go back to step 1 or stop if $s \geq S$.

The step parameter (also called learning rate) δ_s is a heuristic proposed by Pagès and Printems (2003). This sequence ensures convergence for the univariate uniform law and is arguably relevant near the mode of *any* distribution i.e. where a distribution is locally flat and mimics the uniform distribution. The number of simulations S is specified as a function of P to ensure the last step reaches a given threshold i.e. $\delta_S \leq \delta^\star$. This is consistent with larger quantizers requiring more simulations in order to meet some minimum level of precision and later allows us to compare optimal quantizers of different sizes. Under the standardization introduced in Section 3.3, the problem is well scaled and it generally makes sense to let $\delta_0 = 1$ (unless otherwise specified).

For practical degrees of precision (say $\delta^\star = 1\text{e}{-2}$), we observe significant variability in quantizers obtained from consecutive CLVQ runs with different random variate sequences. White bars in Figure 3.2 [CLVQ] display the dispersion in distortion obtained from 1,000 CLVQ runs. RMS are highly variable with a total range of roughly 2-3%. We consider two additional sub-algorithms which improve the stability of optimal quantizers, namely the randomized Lloyd's method I (or simply *Lloyd's method*) and the splitting method —respectively given by Sub-Algorithms 2 and 3.

Figure 3.2: Histogram of RMS at 21 trading days for 1,000 stochastic optimization runs with $P = 100$ and $\delta_S^\star = 1\text{e}{-2}$. Optimal quantizers are obtained using unconditional random variates according to Algorithm B.5. White bars [CLVQ] are obtained using the CLVQ algorithm only (Sub-Algorithm 1). Grey bars [CLVQ+Lloyd] are obtained using the CLVQ algorithm followed by 10 iterations of Lloyd's method (Sub-Algorithm 2). Black bars [Split+Lloyd] are obtained using the splitting method (Sub-Algorithm 3) followed by 10 iterations of Lloyd's method.



Optimal quantizers are stationary i.e. $\text{E}[x|\hat{x}] = \hat{x}$ under a quadratic criterion, as a direct consequence of $\nabla D_{x_t} = 0$. Lloyd's method as proposed by Lloyd (1982) relies on $x^{(i)} \to \text{E}[x|\hat{x} = x^{(i)}]$ admitting a *unique* fixed point when x follows a univariate log-concave law. In the multivariate case, optimal quantizers are stationary, but *multiple* stationary quantizers could and in fact do exist in a GARCH setting.

Still, we empirically observe that Lloyd's method converges (at least locally) to a stationary quantizer[7]. The method is hence useful for refining a quantizer already near

[7] Formal convergence results related to Lloyd's method in a multivariate setting have yet to be established.

a local minimum (i.e. near stationary) such as one obtained via CLVQ. Since analytical expressions for $\mathrm{E}[x|\hat{x}]$ are unavailable here, we consider a randomized adaptation under which conditional expectations are estimated via Monte Carlo as discussed by Pagès et al. (2004b). For consistency, we use a Monte Carlo sample size given by (3.6).

Sub-Algorithm 2. *Randomized Lloyd's Method I*

Let U be a fixed number of iterations and $u := 0$,

1. *Draw a sequence of random variates $\{\xi_s\}_{s=1}^{S}$ and compute for $s = 1, \ldots, S$*

$$i_s = \arg\min_{1 \leq j \leq P} |\xi_s - x_u^{(j)}|;$$

2. *For $i = 1, \ldots, P$, let*

$$x_{u+1}^{(i)} = \frac{x_u^{(i)} + \sum_{s=1}^{S} I(i_s = i)\xi_s}{1 + \sum_{s=1}^{S} I(i_s = i)};$$

let $u := u + 1$ and go back to step 1 or stop if $u \geq U$.

Figure 3.2 shows that performing as few as 10 iterations of Lloyd's method following a CLVQ run [CLVQ+Lloyd] leads to a significant decrease in RMS. But optimal quantizers remain quite variable. Preliminary observations (not shown) suggest little benefit to considering more than 10 to 20 iterations as subsequent variations in quantizers are quickly dominated by (conditional expectation) sampling errors —especially in the tails. This is consistent with observations from Pagès and Printems (2003) in the pure Gaussian case. We let $U = 10$ throughout.

Pagès (1997) further suggests initializing the optimization of a P-quantizer using an optimal $P-1$-quantizer. This method —referred to as the splitting method— is formalized in Sub-Algorithm 3. The updating method for the initial step parameter δ_0 allows stochastic gradient descents to adapt to decreasing scales as additional knots are added[8].

[8] Following Pagès and Printems (2003), this choice is motivated by the inequality $1/2 \min_{i \neq j} |x^{(j)} - x^{(i)}| \leq (D_x)^{1/2}$. When the bound is close to being tight, we protect progress made in past iterations and otherwise fall back to $\delta_0 = 1$. Since only a rough estimate of D_x is needed, step 2 of the splitting method is typically performed as a companion procedure to step 1 for numerical efficiency.

Sub-Algorithm 3. *Splitting Method*

Let $\delta^\star$ be a desired precision, $\delta_0 := 1$ and $p := 1$,

1. *Set Γ_p to an optimal p-quantizer obtained via CLVQ (see Sub-Algorithm 1) with precision $\delta^\star$, δ_0 as the initial step parameter and $\Gamma_{p-1} \cup \{x_0^{(p)}\}$ as the initial quantizer if $p > 1$ or $\{x_0^{(1)}\}$ otherwise.*

2. *Using the sequence of random variates $\{\xi_s\}_{s=1}^S$ used in step 2, let*

$$\delta_0 := \min\left(\left(\frac{1}{S}\sum_{s=1}^{S} \min_{1\le j\le p} |\xi_s - x_p^{(j)}|^2\right)^{1/2}, 1\right)$$

   where $\Gamma_p = \{x_p^{(1)}, \ldots, x_p^{(p)}\}$;

let $p := p+1$ and go back to step 1 or stop if $p > P$.

Benefits from the splitting method are readily apparent in Figure 3.2. The splitting method (followed by 10 iterations of Lloyd's method) [Split+Lloyd] consistently yields by far the lowest distortion —even lower than the best runs of the other two methodologies. Figure B.1 in the Appendix shows that most of the remaining variability comes from sampling errors in RMS metrics. In other words under a precision of $\delta^\star = 1\text{e}{-2}$, quantization improvements become difficult to detect using a sample size of 1,000,000[9]. Since such a sample size is reasonable for most practical applications, we let $\delta^\star = 1e-2$ throughout. We are now ready to specify marginal and Markovian quantization methods.

3.4.2 Marginal Quantization

Algorithm 1 introduces the marginal quantization method as first proposed by Bally et al. (2001)[10]. We find an optimal quantizer for each time t marginal law —hence the name— and build a time-inhomogeneous Markov chain using Monte Carlo estimators for transition probabilities. Even though formulated sequentially, optimizations performed in the

[9] Preliminary observations (not shown) suggest this conclusion is robust to quantizer size (here $P =$ 100).

[10] See also Bally et al. (2005).

first step can be performed in parallel. For consistency, transition probabilities are estimated using a sample size given by (3.6).

Algorithm 1. *Marginal Quantization*

Draw unconditional random variates $\{\xi_{t,s}\}_{s=1}^{S}$ *according to Algorithm B.5. For* $t = 1,\ldots,T$,

1. *Set* $\Gamma_t = \{x_t^{(1)},\ldots,x_t^{(P)}\}$ *to an optimal P-quantizer obtained via a splitting method (Sub-Algorithm 3) followed by Lloyd's method (Sub-Algorithm 2) using unconditional random variates;*

2. *Compute transition probabilities for* $i,j = 1,\ldots,P$ *according to*

$$p_t^{(i,j)} = \frac{\sum_{s=1}^{S} I(i_s = i, j_s = j)}{\sum_{s=1}^{S} I(i_s = i)} \tag{3.8}$$

   where

$$i_s = \underset{1\le k\le P}{\arg\min}\,|\xi_{t-1,s} - x_{t-1}^{(k)}|, \quad j_s = \underset{1\le k\le P}{\arg\min}\,|\xi_{t,s} - x_t^{(k)}|$$

   with Γ_{t-1} *known by recursion when* $t > 1$ *or* $p_1^{(i,j)} = (1/S)\sum_{s=1}^{S} I(j_s = j)$ *otherwise;*

3. *Compute unconditional probabilities according to* $p_t^{(j)} = \sum_{i=1}^{P} p_{t-1}^{(i)} p_t^{(i,j)}$ *for* $j = 1,\ldots,P$ *with* $p_0^{(1)} = 1$ *and* 0 *otherwise.*

Figure 3.3 displays an optimal marginal quantizations at 21 trading days for both symmetric and asymmetric cases. When displaying quantizations, we first de-standardize according to (3.5) and then convert log-prices s and conditional variances h according to respectively $100\exp(s)$ and $100\sqrt{252h}$. Each colored patch represents a Voronoi tile $C^{(i)}$ associated to a given quantizer element $x^{(i)}$ displayed as a black dot. Color intensities reflect unconditional probabilities.

While Voronoi tiles are delimited by straight segments, parabolic boundaries observed in Figure 3.3 come from the conversion of variances into volatilities. The negative correlation between variances and log-prices is readily apparent in the asymmetric case. We may also visualize the lower support bound for conditional variances, taking a parabolic smile and smirk shape in respectively the symmetric and asymmetric case. Under *non-standardized* variables (i.e. (s_t, h_{t+1}) as opposed to x_t), distances would be dominated by log-prices and Voronoi tiles would be delimited by nearly vertical segments.

Figure 3.3: Optimal marginal quantization (Algorithm 1) at 21 trading days with $P = 100$ and $\delta^\star = 1\mathrm{e}{-6}$.

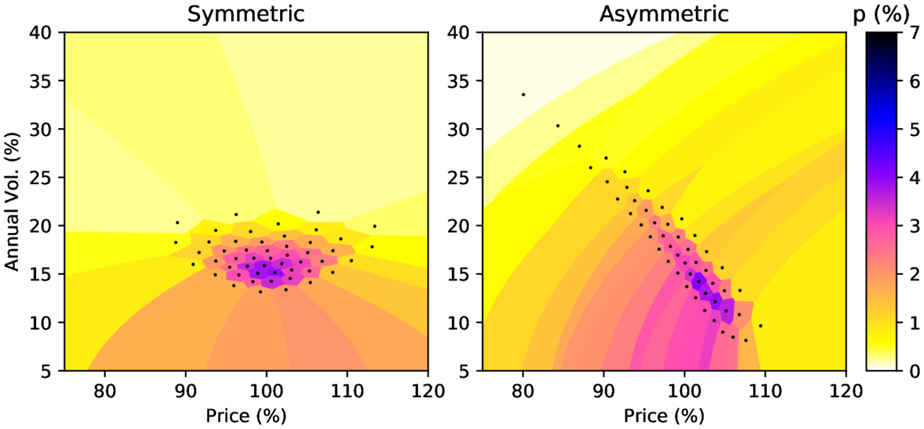
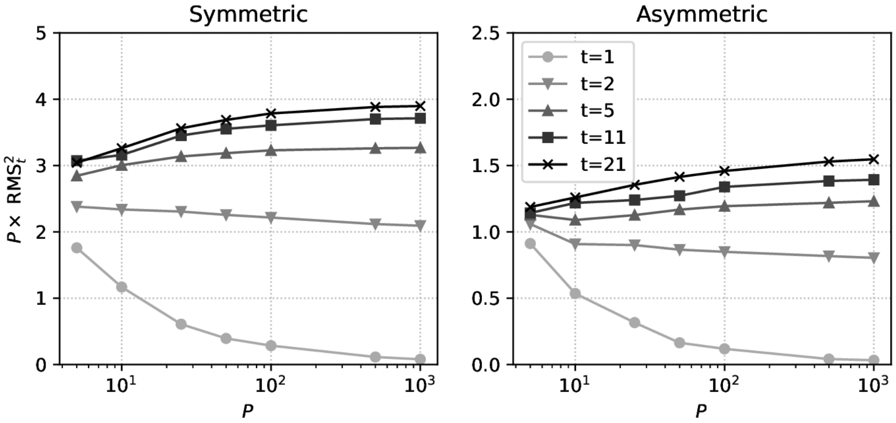


Figure 3.4 displays the empirical counterpart of Zador's limiting constant in two dimensions, namely $P(\mathrm{RMS}_t)^2$. We readily confirm Zador's prediction. We further observe a significant increase in the constant G through *time* with the first time step undoubtedly being the easiest for quantization. Figure B.2 in the Appendix shows RMS for each future trading day for a large quantizer size i.e. with Zador's rate nearly attained. We see daily empirical errors quickly increase in the first few days but then stabilize —and even plateau in the symmetric case.

Strong performance observed in the first few days appear to be related to the fact GARCH models are driven by a single innovation. The support of $(s_t, h_{t+1})|(s_{t-1}, h_t)$ depicts a parabola in price-variance space parameterized by z_t. Since (s_0, h_1) is assumed known, the support of (s_1, h_2) is a line as opposed to a surface for later time steps. Supports indeed converge towards the first quadrant of $\mathbb{R}^2$ as $t \to \infty$ i.e. towards the continuous-time limit of Heston (1993). This explains wide early gaps in RMS e.g. between $t = 1$ and $t = 2$.

Figure 3.4 suggests the limiting constant G is higher in the symmetric case. This is consistent with Pagès and Printems (2003) finding independent variables are the most dif-

Figure 3.4: Empirical estimates for Zador's constant under marginal quantization (Algorithm 1) as a function of quantizer size (P) for different trading days ($t = 1, 2, 5, 11, 21$).



ficult for quantization in the Gaussian case. The asymmetric i.e. strongly dependent case is increasingly difficult for more distant trading day, as evidenced by the significant positive slope for $t \approx 21$ in the right panel of Figure B.2. As a likely explanation, the leverage effect in the asymmetric case generates strongly skewed and fat-tailed distributions which are presumably more difficult for quantization. These distribution features are readily apparent by comparing the top-left corners in Figure 3.3 where worst (quantized) market losses are given by respectively -12% and -20% in the symmetric and asymmetric cases.

Strictly speaking, the set of quantizations $\{\hat{x}_t\}$ in Algorithm 1 is not a Markov chain. Marginal quantization indeed imposes an *ad hoc* Markovian structure by relying on quantized transition probability estimators (3.8). Concerns related to the optimality of such an *ad hoc* Markovian structure are typically understated in practice as a Markov chain can always be built in this way and does converge to GARCH dynamics as $P \to \infty$[11]. We next present the *recursive* approach of Pagès et al. (2004a) which targets Markovian laws and is later used under deterministic methodologies.

[11] See Theorem 5.1 of Pagès and Pham (2005) for convergence results.

3.4.3 Markovian Quantization

Pagès et al. (2004a) propose to recursively optimize distortions under time t laws induced by time $t-1$ optimal quantizations. Non-standardized random variates targeted by Markovian quantization (also referred to as *recursive* quantization) in Algorithm 2 are more precisely given at time t by

$$\left(\mathscr{S}(\hat{s}_{t-1},\hat{h}_t,z_t),\mathscr{H}(\hat{h}_t,z_t)\right)$$

where $(\hat{s}_{t-1},\hat{h}_t)$ is an optimal (time $t-1$) quantization known by recursion and z_t is still a standard normal random variable. We henceforth simply refer to the resulting law as the *Markovian* law. It is a mixture of P conditional GARCH laws, which as previously noted are defined over parabolas in price-variance space. Markovian quantization yields true Markov chains that converge to GARCH dynamics as $P \to \infty$[12], but is suboptimal from a marginal standpoint.

Algorithm 2. *Markovian (Recursive) Quantization*
For $t = 1,\ldots,T$, assuming the optimal time $t-1$ quantizer $\Gamma_{t-1} = \{x_{t-1}^{(1)},\ldots,x_{t-1}^{(P)}\}$ and corresponding unconditional probabilities $p_{t-1}^{(i)} = \mathbb{P}(\hat{x}_{t-1} = x_{t-1}^{(i)})$ are known at $t = 1$ and otherwise known by recursion

1. *Draw a sequence of standard normal random variates $\{z_s\}_{s=1}^S$;*
2. *Draw a sequence of integer random variates $\{i_s\}_{s=1}^S$ with probability $\mathbb{P}(i_s = i) = p_{t-1}^{(i)}$ for $i \in \{1,\ldots,P\}$ and 0 otherwise;*
3. *Compute standardized Markovian random vectors for $s = 1,\ldots,S$*

$$\xi_s = \left(\frac{\mathscr{S}(s_{t-1}^{(i_s)},h_t^{(i_s)},z_s)}{\sqrt{Var[s_t]}},\frac{\mathscr{H}(h_t^{(i_s)},z_s)}{\sqrt{Var[h_{t+1}]}}\right);$$

where $Var[h_{t+1}]$ and $Var[s_t]$ are predetermined constants respectively given by (**??**) *and* (**??**)*;*

[12] See Footnote 11.

4. *Set $\Gamma_t = \{x_t^{(1)}, \ldots, x_t^{(P)}\}$ to an optimal P-quantizer obtained via a splitting method (see Sub-Algorithm 3) followed by Lloyd's method (see Sub-Algorithm 2) using standardized Markovian random variate sequences $\{\xi_s\}_{s=1}^S$;*

5. *Compute transition probabilities*

$$p_t^{(i,j)} = \frac{\sum_{s=1}^S I(i_s = i, j_s = j)}{\sum_{s=1}^S I(i_s = i)}$$

for $i, j = 1, \ldots, P$, where $j_s = \underset{1 \leq k \leq P}{\arg\min} |\xi_s - x_t^{(k)}|$ and i_s was previously computed at step 2 (with $p_1^{(i,j)} = p_1^{(1,j)}$ by convention);

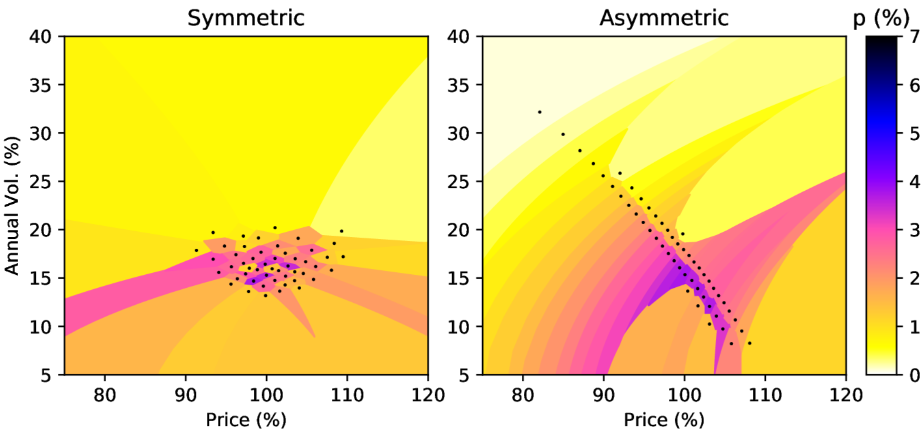
6. *Compute unconditional probabilities according to $p_t^{(j)} = \sum_{i=1}^P p_{t-1}^{(i)} p_t^{(i,j)}$ for $j = 1, \ldots, P$.*

Regarding numerical efficiency of Algorithm 2, some authors such as Bally et al. (2005) suggest harvesting probability estimators as companion parameters in the CLVQ algorithm. We find the computational impact to be negligible. Once optimal quantizers are computed, we may indeed rely on fast KD-trees to find nearest neighbors when estimating probabilities; see Bentley (1975). In contrast to marginal quantization, the recursive formulation of the Markovian algorithm prohibits parallel computations.

Markovian laws are fundamentally different from their *unconditional* counterparts considered previously under marginal quantization. The Markovian support is indeed given by a union of P parabolas as opposed to the first quadrant of $\mathbb{R}^2$ in the limit under unconditional laws. We expect such discrepancies to be exacerbated by the inherent accumulation of quantization errors under the Markovian approach i.e. as errors in time $t-1$ quantizers recursively impact time t Markovian laws. This distinguishes our work from existing stochastic volatility quantization approaches such as Callegaro et al. (2016) under which unconditional and Markovian laws are structurally similar i.e. defined over the same support. For example here we expect pre-asymptotic Markovian quantizer elements *not* to be in the targeted support, but still converge towards it.

Figure 3.5 shows an optimal Markovian quantization at 21 trading days. The geometry in the asymmetric case is quite different than its marginal counterpart i.e. with

Figure 3.5: Optimal Markovian quantization (Algorithm 2) at 21 trading days with $P = 100$ and $\delta^{\star} = 1\text{e}{-6}$.



optimal quantizers forming multiple diagonal segments as opposed to being much more diffused previously. Quantizers generally appear noisier on any given day and display erratic variations in time (not shown). This noise is *not* related to the stochastic nature of the algorithm which did in fact converge (here $\delta^{\star} = 1\text{e}{-6}$ corresponding to roughly 140,000,000 simulations in the last iteration of the splitting method). These preliminary observations instead suggest distortion functions under Markovian laws have many attracting local minima and are particularly difficult to optimize.

Deterministic approaches (introduced next in Section 3.5) must be solved in a Markovian setting since time $t > 1$ marginal laws quickly become too convoluted for analytical solutions. Markovian quantization is hence a relevant benchmark for deterministic approaches to come, whereas marginal quantization provides best results under marginal metrics (by construction).

3.5 Deterministic Methods

Nearest neighbors are highly tractable in one dimension with Voronoi tiles simply delimited by mid-points of adjacent quantizer elements. Gradients for log-price and conditional variance distortions may hence be derived under the recursive quantization approach. We do so towards implementing numerically efficient optimization routines. We first adapt product quantizations of Fiorin et al. (2017) to GARCH dynamics in Section 3.5.1. This approach intuitively wastes quantizer elements over effectively null probability areas — more so in the asymmetric case e.g. in the top-right and bottom-left areas in the right panel of Figure 3.3. We thus propose a novel deterministic alternative which targets multiple *conditional* laws in Section 3.5.2. Scaling issues related to the Euclidean norm in price-variance space (discussed in Section 3.3) do not apply here since quantizers to be optimized are one-dimensional.

3.5.1 Product Quantization

For $t = 1, \dots, T$, we preliminarily define $\Gamma_{s_t} = \{s_t^{(1)}, \dots, s_t^{(K)}\}$ and $\Gamma_{h_{t+1}} = \{h_{t+1}^{(1)}, \dots, h_{t+1}^{(N)}\}$ for some price and variance quantizer sizes denoted by respectively K and N. For notational convenience, we further assume initial quantizers $\Gamma_{s_0} = \{0, \dots, 0\}$ (containing K elements) and $\Gamma_{h_{t+1}} = \{h_1, \dots, h_1\}$ (containing N elements). Without loss of generality, we assume all time $t > 0$ quantizers are ordered, i.e. with $s_t^{(i+1)} > s_t^{(i)}$ for $i = 1, \dots, K-1$ and $h_{t+1}^{(l+1)} > h_{t+1}^{(l)}$ for $l = 1, \dots, N-1$.

The Cartesian product quantizer $\Gamma_t = \Gamma_{s_t} \times \Gamma_{h_{t+1}}$ is a typical rectangular grid over $\mathbb{R}^2$ containing KN elements. In particular, we let $KN \approx P$ when later comparing deterministic and stochastic methodologies. Regarding the GARCH quantization, we drop optimal Voronoi tessellations in favor of much more tractable *product* tessellations,

$$\begin{pmatrix} \hat{s}_t \\ \hat{h}_{t+1} \end{pmatrix} = \begin{pmatrix} \sum\limits_{i=1}^{K} s_t^{(i)} \mathrm{I}(s_t \in C_{s_t}^{(i)}) \\ \sum\limits_{l=1}^{N} h_{t+1}^{(l)} \mathrm{I}(h_{t+1} \in C_{h_{t+1}}^{(l)}) \end{pmatrix}$$

where *one-dimensional* Voronoi tessellations for $t > 0$ are simply

$$C_{s_t}^{(i)} = \left(\frac{(s_t^{(i-1)} + s_t^{(i)})}{2}, \frac{(s_t^{(i)} + s_t^{(i+1)})}{2} \right), \quad C_{h_{t+1}}^{(l)} = \left(\frac{h_{t+1}^{(l-1)} + h_{t+1}^{(l)})}{2}, \frac{h_{t+1}^{(l)} + h_{t+1}^{(l+1)})}{2} \right)$$

for $i = 1, \dots, K$ and $l = 1, \dots, N$ with $s^{(0)} = -\infty$ and $s^{(K+1)} = \infty$ and $h_t^{(0)} = 0$ and $h_t^{(N+1)} = \infty$ by convention. We do so towards *analytically* solving transition probabilities $p_t^{(il,jm)} = \mathbb{P}(\hat{s}_t = s_t^{(j)}, \hat{h}_{t+1} = h_{t+1}^{(m)} | \hat{s}_{t-1} = s_{t-1}^{(i)}, \hat{h}_t = h_t^{(l)})$ and unconditional probabilities $p_{t-1}^{(il)} =$

$\mathbb{P}(\hat{s}_{t-1} = s_{t-1}^{(i)}, \hat{h}_t = h_t^{(l)})$ where by convention $p_0^{(11)} = 1$ (and 0 otherwise). Algorithm 3 presents the product quantization optimization method. Henceforth, $d\varphi(z) = e^{-z^2/2}/\sqrt{2\pi}dz$ is the standard normal probability density function [PDF] and $\Phi(z)$ is the associated cumulative density function [CDF].

Algorithm 3. *Product Quantization.*

For $t = 1, \ldots, T$, assuming the optimal time $t-1$ Cartesian product quantizer $\Gamma_{t-1} = \{(s_{t-1}^{(i)}, h_t^{(l)}), \quad i = 1, \ldots, K, \quad l = 1, \ldots, N\}$ and corresponding unconditional probabilities $p_{t-1}^{(il)}$ are known at $t = 1$ and otherwise known by recursion,

1. *Find a log-price quantizer $\Gamma_{s_t} = \{s_t^{(1)}, \ldots, s_t^{(K)}\}$ minimizing the distortion for $\tilde{s}_t = \mathscr{S}(\hat{s}_{t-1}, \hat{h}_t, z_t)$,*

$$\begin{aligned} D_{s_t} &= E\left[\sum_{i=1}^{K} I(\tilde{s}_t \in C_{s_t}^{(i)})(\tilde{s}_t - s_t^{(i)})^2\right] \\ &= \sum_{i=1}^{K}\sum_{l=1}^{N} p_{t-1}^{(il)} \int_{-\infty}^{\infty} \sum_{j=1}^{K} I(\mathscr{S}(s_{t-1}^{(i)}, h_t^{(l)}, z) \in C_{s_t}^{(j)})(\mathscr{S}(s_{t-1}^{(i)}, h_t^{(l)}, z) - s_t^{(j)})^2 d\varphi(z); \end{aligned} \tag{3.9}$$

2. *Find a conditional variance quantizer $\Gamma_{h_{t+1}} = \{h_{t+1}^{(1)}, \ldots, h_{t+1}^{(N)}\}$ minimizing the distortion for $\tilde{h}_{t+1} = \mathscr{H}(\hat{h}_t, z_t)$,*

$$\begin{aligned} D_{h_{t+1}} &= E\left[\sum_{l=1}^{N} I(\tilde{h}_{t+1} \in C_{h_{t+1}}^{(l)})(\tilde{h}_{t+1} - h_{t+1}^{(l)})^2\right] \\ &= \sum_{i=1}^{K}\sum_{l=1}^{N} p_{t-1}^{(il)} \int_{-\infty}^{\infty} \sum_{m=1}^{N} I(\mathscr{H}(h_t^{(l)}, z) \in C_{h_{t+1}}^{(m)})(\mathscr{H}(h_t^{(l)}, z) - h_{t+1}^{(m)})^2 d\varphi(z); \end{aligned}$$

3. *Compute transition probabilities according to*

$$\begin{aligned} p_t^{(il,jm)} &= \mathbb{P}(\tilde{s}_t \in C_{s_t}^{(j)}, \tilde{h}_{t+1} \in C_{h_{t+1}}^{(m)} | \hat{s}_{t-1} = s_{t-1}^{(i)}, \hat{h}_t = h_t^{(l)}) \\ &= \int_{-\infty}^{\infty} I(\mathscr{S}(s_{t-1}^{(i)}, h_t^{(l)}, z) \in C_{s_t}^{(j)}) I(\mathscr{H}(h_t^{(l)}, z) \in C_{h_{t+1}}^{(m)}) d\varphi(z); \end{aligned}$$

4. *Compute unconditional probabilities according to $p_t^{(jm)} = \sum_{i=1}^{K}\sum_{l=1}^{N} p_{t-1}^{(il)} p_t^{(il,jm)}$.*

In Appendix A.3, we explicitly solve distortions, gradients and transition probabilities for the HN-GARCH case. We experimented with various optimization methods, including trust-region methods using both analytical gradients and Hessians[13]. We generally find the fastest convergence is achieved by a simple BFGS algorithm under a proposed change of variable (presented in Appendix A.5) and analytical gradients. Preliminary experiments suggest global optimization routines such as the Basin-Hopping algorithm[14] fail to improve optimal distortions obtained via the BFGS algorithm, presumably due to the large number of dimensions.

Figure 3.6 shows an optimal product quantization at 21 trading days. As anticipated, quantizers in low(high) price and low(high) variance areas have very low probability especially in the asymmetric case; see white areas in the top-right and bottom-left corners of the right panel. Similarly to stochastic Markovian quantization presented in Section 3.4.3, we observe erratic variations in time (not shown) for conditional variance quantizers, but deterministic quantizers are significantly less noisy than their stochastic counterparts. Overall, these observations suggest supports of Markovian laws (informally characterized in Sections 3.4.2 and 3.4.3) are an inherently difficult case for quantization. We next consider a quantization that allows conditional variance quantizers to vary with log-price quantizers and hence better accommodates strong dependence effects.

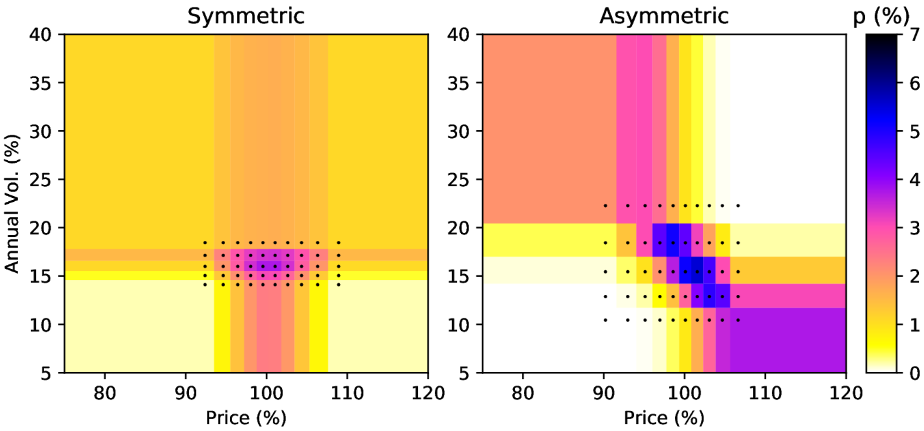
3.5.2 Conditional Quantization

We consider quantizers of the form $\Gamma_t = \{(s_t^{(i)}, h_{t+1}^{(il)}), \quad i = 1, \ldots, K, \quad l = 1, \ldots, N\}$ for $t = 1, \ldots, T$ where univariate quantizers Γ_{s_t} and $\Gamma_{h_{t+1}}^{(i)}$ are assumed ordered. More precisely, we have one variance quantizer per price quantizer element given by $\Gamma_{h_{t+1}}^{(i)} = \{h_{t+1}^{(i1)}, \ldots, h_{t+1}^{(iN)}\}$ for $i = 1, \ldots, K$. Quantizers are still assumed ordered for $t > 0$, i.e. $h_{t+1}^{(il+1)} > h_{t+1}^{(il)}$ for $l = 1, \ldots, N$ with $h_t^{(i0)} = 0$ and $h_t^{(iN+1)} = \infty$ by convention. As usual, we consider initial variance quantizers $\Gamma_{h_1}^{(i)} = \{h_1, \ldots, h_1\}$ (with N elements) for $i = 1, \ldots, K$

[13] See Nocedal and Wright (2006) for more on numerical optimization routines such as trust-region or BFGS algorithms.

[14] See e.g. of Wales and Doye (1997).

Figure 3.6: Optimal product quantization (Algorithm 3) at 21 trading days with $K = 10$ and $N = 5$.



for notational convenience.

The bi-variate quantizer Γ_t is now an *irregular* rectangular grid over the price-variance plane. Introducing

$$C^{(il)}_{h_{t+1}} = \left(\frac{h^{(il-1)}_{t+1} + h^{(il)}_{t+1}}{2}, \frac{h^{(il)}_{t+1} + h^{(il+1)}_{t+1}}{2} \right),$$

for $t > 0$, $i = 1, \ldots, K$ and $l = 1, \ldots, N$, we propose the following *conditional* quantization

$$\begin{pmatrix} \hat{s}_t \\ \hat{h}_{t+1} \end{pmatrix} = \begin{pmatrix} \sum\limits_{i=1}^{K} s^{(i)}_t \mathrm{I}(s_t \in C^{(i)}_{s_t}) \\ \sum\limits_{i=1}^{K} \mathrm{I}(s_t \in C^{(i)}_{s_t}) \sum\limits_{l=1}^{N} h^{(il)}_{t+1} \mathrm{I}(h_{t+1} \in C^{(il)}_{h_{t+1}}) \end{pmatrix} \tag{3.10}$$

where variance tessellations are selected *conditionally* —hence the name— on the contemporaneous price quantization. Similarly to the product quantization approach, we derive analytical expressions for transition probabilities $p^{(il,jm)}_t = \mathbb{P}(\hat{s}_t = s^{(j)}_t, \hat{h}_{t+1} = h^{(jm)}_{t+1} | \hat{s}_{t-1} = s^{(i)}_{t-1}, \hat{h}_t = h^{(il)}_t)$ and unconditional probabilities $p^{(il)}_{t-1} = \mathbb{P}(\hat{s}_{t-1} = s^{(i)}_{t-1}, \hat{h}_t = h^{(il)}_t)$ with $p^{(11)}_0 = 1$ (and 0 otherwise). We propose a conditional quantization optimization procedure in Algorithm 4.

Algorithm 4. *Conditional Quantization*

For $t = 1, \ldots, T$, assuming the optimal time $t-1$ conditional quantizer Γ_{t-1} and corresponding unconditional probabilities $p^{(il)}_{t-1}$ are known at $t = 1$ and otherwise known by recursion,

1. *Find a log-price quantizer* $\Gamma_{s_t} = \{s_t^{(1)}, \dots, s_t^{(K)}\}$ *minimizing the recursive distortion given by* (3.9)*;*

2. *For* $j = 1, \dots, K$*, find a conditional variance quantizer* $\Gamma_{h_{t+1}}^{(j)} = \{h_{t+1}^{(j1)}, \dots, h_{t+1}^{(jN)}\}$ *minimizing the distortion of* $\tilde{h}_{t+1} = \mathscr{H}(\hat{h}_t, z_t)$ conditional *on* $\{\tilde{s}_t \in C_{s_t}^{(j)}\}$*,*

$$\begin{aligned} D_{h_{t+1}}^{(j)} &= E\left[\sum_{l=1}^{N} I(\tilde{h}_{t+1} \in C_{h_{t+1}}^{(jl)})(\tilde{h}_{t+1} - h_{t+1}^{(jl)})^2 \middle| \tilde{s}_t \in C_{s_t}^{(j)}\right] \\ &\propto \sum_{i=1}^{K}\sum_{l=1}^{N} p_{t-1}^{(il)} \int_{-\infty}^{\infty} I(\mathscr{S}(s_{t-1}^{(i)}, h_t^{(il)}, z) \in C_{s_t}^{(j)}) \\ &\qquad \sum_{m=1}^{N} I(\mathscr{H}(h_t^{(il)}, z) \in C_{h_{t+1}}^{(jm)})(\mathscr{H}(h_t^{(il)}, z) - h_{t+1}^{(jm)})^2 d\varphi(z) \end{aligned}$$

where the final expression holds up to a normalization factor which may safely be disregarded during the optimization process;

3. *Compute transition probabilities according to*

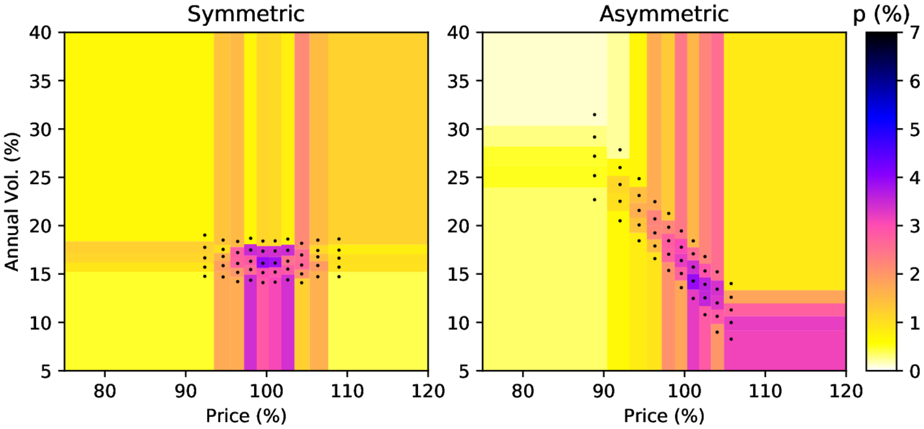
$$\begin{aligned} p_t^{(il,jm)} &= \mathbb{P}(\tilde{s}_t \in C_{s_t}^{(j)}, \tilde{h}_{t+1} \in C_{h_{t+1}}^{(jm)} | s_{t-1} = s_{t-1}^{(i)}, h_t = h_t^{(il)}) \\ &= \int_{-\infty}^{\infty} I(\mathscr{S}(s_{t-1}^{(i)}, h_t^{(il)}, z) \in C_{s_t}^{(j)}) I(\mathscr{H}(h_t^{(il)}, z) \in C_{h_{t+1}}^{(jm)}) d\varphi(z); \end{aligned}$$

4. *Compute unconditional probabilities according to* $p_t^{(jm)} = \sum_{i=1}^{K}\sum_{l=1}^{N} p_{t-1}^{(il)} p_t^{(il,jm)}$.

Explicit expressions for distortions, gradients and transition probabilities are solved under the HN-GARCH specification in Appendix A.4 and implementation concerns discussed in Appendix A.5 also apply to conditional quantization. Step 1 of Algorithm 4 is identical to product quantization, but Step 2 iterates over $j = 1, \dots, K$ and is thus significantly more demanding. This step is *embarrassingly* parallel so that the burden under large K may be mitigated by additional processing units. Also, transition probabilities are obtained as a by-product of Step 2 such that no additional computations are required for Step 3 as opposed to product quantization.

Figure 3.7 displays an optimal conditional quantization at 21 trading days. We readily visualize the significant advantage of conditional quantization over product quantization for the asymmetric case. Optimal quantizers are indeed placed diagonally as opposed to being constrained to a rectangle. This leads to a more efficient utilization of resources i.e. with significantly less low probability (white) areas. The added flexibility appears less exploited in the symmetric case. We next confirm these intuitions and compare all methodologies presented so far under marginal metrics.

Figure 3.7: Optimal conditional quantization (Algorithm 4) at 21 trading days with $K = 10$ and $N = 5$.



3.6 Summary of Numerical Results

As discussed by Pagès and Sagna (2015), uniform dispatching (i.e. $P_t = P$) is suboptimal, but minimizes overall complexity and hence estimation/storage costs of transition probabilities. Under deterministic approaches, an additional practical concern is the allocation of price-variance quantizer sizes (K,N) for some budget $P = KN$. We assume K and N are uniform through time with one notable exception; for conditional quantizations we let $(K = P,\, N = 1)$ at $t = 1$ towards accommodating the parabolic shape of optimal quantizers. We emphasize that this shape is *a priori* inferred from supports of targeted recursive distributions; see e.g. top panels of Figure B.4 in the Appendix.

Figure 3.8 presents total quantization errors over 21 trading days for different (K,N) allocations under both product and conditional approaches. In order to mitigate overfitting concerns, we limit allocations to three scenarios: the same number of quantizer elements, twice the number of quantizer elements for prices and twice the number of quantizer elements for variances. We find the same number of elements for both price and variance quantizers is typically best. The only exception is under the asymmetric case for conditional quantization where allocating twice the number of quantizer elements to prices is better[15].

Figure 3.8: TRMS at 21 trading days for product and conditional quantizations under three (K,N) allocations: the same number of elements for both price and variance quantizers with ($K = 17$, $N = 17$) [$K = N$], twice the number of price elements as variance elements with ($K = 24$, $N = 12$) [$K = 2N$], and vice-versa with ($K = 12$, $N = 24$) [$N = 2K$].

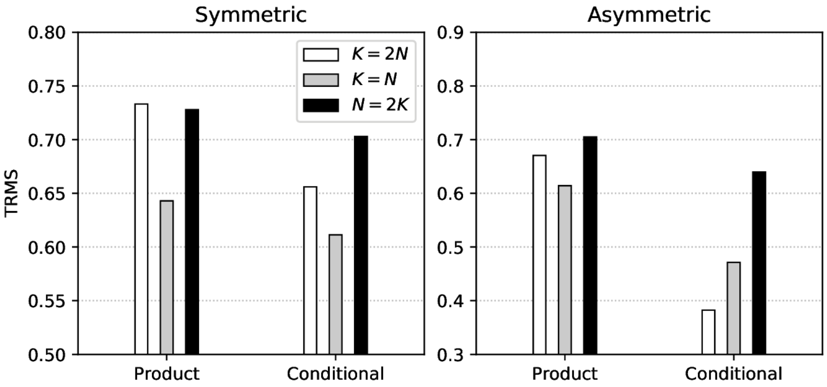
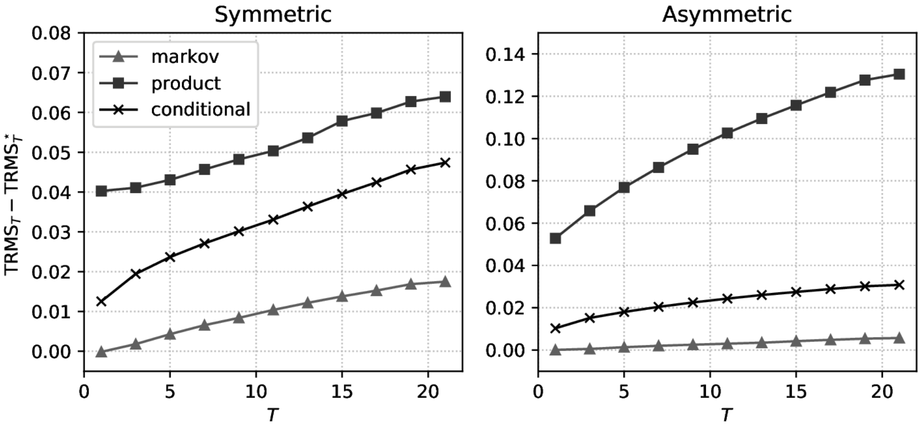


Figure 3.9 shows total recursive quantization errors in excess of the marginal approach. We observe roughly *linear* trends in total errors past the first few time steps. Figure B.3 in the Appendix shows corresponding *daily* excess recursive quantization errors. Daily errors stabilize in time —and even slightly decrease for deterministic approaches in the asymmetric case. These observations mitigate concerns related to the transmission and possible accumulation of errors in time under recursive approaches. Figure B.3 makes apparent the instability in time faced by recursive approaches when compared to corresponding —much smoother— curves under marginal quantization in Figure B.2. For example, the left panel of Figure B.3 shows an anomaly under product quantizations at $t = 15$ which is most likely related to the BFGS algorithm converging to a local minimum.

[15] This is a desirable feature for dynamic programming as the price dimension is often the primary source of convexity (and hence of errors) in value functions e.g. when pricing options.

Figure 3.9: TRMS of all recursive approaches over the marginal approach through time i.e. $\mathrm{TRMS}_T - \mathrm{TRMS}^\star_T$ where $\mathrm{TRMS}^\star$ represents total quantization errors under marginal quantizations of Section 3.4.2. We use $P = 1681$ for both marginal and Markovian stochastic approaches, ($K = 41$, $N = 41$) for both product and conditional quantization under the symmetric case and ($K = 58$, $N = 29$) under the asymmetric case for conditional quantization.



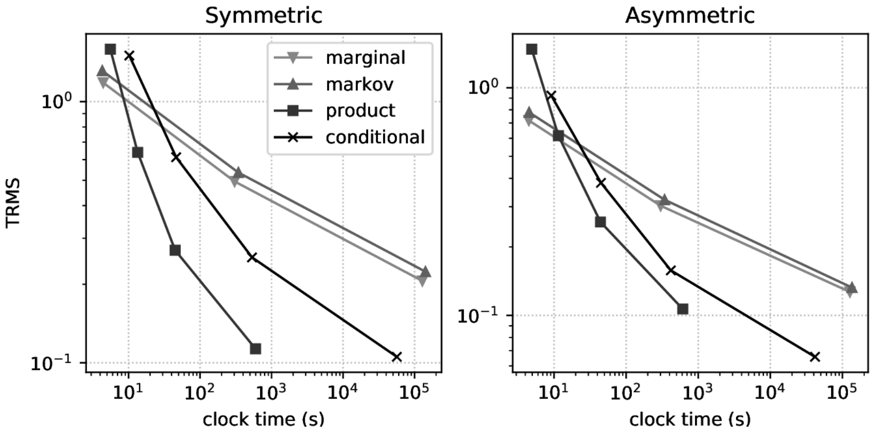
The stochastic Markovian approach significantly over-performs deterministic approaches for a given budget $P \approx KN$. This is not surprising given strong rectangular (i.e. grid-like) constraints imposed on deterministic quantizations. Conditional quantization significantly over-performs product quantization in both cases. While this is expected — and undoubtedly confirmed— for the asymmetric case, the conditional approach further roughly halves errors under product quantization for the *symmetric* case. This is mostly due to letting ($K = P$, $N = 1$) at $t = 1$ under the conditional approach. The flexibility to reallocate variance elements (N) to price elements (K) during early time steps appears as a considerable advantage of conditional quantization over product quantization.

At this point, it is unclear why one would rely on deterministic approaches in practice. Figure 3.10 presents main results of practical interest, namely total quantization errors against computation times —with lower curves corresponding to higher efficiency. While both marginal and conditional methods are embarrassingly parallel, we focus on serial implementations to avoid (as much as possible) machine-dependent effects[16]. The large-

[16] More precisely, the CLVQ algorithm under stochastic approaches and distortion functions/gradients under deterministic approaches are implemented in pure C and linked using Python/Cython. Computations are done on two Intel Xeon clocked at 2.4GHz using 15MB of (L3) cache and 64GB of random access memory.

scale scenario $P = 9801$ could not be computed in a reasonable time (less than a few weeks) for stochastic methods and was hence omitted[17].

Figure 3.10: TRMS for 21 trading days against corresponding computational times in clock seconds. For clarity, both axes are shown in logarithmic scale. For each approach, markers represent (P, K, N) triplets (49,7,7), (289,17,17), (1681,41,41) and (9801,99,99), with the exception of the conditional approach for the asymmetric case where triplets are (49,10,5), (289,24,12), (1681,58,29) and (9801,140,70). The case $P = 9801$ is omitted for stochastic approaches as it could not be computed in a reasonable time.



Stochastic Markovian quantization performs poorly for all sizes. The increase in computation times over the marginal approach is mostly driven by the cost of performing simulations at each time step, rather than in a single stride[18]. Simulation and nearest neighbor search costs overwhelm probability estimation costs under both stochastic approaches. In particular, harvesting CLVQ companion parameters has a negligible impact due to fast $\mathcal{O}(\log(P))$ KD-trees being utilized in Monte Carlo estimators. For small quantizers (e.g. with 100 elements), marginal stochastic quantization is the most efficient and definite choice in practice.

For larger quantizer sizes however, we readily see deterministic approaches benefit from exponential efficiency gains, surpassing stochastic methods for quantizers with more

[17] One could relax the splitting method e.g. by adding 10 to 100 elements instead of a single element at each step. But this approach raises additional concerns related to the optimal trade-off between the number of elements added and quantization quality.

[18] Simulations are performed using *numpy*, a widespread Python package warping low-level C arrays. This package implicitly parallelizes matrix operations (e.g. similarly to MATLAB) such that performing a single GARCH time-series simulation for a given T is much more efficient than performing T one-step Markovian simulations.

than roughly 300 elements. Product quantization performs surprisingly well, especially under symmetric models. The conditional approach offers the lowest achievable (i.e. in a reasonable time) distortion under both symmetric and asymmetric cases by a margin of respectively 0.77% and 4.1% TRMS. Given sufficient computational units, times under conditional quantization could theoretically approach times under product quantization. A parallelized implementation of conditional quantization hence appears as the most efficient, stable and versatile (i.e. under leverage effect or not) choice in practice. The simplicity of product quantization could still warrant its use e.g. as a direct replacement of ad hoc regular grids typically used in dynamic programming applications; as discussed earlier in Section 3.2.

Marginal metrics under recursive approaches are only indirectly impacted by transition probabilities through targeted laws. In the next section, we turn to practical applications which explicitly rely on GARCH dynamics.

3.7 Applications

We consider three applications under the asymmetric case using conditional quantization: European option pricing, American option pricing, and variance-optimal hedging. Our focus on conditional quantization is motivated by its overwhelming marginal efficiency in the asymmetric case, the availability of analytical expressions for transition probabilities (not explicitly tested so far), and the fact large-scale ($KN > 5000$) quantizers may be optimized in a reasonable time. Experimentations are performed in highly stylized settings in the context of the S&P 500 index and numerical results are provided as *proofs of concept*. Thorough empirical (i.e. using actual market data) and numerical work in any particular setting (i.e. pricing or hedging) is left for future research.

Given the results in Section 3.6, we henceforth use twice the number of elements for log-prices as for conditional variances, with the exception of the first time step.

3.7.1 European Call Option Pricing

We compare European call option prices obtained via the quantization of a risk-neutral HN-GARCH model against semi-analytical expressions of Heston and Nandi (2000). Under the change of numéraire introduced in Section 3.3, call option prices are given by

$$c(k,T,h_1) = \mathrm{E}\left[\left(e^{s_T} - e^k\right)^+\right]$$

where $(x)^+ = \max(x,0)$, k is a *forward* log-strike price related to observable strike prices K by $k = \log(K/S_0) - (r-\delta)T$ where r is a risk-free rate and δ is a dividend yield[19]. Observable option prices may be recovered according to $C = cS_0e^{-\delta T}$.

According to empirical results of Heston and Nandi (2000), the GARCH option pricing model is most relevant when variances are strongly correlated with prices i.e. under a leverage effect. We hence risk-neutralize the previously introduced *asymmetric* model. Risk-neutral parameters are obtained directly under the variance-dependent pricing kernel of Christoffersen et al. (2013) which allows for a variance risk premium. This premium is due to higher volatility states being perceived (in equilibrium) as worst market scenarios; see e.g. Gârleanu et al. (2009) and references therein for more empirical insights. We use a variance aversion parameter which roughly matches the estimate on S&P 500 option and log-return data from Christoffersen et al. (2013)[20]. Table 3.2 shows risk-neutral parameters.

The risk-neutral persistence is slightly higher (0.965 versus 0.962) and the long-run log-return variance is markedly higher (23.5% versus 18%) as expected under variance aversion. The γ parameter is slightly lower (152 versus 190) which —while surprising— is consistent with variance aversion[21]. Investors are indeed averse to surges in variance caused by extreme *positive* log-returns, hence shifting some of the pricing kernel weight

[19] See the empirical supplement of Lamarre et al. (2017) for hypotheses underlying this change of numéraire and other empirical implications.

[20] More precisely, we set the ξ parameter of Christoffersen et al. (2013) to $23,000$ in line with their estimate of $(1-2\alpha\xi)^{-1} \approx 1.26$ and adjust parameters according to their Equation (11) where $\phi = -(\mu - 1/2 + \gamma)(1-2\alpha\xi) + \gamma - 1/2$.

[21] In particular, the classical pricing principle under no variance aversion of Duan (1995) would increase the γ parameter; see Proposition 1 of Heston and Nandi (2000).

towards the right log-price tail. This setting does not rely on the local risk-neutral valuation relationship of Duan (1995) which in contrast leaves conditional variance dynamics untouched.

Table 3.2: Parameters of HN-GARCH model (3.3) in the risk-neutral (asymmetric) case.

Parameter	Risk-Neutral
μ	0
ω	4.14e−7
α	7.16e−6
β	0.8
γ	152

Following empirical option pricing convention, we present results as Black-Scholes implicit volatilities [IVs] defined as the σ parameter solving (for k and T fixed) $c = \Phi(d_+(0,k,\sigma,T)) - e^k\Phi(d_-(0,k,\sigma,T))$ for a given c where $d_\pm(s,k,\sigma,\tau) = (s-k\pm\sigma^2\tau/2)/(\sigma\sqrt{\tau})$. We underline that the change of numéraire used here greatly streamlines the Black-Scholes framework. Due to put-call parity, there exists a single IV for any put and call pair of a given maturity and strike price. Without loss of generality, we hence consider call options only. B&S is typically not a valid empirical pricing model and IV is merely used as a normalization function towards controlling variations in days-to-maturity and/or strike price.

Under an optimal conditional quantization as obtained in Section 3.5.2, option prices are given by

$$\hat{c} = \sum_{i=1}^{K}\sum_{l=1}^{N} p_T^{(il)} \left(e^{s_T^{(i)}} - e^k\right)^+ .$$

The left and right panels of Figure 3.11 display the convergence of IVs in respectively strike and days-to-maturity space. Figure 3.12 shows more accurately the convergence for two monthly contracts: at-the-money ($k = 0\%$, $T = 21$) in the left panel and out-of-the-money ($k = 2\%$, $T = 21$) in the right panel.

First, we note prices obtained from the quantization approach underestimate true prices, as evidenced by IVs converging from below in Figures 3.11 and 3.12[22]. This

[22] IV is a strictly monotonic increasing function of option prices.

Figure 3.11: Annual IVs for European call options obtained via the conditional quantization methodology. In the left panel, we show IVs as a function of k for 21 trading-days-to-maturity options and in the right panel as a function of trading-days-to-maturity for an at-the-money ($k = 0$) option. True prices are computed according to semi-analytical expressions of Heston and Nandi (2000).

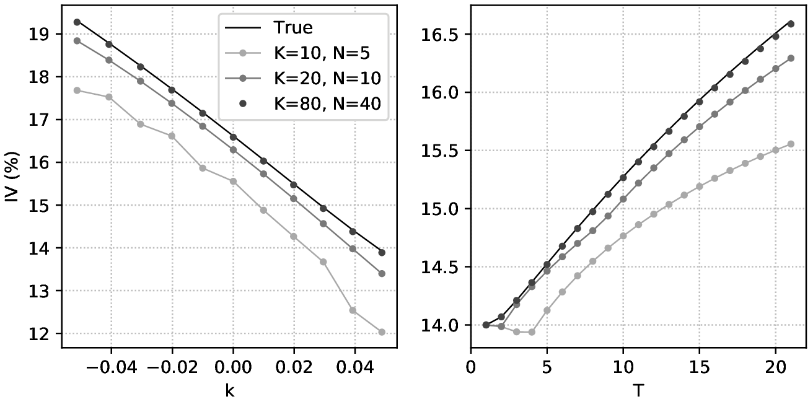
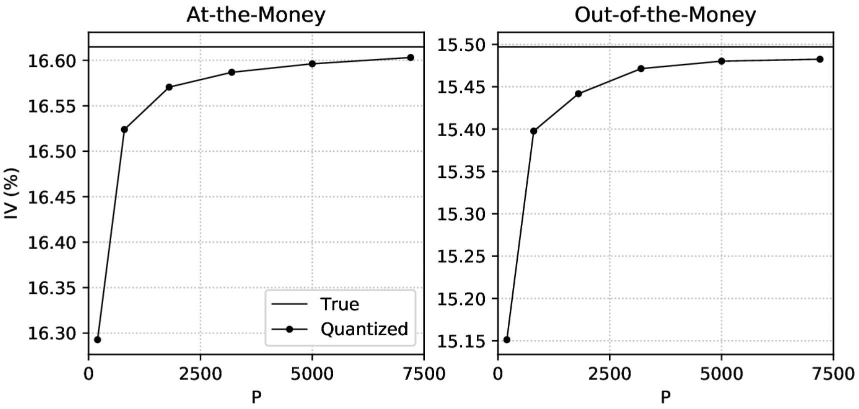


Figure 3.12: Annual IVs for 21-trading-days-to-maturity European call options obtained via the conditional quantization methodology as a function of quantizer size ($KN = 200$, 800, 1800, 3200, 5000, 7200). In the left panel, we show an at-the-money option ($k = 0$) and in the right panel an out-of-the-money option ($k = 0.02$). True prices are computed according to semi-analytical expressions of Heston and Nandi (2000).



is consistent with Jensen's inequality. As noted by Pagès and Printems (2003), we indeed have $\mathrm{E}[f(\hat{x}_T)] = \mathrm{E}[f(\mathrm{E}[x_T|\hat{x}_T])] \leq \mathrm{E}[\mathrm{E}[f(x_T)|\hat{x}_T]] = \mathrm{E}[f(x_T)]$ due to optimal quantizers being stationary i.e. $\mathrm{E}[x_T|\hat{x}_T] = \hat{x}_T$. Second, the at-the-money error is 0.012% when $KN = 7200$. This error translates to 5 cents for the S&P 500 index with $S_0 \approx 2800$, significantly less than typical bid-ask spreads of roughly 30 cents as of late 2018.

Even though semi-analytical expressions are available, quantized European option pricing still has practical uses when computation times are critical. Given an optimal quantization, we indeed compute a hundred prices in roughly 0.5 seconds for $KN = 7200$ as opposed to roughly 35 seconds under Heston and Nandi (2000). These computation times are achieved using a naive sequential python implementation. More sophisticated implementations would most likely be suitable for *real-time* applications with numerous option contracts such as live data streaming or high-frequency option trading. When markets fluctuate greatly intra-daily, a quantization previously optimized for some initial daily variance h_1 likely becomes irrelevant for real-time applications. To prevent such scenarios, we may devise *ad hoc* schemes which account for intra-daily market realizations[23].

3.7.2 American Put Option Pricing

We now turn to American options i.e. allowing for the possibility of early exercise. Our benchmark is given by the least squares Monte Carlo [LSM] methodology of Stentoft (2005) which is an adaptation for GARCH models of Longstaff and Schwartz (2001). This choice is motivated by LSM being the most popular in practice. But LSM unfortunately yields biases of *opposite* signs to the quantization approach[24]. Significant pre-asymptotic differences are hence anticipated, with true prices lying somewhere between both estimators. We emphasize biases are *not* an artifact of the quantization approach *per se*, but an inherent feature of the two main American option pricing frameworks; value function approaches such as Carriere (1996), Duan and Simonato (2001) and our current approach have positive biases, whereas stopping time approaches such as Longstaff and Schwartz (2001) have negative biases[25].

[23] For completeness, we propose such a scheme here. We find one optimal quantization for each element of a pre-defined *grid* of h_1. Intra-daily quantizations are then selected by predicting h_1 using the real-time partial daily market log-return (and h_0) and selecting quantizations according to the nearest neighbor rule. An obvious choice for the conditional variance grid is an optimal quantizer targeting $h_1|h_0$. The set of quantizations is then re-optimized and stored following the re-estimation of model parameters, typically on a daily basis when markets are closed (i.e. overnight).

[24] See e.g. Chapter 5 of Bell et al. (2013).

[25] In the former case, the positive bias is related to Jensen's inequality and the maximum operator being convex; see the first proof in Section 5.1 of Carriere (1996). In the latter case, optimal stopping time estimators are by construction suboptimal, hence yielding lower continuation values. See Stentoft (2010) for a thorough numerical comparison of both approaches.

American option pricing is relevant in the presence of interests for in-the-money put options and/or dividends for in-the-money call options as the early exercise premium is otherwise worthless. Here, we focus on the former case i.e. put options under interests. To do so, we first introduce *spot* log-prices $\tilde{s}_t = s_t + tr$ for $t = 0, \dots, T$ where r is a constant risk-free rate (and $\tilde{s}_0 = 0$). To ensure early exercise effects are significant, we consider a high annual risk-free rate of 5% i.e. $r \approx 0.0002$ under no dividends. Observable asset prices are $S_t = S_0 e^{\tilde{s}_t}$ for some initial asset price S_0.

Following the notation introduced in Section 3.3, we formalize the valuation function,

$$v_t(s_t, h_{t+1}) = \max\left(e^{-r}\mathrm{E}\left[v_{t+1}\middle|\mathscr{F}_t\right], (e^{\tilde{k}} - e^{\tilde{s}_t})^+\right)$$

for $t = 0, \dots, T-1$ where $\tilde{k} = \log(K/S_0)$ is a *spot* log-strike price for some observable strike price K and $v_T = (e^{\tilde{k}} - e^{\tilde{s}_T})^+$. The conditional expectation is computed under the previously introduced *risk-neutral* model under variance aversion; see Table 3.2 for parameters. Strictly speaking, this setting characterizes Bermudian options i.e. with early exercise only allowed on market closes. Under conditional quantizations of Section 3.5.2, the problem is solved similarly to Duan and Simonato (2001) or Bally et al. (2005) as a backward recursion for $t = T-1, \dots, 0$,

$$\hat{v}_t^{(il)} = \max\left(e^{-r}\sum_{j=1}^{K}\sum_{m=1}^{N} p_t^{(il,jm)}\hat{v}_{t+1}^{(jm)}, (e^{\tilde{k}} - e^{\tilde{s}_t^{(i)}})^+\right),$$

for $i = 1, \dots, K$ and $l = 1, \dots, N$[26] where $\hat{v}_T^{(jm)} = (e^{\tilde{k}} - e^{\tilde{s}_T^{(j)}})^+$. Quantizers are optimized for *forward* log-prices s_t first —as done throughout this paper— and then converted to spot values $\tilde{s}_t^{(i)} = s_t^{(i)} + rt$ for $i = 1, \dots, K$. The option price is finally given by $\hat{c} = \hat{v}_0^{(11)}$ with observable option prices recovered according to $C = S_0 c$.

Regarding LSM, we first obtain unconditional GARCH simulations using Algorithm B.5 (omitting the third step) and then accumulate interests according to $\tilde{s}_t = s_t + tr$. When

[26] For notational convenience, we assume fixed quantizer sizes under conditional quantization. Remember from Section 3.6, we set $N = 1$ for $t = 1$ such that $\hat{v}_1^{(il)}$ is actually defined over $i = 1, \dots, P$ and $l = 1$ for some fixed (overall) number of elements P. The shape of transition probability matrices for $t = 0, 1$ and corresponding summation indices in $\hat{v}_0^{(il)}$ are easily adjusted.

estimating continuation values, we consider quadratic and cross-product terms

$$\{1, e^{\tilde{s}_t}, h_{t+1}, e^{2\tilde{s}_t}, h_{t+1}^2, e^{\tilde{s}_t} h_{t+1}\}$$

as regressors and discounted cash flows (as opposed to previously estimated continuation values) of currently in-the-money options as the dependent variable. This setting corresponds to the standard algorithm of Longstaff and Schwartz (2001).

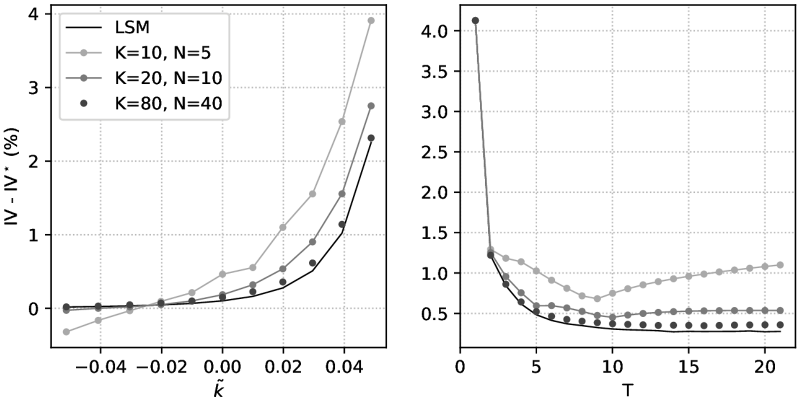
We present empirical results in terms of IV as introduced in Section 3.7.1. The Black-Scholes model is given by $c = e^k \Phi(-d_-(0,k,\sigma,T)) - \Phi(-d_+(0,k,\sigma,T))$ for put options where we first appropriately recover *forward* log-strike prices from *spot* log-strike prices according to $k = \tilde{k} - rT$. American option prices are theoretically bounded from below by their European counterparts and American option IVs are well-defined if corresponding European IVs are well-defined[27]. It hence makes sense to empirically focus on the IV early exercise premium, defined as $\text{IV} - \text{IV}^\star > 0$ where IV and $\text{IV}^\star$ are respectively computed from matching American and European option prices.

The left and right panels of Figure 3.13 show the IV early exercise premia in respectively strike and days-to-maturity space. The LSM algorithm is performed using 10,000,000 simulations such that resulting 95% confidence bounds are hardly distinct (and hence omitted for clarity). We confirm American option pricing is most relevant for in-the-money options as early exercise quickly becoming otherwise worthless (i.e. for $\tilde{k} \ll 0$).

Figure 3.14 shows the convergence for two monthly in-the-money option contracts where the LSM algorithm is now performed under 1,000,000 simulations with resulting 95% confidence intervals shown as dashed lines. As anticipated, we observe significant differences between both approaches due to opposite biases. For American options written on SPY, the premium difference of roughly 0.075% (for $KN = 7200$) in the left panel of Figure 3.14 translates to 2.3 cents (with $S_0 \approx 280$), as compared to corresponding bid-

[27] Strictly speaking, European put option prices are bounded from above by $c < e^k$ from absence of static arbitrage opportunities. European prices are typically sufficiently away from upper bounds that American prices also remain within static arbitrage bounds —at least under realistic interest rates.

Figure 3.13: Annual early exercise IV premia for American put options obtained via the conditional quantization methodology. In the left panel, we show early exercise premia as a function of $\tilde{k}$ for 21 trading-days-to-maturity options and in the right panel as a function of trading-days-to-maturity for an in-the-money option ($\tilde{k} = 0.02$) option. LSM premia are computed according to Stentoft (2005) under 10,000,000 simulations with 95% confidence bounds omitted for clarity.



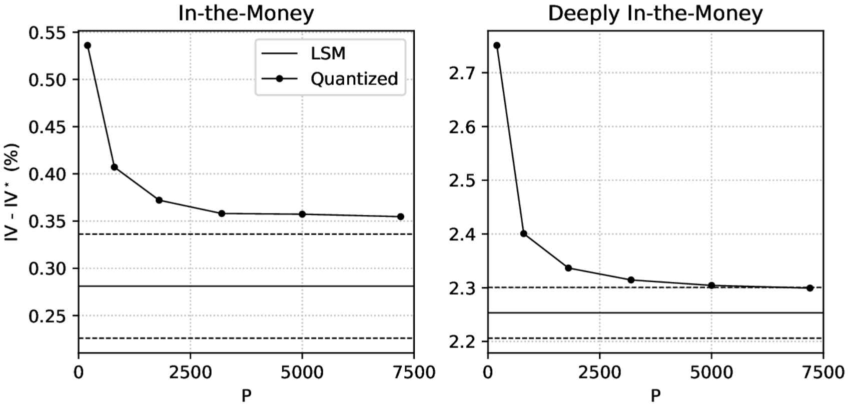
ask spreads of roughly 10 cents as of late 2018[28]. Remember that the actual error is slightly less than 2.3 cents since LSM is negatively biased.

This accuracy is quite remarkable given a single American option price is computed in less than 0.3 seconds under the quantization approach (for $KN = 7200$) as opposed to roughly 6 seconds under LSM with $1,000,000$ simulations[29]. Given the very fast convergence of the quantization approach, the accuracy provided by $KN = 3200$ might suffice for most practical applications —in which case an option is priced under roughly 0.05 seconds. While numerical results from Stentoft (2010) suggest LSM-type algorithms have the smallest *absolute* biases, our preliminary results undoubtedly illustrate the potential of quantization towards real-time and/or large-scale applications such as discussed in the last paragraph of Section 3.7.1.

[28] SPY is an exchanged-traded fund which is designed to track roughly one tenth of the S&P 500 index level. Due to the scale invariance property of GARCH models, GARCH parameters for SPY are presumably very similar to parameters for the S&P 500 index i.e. our asymmetric case. The back-of-the-envelope calculation presented here is for illustration purposes only. In particular, monthly rates were much closer to 2% in late 2018 and dividend yields were also significant. Both effects would have a negative impact on early exercise premia and would hence *decrease* absolute errors in price.

[29] Computation times are obtained under a naive Python implementation and could most likely be improved (in absolute) using pure C.

Figure 3.14: Annual early exercise IV premia for 21-trading-days-to-maturity American put options obtained via the conditional quantization methodology as a function of quantizer size ($KN = 200, 800, 1800, 3200, 5000, 7200$). In the left panel, we show an in-the-money option ($\tilde{k} = 0.02$) and in the right panel a *deeply* in-the-money option ($\tilde{k} = 0.05$). LMS premia are computed according to Stentoft (2005) under 1,000,000 simulations. Corresponding 95% confidence bounds are estimated by bootstrap using 10,000 LSM runs and are shown as dashed lines.



3.7.3 Variance-Optimal Hedging

For the final application, we consider variance-optimal hedging as proposed by Schweizer (1995). While this setting provides a tractable solution to the challenging problem of market incompleteness, it does so under an assumed quadratic criterion which is inconsistent with increasing utility of wealth because of large hedging gains being penalized as much as large hedging losses. Still the fact variance-optimal hedging ratios converge to their complete market counterparts as time intervals shrink to zero in the Black-Scholes framework legitimizes the approach; see e.g. Prigent (2003).

We consider an empirical application to European put payoffs in the context of monetizing the variance risk premium by selling and hedging *at-the-money* (i.e. $k = 0\%$) S&P 500 index options similarly to Lamarre et al. (2017). Introducing a dynamically rebalanced hedging portfolio value at time $t = 0, \dots, T$

$$v_t = v_0 + \sum_{k=1}^{t} \Delta_k (e^{s_k} - e^{s_{k-1}}),$$

where v_0 is an amount that must be set aside in risk-free instruments at inception and Δ_k represents a predictable exposure over $[t-1, t)$, we focus on the overall profits and losses

[PNL],

$$\mathrm{PNL} = c_0 + v_T - v_0 - (e^k - e^{s_T})^+,$$

where c_0 is a put option price computed under the *risk-neutral* model. Under our change of numéraire, PNLs may be interpreted as *excess* returns from committing $S_0 e^{-\delta T}$ capital for a dividend yield δ and selling and hedging a single option until maturity.

Regarding the benchmark, we consider classical delta-hedging [DH] i.e. the partial derivative of option prices with respect to the asset price. DH presupposes continuous-time portfolio rebalancing. Alexander and Nogueira (2007) further show DH is variance-optimal if and only if price-variance correlation is zero. DH indeed fails to account for the partial derivative of option prices with respect to conditional variances; see e.g. Garcia and Renault (1998). By construction, we hence expect variance-optimal hedging [OH] of Schweizer (1995) to decrease risk with respect to DH under a leverage effect when rebalancing at predetermined discrete time intervals (here at market close).

Variance-optimal hedging ratios indeed minimize the expected squared distance between the portfolio and the payoff,

$$\min_{v_0, \{\Delta\}} \mathrm{E}\left[\left(v_T - (1 - e^{s_T})^+\right)^2\right],$$

where the expectation is taken under a *physical* model i.e. with model parameters given by the asymmetric case in Table 3.1. From Example 1 of Schweizer (1995), such a protocol exists (i.e. their non-degeneracy condition is respected) under conditional quantizations of Section 3.5.2 when $KN > 1$. The solution is then given by[30] $v_0 = c_0^{(11)}$ and

$$\Delta_t^{\mathrm{OH}}(v_{t-1}, s_{t-1}, h_t) = \sum_{i=1}^{K} \mathrm{I}(s_{t-1} \in C_{s_{t-1}}^{(i)}) e^{-s_{t-1}^{(i)}} \sum_{l=1}^{N} \mathrm{I}(h_t \in C_{h_t}^{(il)}) \frac{q_{t-1,1}^{(il)} - v_{t-1} m_{t,1}^{(il)}}{m_{t,2}^{(il)}}$$

[30] This solution is obtained by rearranging expressions in Section 2.1 and 2.2 of Rémillard and Rubenthaler (2013) in the univariate case and solving conditional expectations under quantizations. In particular, γ_t matches their definition, $m_{t,2}$ corresponds to their A_t/S_t^2, $m_{t,1}$ to their μ_t/S_t and $m_{t,0}$ to $\mathrm{E}[\gamma_{t+1}|\mathscr{F}_t]$. Also, their $P_{t+1}|\mathscr{F}_{t-1}$ may be replaced by $\gamma_{t+1}|\mathscr{F}_{t-1}$ as already pointed out by their alternative formulation of Eq. (2.3).

where we compute by backward recursion (i.e. for $t = T, \ldots, 1$),

$$m_{t,p}^{(il)} = \sum_{j=1}^{K} \sum_{m=1}^{N} p_t^{(il,jm)} (e^{s_t^{(j)} - s_{t-1}^{(i)}} - 1)^p \gamma_{t+1}^{(jm)} \quad \text{for } p = 0, 1, 2,$$

$$q_{t-1,p}^{(il)} = \sum_{j=1}^{K} \sum_{m=1}^{N} p_t^{(il,jm)} (e^{s_t^{(j)} - s_{t-1}^{(i)}} - 1)^p c_t^{(jm)} \gamma_{t+1}^{(jm)} \quad \text{for } p = 0, 1,$$

with $\gamma_t^{(il)} = m_{t,0}^{(il)} - (m_{t,1}^{(il)})^2 / m_{t,2}^{(il)} \leq 1$ where $\gamma_{T+1}^{(il)} = 1$ and with

$$c_{t-1}^{(il)} = \frac{q_{t-1,0}^{(il)} - \left(m_{t,1}^{(il)} / m_{t,2}^{(il)} \right) q_{t-1,1}^{(il)}}{\gamma_t^{(il)}}$$

where $c_T^{(il)} = (e^k - e^{s_T^{(i)}})^+$ for $i = 1, \ldots, K$ and $l = 1, \ldots, N$[31]. $m_{t,p}^{(il)}$ and $\gamma_t^{(il)}$ are computed off-line following optimal quantizations. v_0 is interpreted as the *expected cost of hedging* due to $E[v_T - (e^k - e^{s_T})^+] = 0$[32]. Since we have $c_0 = 0.01913$ and $v_0 = 0.01737$, we expect excess monthly PNLs of roughly 0.18% i.e. 2.1% per annum —a premium which is motivated by variance aversion in our toy economy. Under no variance aversion (i.e. under the valuation principle of Duan (1995) and Heston and Nandi (2000)), we instead have $c_0 = 0.01738$ and the framework of Schweizer (1995) appears suitable for *option pricing*.

Regarding the benchmark, we consider

$$\Delta_t^{\text{DH}}(s_{t-1}, h_t) = -\text{E}[e^{s_T - s_t} \text{I}(s_T \leq k) | \mathscr{F}_t]$$

where the expectation is now computed under the *risk-neutral* model and explicit expressions are found in Appendix A.6 for completeness[33]. We let $v_0 = c_0$ i.e. proceeds from the sale of the option.

We perform a numerical experiment under the *physical* model using $S = 10,000,000$ unconditional GARCH simulations —still obtained via Algorithm B.5 (omitting the third

[31] See Footnote 26.

[32] The expectation is taken under the physical measure. See also Remark (3) following Corollary 3.2 of Schweizer (1995).

[33] In Appendix A.6, we further show this choice corresponds to $\partial C_t / \partial \tilde{S}_t$ where $C_t = S_0 e^{rt - \delta T} \text{E}[(e^k - e^{s_T})^+ | \mathscr{F}_t]$ is the observed option price at time t and $\tilde{S}_t = e^{-\delta(T-t)} S_t$ is a total-return index at time t i.e. including dividends.

step). Our focus is on 21 trading days at-the-money put options. For each path $s = 1, \dots, S$, we retrieve hedging protocols $\{\Delta_{t,s}^{\mathrm{OH}}(v_{t-1,s}, s_{t-1,s}, h_{t,s}), \Delta_{t,s}^{\mathrm{DH}}(s_{t-1,s}, h_{t,s})\}_{t,s}$ and compute resulting profits and losses after 21 trading days i.e. $\mathrm{PNL}_s^{\mathrm{OH}}$ and $\mathrm{PNL}_s^{\mathrm{DH}}$.

The left panel of Figure 3.15 shows a histogram of hedging ratios at $t = 10$, whereas the right panel displays estimated PNL densities at maturity. First, we note that OH ratios are typically lower than DH ratios. This is consistent with OH shorting additional units of the underlying asset towards mitigating adverse effects from spikes in volatility (due to the negative price-variance relationship under a leverage effect). From density estimations, we readily observe PNLs under OH are much closer to normality and are slightly less (negatively) skewed. In particular, the very odd shape of the density under DH suggests misspecification.

Figure 3.15: Numerical results from selling and hedging at-the-money monthly put options under delta-hedging [DH] and optimal hedging [OH] obtained via the conditional quantization methodology. The left panel displays hedging ratios under both DH and OH after roughly two weeks. More precisely, we show a two-dimensional histogram with darker colors associated to more observations and where the black line is the identity function (both axes are truncated at -1 and 0 for clarity). The right panel displays Gaussian kernel density estimators for PNLs (truncated at $\pm 2\%$ for clarity).

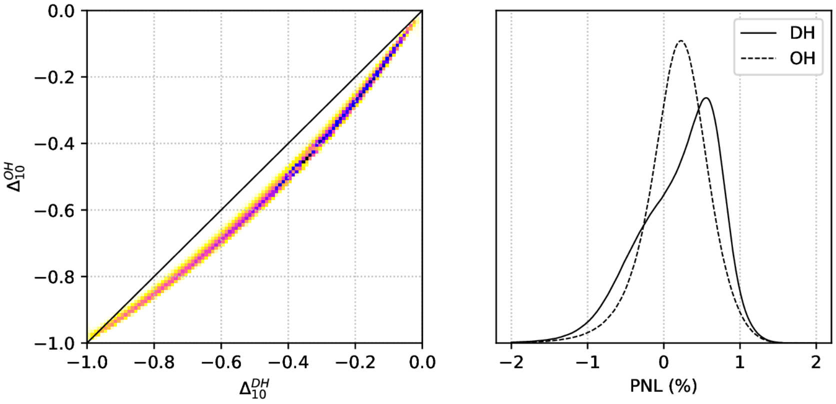


Table 3.3 shows descriptive statistics for PNLs in percent. Mean PNL under OH is consistent with an annual premia of roughly 2.1%, which is the empirical manifestation of $E[v_T - (e^k - e^{s_T})^+] = 0$. Standard deviations further confirm OH minimizes risk. While DH shoulders more risk and reaps more reward (roughly 2.4% per annum), the overall risk-reward trade-off is interestingly best under OH as evidenced by a Sharpe(Omega)

ratio of 0.42(2.90) versus 0.38(2.46). This last observation is in line with empirical observations from Lamarre et al. (2017) using actual S&P 500 data.

The high excess kurtosis under OH (4.67) is mostly explained by 59 (out of 10,000,000) extreme PNL observations falling outside $\pm 5\%$. All such observations correspond to large quantization error realizations, typically with $s_T \gg 0.1$. Roughly speaking, extreme PNLs are related to undue extrapolations caused by market realizations falling *outside* a quantization. Such extrapolations are easily avoided in practice by re-optimizing quantizers following significant market returns. As a final comment, we underline that the numerical experiment is completed under two minutes for the OH quantization approach, whereas DH ratios under Heston and Nandi (2000) are computed in roughly three hours.

Table 3.3: Descriptive statistics of PNL (in percent) when selling and hedging monthly at-the-money ($k = 0\%$) put options under delta-hedging [DH] and optimal hedging [OH] obtained via the conditional quantization methodology.

	DH (%)	OH (%)
Mean	0.1992	0.1763
Std.Dev.	0.5239	0.4207
Sharpe	0.38	0.42
Min.	−4.83	−19.31
Max.	1.77	25.28
Skew.	−0.70	−0.67
Kurt.	0.44	4.67
Omega	2.46	2.90

3.8 Conclusion

We obtain Markov chains approximating GARCH dynamics under a quadratic optimality criterion known as the distortion. We do so both under stochastic and deterministic methodologies. When assuming resulting Markov chains as market models, many financial problems relying on dynamic programming are solved as straightforward recursive summations. When instead assuming Markov chains as *approximating* GARCH dynamics, our approach is relevant towards improving the pre-asymptotic behavior of existing Markov chain-based approximations such as Duan and Simonato (2001).

Numerical results suggest a novel conditional quantization approach is the most efficient in terms of marginal metrics, especially for large quantizer sizes under leverage effects. This approach is successfully applied to three different scenarios corresponding to option pricing and hedging for the S&P 500 index and appears suitable for real-time and large-scale applications such as high-frequency option trading.

Further avenues of research include quantizations under non-normal innovations (e.g. using the empirical law inferred from model residuals) and under other GARCH specifications. The quantization of dynamics based on high-frequency datasets is of particular interest; see e.g. the HEAVY model of Shephard and Sheppard (2010). As pointed out by Pagès and Printems (2003), quantization extensions to high dimensional processes have limited potential when compared to Monte Carlo approaches, but remain relevant e.g. when dynamically optimizing portfolio allocations across a *maximum* of 5 to 7 asset classes.

References

Alexander, C. and Nogueira, L. M. (2007). Model-free hedge ratios and scale-invariant models. *Journal of Banking and Finance*, 31(6):1839–1861.

Bally, V., Pagès, G., and Printems, J. (2001). A stochastic quantization method for non-linear problems. *Monte Carlo Methods and Applications*, 7:21–34.

Bally, V., Pagès, G., and Printems, J. (2005). A quantization tree method for pricing and hedging multidimensional American options. *Mathematical Finance*, 15(1):119–168.

Basak, S. and Chabakauri, G. (2010). Dynamic mean-variance asset allocation. *The Review of Financial Studies*, 23(8):2970–3016.

Bell, A. R., Brooks, C., and Prokopczuk, M. (2013). *Handbook of Research Methods and Applications in Empirical Finance*. Edward Elgar, UK.

Ben-Ameur, H., Breton, M., and Martinez, J. (2009). Dynamic programming approach for valuing options in the GARCH model. *Management Science*, 55(2):252–266.

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Callegaro, G., Fiorin, L., and Grasselli, M. (2016). Pricing via recursive quantization in stochastic volatility models. *Quantitative Finance*, 17(6):855–872.

Carriere, J. F. (1996). Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19:19–30.

Christoffersen, P., Heston, S., and Jacobs, K. (2013). Capturing option anomalies with a variance-dependent pricing kernel. *Review of Financial Studies*, 26(8):1963–2006.

Duan, J.-C. (1995). The GARCH option pricing model. *Mathematical Finance*, 5(1):13–32.

Duan, J.-C. and Simonato, J.-G. (2001). American option pricing under GARCH by a Markov chain approximation. *Journal of Economic Dynamics & Control*, 25:1689–1718.

Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, 48(5):1749–1778.

Fiorin, L., Pages, G., and Sagna, A. (2017). Product Markovian quantization of an R^d -valued Euler scheme of a diffusion process with applications to finance. Working Paper. Available online at http://arxiv.org/abs/1511.01758.

Garcia, R. and Renault, E. (1998). A note on hedging in ARCH and stochastic volatility option pricing models. *Mathematical Finance*, 8(2):153–161.

Gârleanu, N., Pedersen, L. H., and Poteshman, A. M. (2009). Demand-based option pricing. *Review of Financial Studies*, 22(10):4259–4299.

Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.

Graf, S. and Luschgy, H. (2000). *Foundations of Quantization for Probability Distributions*, volume 1730 of *Lecture Notes in Mathematics*. Springer, Berlin.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343.

Heston, S. L. and Nandi, S. (2000). A closed-form GARCH option valuation model. *Review of Financial Studies*, 13(3):585–625.

Kushner, H. J. and Yin, G. G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, New York, N.Y., 2nd edition.

Lamarre, H., Dupuis, D. J., and Rémillard, B. (2017). The economic value of volatility timing using realized volatility for hedged S&P 500 options. Working Paper, HEC Montréal.

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137.

Longstaff, F. A. and Schwartz, E. S. (2001). Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies*, 14(1):113–147.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer-Verlag, Berlin, New York, 2nd edition.

Pagès, G. (1997). A space quantization method for numerical integration. *Journal of Computational and Applied Mathematics*, 89(1):1–38.

Pagès, G. and Pham, H. (2005). Optimal quantization methods for nonlinear filtering with discrete-time observations. *Bernoulli*, 11(5):893–932.

Pagès, G., Pham, H., and Printems, J. (2004a). An optimal Markovian quantization algorithm for multidimensional stochastic control problems. *Stochastics and Dynamics*, 4(4):501–545.

Pagès, G., Pham, H., and Printems, J. (2004b). *Handbook of Computational and Numerical Methods in Finance*. Birkhäuser, Boston, MA.

Pagès, G. and Printems, J. (2003). Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods and Applications*, 9(2):135–165.

Pagès, G. and Sagna, A. (2015). Recursive marginal quantization of the Euler scheme of a diffusion process. *Applied Mathematical Finance*, 22(5):463–498.

Prigent, J.-L. (2003). *Weak Convergence of Financial Markets*. Springer-Verlag Berlin Heidelberg, 1st edition.

Rémillard, B. and Rubenthaler, S. (2013). Optimal hedging in discrete time. *Quantitative Finance*, 13(6):819–825.

Schweizer, M. (1995). Variance-optimal hedging in discrete time. *Mathematics of Operations Research*, 20(1):1–32.

Shephard, N. and Sheppard, K. (2010). Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 25(2):197–231.

Stentoft, L. (2005). Pricing American options when the underlying asset follows a GARCH process. *Journal of Empirical Finance*, (12):576–611.

Stentoft, L. (2010). Value function approximation or stopping time approximation: a comparison of two recent numerical methods for American option pricing using simulation and regression. *Journal of Computational Finance*, 18(1):65–120.

Wales, D. J. and Doye, J. P. (1997). Global optimization by Basin-Hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *The Journal of Physical Chemistry A*, 101(28):5111–5116.

General Conclusion

Firstly, empirical observations from Gârleanu et al. (2009) under market segmentation prompt us to investigate how volatility forecasts of differing statistical quality impact risk-reward trade-offs for a profit-oriented agent extracting risk premia embedded in options. We present robust empirical results corroborating volatility timing results of Fleming et al. (2003); statistical value generated by realized volatility does translate to significant economic value in the option market.

Market option prices are likely the result of a complex equilibrium involving participants with varying degrees of sophistication e.g. with access (or not) to high-frequency datasets. Building theoretical pricing implications under such heterogeneity proves to be challenging. Still, a restrained three-agents model could be sufficiently tractable to yield insights regarding exchanges of un-hedgeable risk factors amongst end-users, market-makers and proprietary traders. Proprietary traders are better positioned for absorbing *long-run* market imbalances, i.e. for hedging underlying market shocks. In contrast, market-makers are naturally better positioned for absorbing *short-lived* imbalances between end-users and proprietary traders, i.e. for hedging option demand shocks. Net welfare gains presumably appear from respectively attributing long- and short-run responsibilities to proprietary traders and market-makers —a conjecture left for future research.

Secondly, we turn to numerical strategies for mitigating computational burdens faced by stochastic optimal control problems under GARCH dynamics. We evidence the strong potential of proposed conditional quantizations with regard to applications for which Monte Carlo is unavailable.

Resulting discrete-state Markov chains are particularly well suited to retirement investment advice in the spirit of Goldstein et al. (2008). Under this framework, investors directly specify *discrete* state probabilities for their retirement income, as opposed to optimizing a presupposed utility function parametrized by risk aversion. The quantization of pricing kernels allows for the construction of *cost-efficient* payoffs in the spirit of Dybvig (1988a,b). Desired retirement income probability distributions can then be targeted under GARCH dynamics in a unified framework e.g. by considering variance-optimal pricing kernel *and* variance-optimal protocols of Schweizer (1995). This quantization application to distributional targeting is promising in regard to some of the challenges raised by *fin-tech* shifts in the financial advisory industry. Recent advances in artificial intelligence paired with proposed Markov chains may indeed help tackle growing customization needs of investors.

References

Dybvig, P. H. (1988a). Distributional analysis of portfolio choice. *The Journal of Business*, 61(3):369–393.

Dybvig, P. H. (1988b). Inefficient Dynamic Portfolio Strategies or How to Throw Away a Million Dollars in the Stock Market. *Review of Financial Studies*, 1(1):67–88.

Fleming, J., Kirby, C., and Ostdiek, B. (2003). The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics*, 67(3):473–509.

Gârleanu, N., Pedersen, L. H., and Poteshman, A. M. (2009). Demand-based option pricing. *Review of Financial Studies*, 22(10):4259–4299.

Goldstein, D. G., Johnson, E. J., and Sharpe, W. F. (2008). Choosing outcomes versus choosing products: Consumer-focused retirement investment advice. *Journal of Consumer Research*, 35(3):440–456.

Schweizer, M. (1995). Variance-optimal hedging in discrete time. *Mathematics of Operations Research*, 20(1):1–32.

Appendix A

Appendix to "On the Marginal and Recursive Quantization of GARCH Models"

HUGO LAMARRE

A.1 Standardization under HN-GARCH

Without further mention, we rely on z_t being independent of h_s for all $t \geq s$, z_t being a standard normal random variate for all $t > 0$ and h_1 being a known constant. In particular, we freely use the following results: $\mathrm{E}[z_t h_s] = \mathrm{E}[z_t]\mathrm{E}[h_s] = 0$ and $\mathrm{E}[z_t^2 h_s] = \mathrm{E}[z_t^2]\mathrm{E}[h_s] = \mathrm{E}[h_s]$ for $t \geq s$ and $\mathrm{E}[h_1] = h_1$ and $\mathrm{Var}[h_1] = 0$. We also follow the usual convention $\sum_{u=1}^{0} x_u = 0$.

A.1.1 Conditional Variance

Lemma B.1. *For $t \geq s > 0$,*

$$E[h_t] = (\omega + \alpha) \sum_{u=1}^{t-s} (\beta + \alpha\gamma^2)^{u-1} + (\beta + \alpha\gamma^2)^{t-s} E[h_s]$$

In particular,

$$E[h_t] = (\omega+\alpha)\sum_{u=1}^{t-1}(\beta+\alpha\gamma^2)^{u-1} + (\beta+\alpha\gamma^2)^{t-1}h_1$$

Proof. We may easily show that $\mathrm{E}[h_{s+1}] = (\omega+\alpha)+(\beta+\alpha\gamma^2)\mathrm{E}[h_s]$ such that the Lemma holds for $t = s+1$. Then by induction,

$$\begin{aligned}\mathrm{E}[h_t] &= (\omega+\alpha)+(\beta+\alpha\gamma^2)\mathrm{E}[h_{t-1}] \\ &= (\omega+\alpha)+(\beta+\alpha\gamma^2)\left((\omega+\alpha)\sum_{u=1}^{t-1-s}(\beta+\alpha\gamma^2)^{u-1}+(\beta+\alpha\gamma^2)^{t-1-s}\mathrm{E}[h_s]\right) \\ &= (\omega+\alpha)+(\omega+\alpha)\sum_{u=1}^{t-1-s}(\beta+\alpha\gamma^2)^{u}+(\beta+\alpha\gamma^2)^{t-s}\mathrm{E}[h_s] \\ &= (\omega+\alpha)\sum_{u=1}^{t-s}(\beta+\alpha\gamma^2)^{u-1}+(\beta+\alpha\gamma^2)^{t-s}\mathrm{E}[h_s]\end{aligned}$$

□

Lemma B.2. *For $t,s > 0$,*

$$E[h_th_s] = (\omega+\alpha)E[h_{s\wedge t}]\sum_{u=1}^{|t-s|}(\beta+\alpha\gamma^2)^{u-1} + E[h^2_{s\wedge t}](\beta+\alpha\gamma^2)^{|t-s|}$$

Proof. Without loss of generality, we assume $t > s$. First note that

$$\begin{aligned}\mathrm{E}[h_th_s] &= \mathrm{E}\left[\left(\omega+\beta h_{t-1}+\alpha(z_{t-1}-\gamma\sqrt{h_{t-1}})^2\right)h_s\right] \\ &= \omega\mathrm{E}[h_s]+\beta\mathrm{E}[h_{t-1}h_s]+\alpha\mathrm{E}[z^2_{t-1}h_s]-2\alpha\gamma\mathrm{E}[z_{t-1}\sqrt{h_{t-1}}h_s]+\alpha\gamma^2\mathrm{E}[h_{t-1}h_s] \\ &= (\omega+\alpha)\mathrm{E}[h_s]+\mathrm{E}[h_{t-1}h_s](\beta+\alpha\gamma^2).\end{aligned}$$

In particular, $\mathrm{E}[h_{s+1}h_s] = (\omega+\alpha)\mathrm{E}[h_s]+\mathrm{E}[h_s^2](\beta+\alpha\gamma^2)$ which proves the Lemma for $t = s+1$. Then for $t > s+1$, we have by induction

$$\begin{aligned}\mathrm{E}[h_th_s] &= (\omega+\alpha)\mathrm{E}[h_s]+\left((\omega+\alpha)\mathrm{E}[h_s]\sum_{u=1}^{t-1-s}(\beta+\alpha\gamma^2)^{u-1}+\mathrm{E}[h_s^2](\beta+\alpha\gamma^2)^{t-1-s}\right)(\beta+\alpha\gamma^2) \\ &= (\omega+\alpha)\mathrm{E}[h_s]\left(1+\sum_{u=1}^{t-1-s}(\beta+\alpha\gamma^2)^{u}\right)+\mathrm{E}[h_s^2](\beta+\alpha\gamma^2)^{t-s} \\ &= (\omega+\alpha)\mathrm{E}[h_s]\sum_{u=1}^{t-s}(\beta+\alpha\gamma^2)^{u-1}+\mathrm{E}[h_s^2](\beta+\alpha\gamma^2)^{t-s}.\end{aligned}$$

□

We are now ready for the main result.

Proposition B.1. *For $t > 0$,*

$$Var[h_t] = 2\alpha^2 \sum_{s=1}^{t-1} (\beta+\alpha\gamma^2)^{2(s-1)} \left(1+2\gamma^2 E[h_{t-s}]\right)$$

Proof. Since $\mathrm{Var}[h_1] = 0$, the result is true for $t = 1$. It is also relatively well-known (see e.g. Equation (7) of Christoffersen et al. (2013)) that $\mathrm{Var}[h_2] = 2\alpha^2 + 4\alpha^2\gamma^2 h_1$ so that the Proposition holds for $t = 2$. Then for $t > 2$,

$$\begin{aligned}
\mathrm{Var}[h_t] &= \mathrm{E}\left[\left(\beta h_{t-1} + \alpha(z_{t-1} - \gamma\sqrt{h_{t-1}})^2\right)^2\right] - \left(\mathrm{E}\left[\beta h_{t-1} + \alpha(z_{t-1} - \gamma\sqrt{h_{t-1}})^2\right]\right)^2 \\
&= \mathrm{E}\big[\alpha^2\gamma^4 h_{t-1}^2 - 4\alpha^2\gamma^3 h_{t-1}^{3/2} z_{t-1} + 6\alpha^2\gamma^2 h_{t-1} z_{t-1}^2 - 4\alpha^2\gamma\sqrt{h_{t-1}} z_{t-1}^3 + \alpha^2 z_{t-1}^4 \\
&\quad + 2\alpha\beta\gamma^2 h_{t-1}^2 - 4\alpha\beta\gamma h_{t-1}^{3/2} z_{t-1} + 2\alpha\beta h_{t-1} z_{t-1}^2 + \beta^2 h_{t-1}^2\big] - \left(\alpha + (\beta+\alpha\gamma^2)\mathrm{E}[h_{t-1}]\right)^2 \\
&= \alpha^2\gamma^4 \mathrm{E}[h_{t-1}^2] + 6\alpha^2\gamma^2\mathrm{E}[h_{t-1}] + 3\alpha^2 + 2\alpha\beta\gamma^2\mathrm{E}[h_{t-1}^2] + 2\alpha\beta\mathrm{E}[h_{t-1}] + \beta^2\mathrm{E}[h_{t-1}^2] \\
&\quad - \left(\alpha^2\gamma^4\mathrm{E}[h_{t-1}]^2 + 2\alpha^2\gamma^2\mathrm{E}[h_{t-1}] + \alpha^2 + 2\alpha\beta\gamma^2\mathrm{E}[h_{t-1}]^2 + 2\alpha\beta\mathrm{E}[h_{t-1}] + \beta^2\mathrm{E}[h_{t-1}]^2\right) \\
&= \left(\alpha^2\gamma^4 + 2\alpha\beta\gamma^2 + \beta^2\right)\left(\mathrm{E}[h_{t-1}^2] - (\mathrm{E}[h_{t-1}])^2\right) + 2\alpha^2 + 4\alpha^2\gamma^2\mathrm{E}[h_{t-1}] \\
&= (\beta+\alpha\gamma^2)^2\mathrm{Var}[h_{t-1}] + 2\alpha^2 + 4\alpha^2\gamma^2\mathrm{E}[h_{t-1}],
\end{aligned}$$

and by induction,

$$\begin{aligned}
\mathrm{Var}[h_t] &= (\beta+\alpha\gamma^2)^2\left(2\alpha^2\sum_{s=1}^{t-2}(\beta+\alpha\gamma^2)^{2(s-1)}\left(1+2\gamma^2\mathrm{E}[h_{t-1-s}]\right)\right) + 2\alpha^2 + 4\alpha^2\gamma^2\mathrm{E}[h_{t-1}] \\
&= 2\alpha^2\left(1 + 2\gamma^2\mathrm{E}[h_{t-1}] + \sum_{s=1}^{t-2}(\beta+\alpha\gamma^2)^{2s}\left(1+2\gamma^2\mathrm{E}[h_{t-1-s}]\right)\right) \\
&= 2\alpha^2\sum_{s=1}^{t-1}(\beta+\alpha\gamma^2)^{2(s-1)}\left(1+2\gamma^2\mathrm{E}[h_{t-s}]\right).
\end{aligned}$$

□

The following Lemma will be useful in the next Section.

Lemma B.3. *For $t, s > 0$,*

$$Cov(h_t, h_s) = (\beta+\alpha\gamma^2)^{|t-s|} Var[h_{t\wedge s}]$$

Proof. Using Lemma B.2 and B.1 and assuming without loss of generality $t > s$,

$$\begin{aligned}
\mathrm{Cov}(h_t, h_s) &= \mathrm{E}[h_t h_s] - \mathrm{E}[h_t]\mathrm{E}[h_s] \\
&= \left((\omega+\alpha)\mathrm{E}[h_s] \sum_{u=1}^{t-s} (\beta+\alpha\gamma^2)^{u-1} + \mathrm{E}[h_s^2](\beta+\alpha^2)^{t-s} \right) \\
&\quad - \left((\omega+\alpha) \sum_{u=1}^{t-s} (\beta+\alpha\gamma^2)^{u-1} + (\beta+\alpha\gamma^2)^{t-s}\mathrm{E}[h_s] \right) \mathrm{E}[h_s] \\
&= (\beta+\alpha\gamma^2)^{t-s}\mathrm{Var}[h_s]
\end{aligned}$$

□

A.1.2 Log-Price

Lemma B.4. *For $t > s > 0$,*

$$E[h_t\sqrt{h_s}z_s] = -2\alpha\gamma\left(\beta+\alpha\gamma^2\right)^{t-s-1} E[h_s]$$

Proof. The Proposition holds for $t = s+1$ since

$$\begin{aligned}
\mathrm{E}[h_{s+1}\sqrt{h_s}z_s] &= \mathrm{E}\left[(\omega + \beta h_s + \alpha(z_s - \gamma\sqrt{h_s})^2)\sqrt{h_s}z_s \right] \\
&= \omega\mathrm{E}[\sqrt{h_s}z_s] + \beta\mathrm{E}[h_s^{3/2}z_s] + \alpha\mathrm{E}[z_s^3\sqrt{h_s}] - 2\alpha\gamma\mathrm{E}[z_s^2 h_s] + \alpha\gamma^2\mathrm{E}[h_s^{3/2}z_s] \\
&= -2\alpha\gamma\mathrm{E}[h_s]
\end{aligned}$$

When $t > s+1$, we similarly have

$$\begin{aligned}
\mathrm{E}[h_t\sqrt{h_s}z_s] &= \mathrm{E}\left[(\omega + \beta h_{t-1} + \alpha(z_{t-1} - \gamma\sqrt{h_{t-1}})^2)\sqrt{h_s}z_s \right] \\
&= \omega\mathrm{E}[\sqrt{h_s}z_s] + \beta\mathrm{E}[h_{t-1}\sqrt{h_s}z_s] + \alpha\mathrm{E}[z_{t-1}^2\sqrt{h_s}z_s] \\
&\quad - 2\alpha\gamma\mathrm{E}[z_{t-1}\sqrt{h_{t-1}h_s}z_s] + \alpha\gamma^2\mathrm{E}[h_{t-1}\sqrt{h_s}z_s] \\
&= (\beta+\alpha\gamma^2)\mathrm{E}[h_{t-1}\sqrt{h_s}z_s]
\end{aligned}$$

since $\mathrm{E}[z_{t-1}^2\sqrt{h_s}z_s] = \mathrm{E}[z_{t-1}^2\sqrt{h_s}]\mathrm{E}[z_s] = 0$ and $\mathrm{E}[z_{t-1}\sqrt{h_{t-1}h_s}z_s] = \mathrm{E}[z_{t-1}]\mathrm{E}[\sqrt{h_{t-1}h_s}z_s] = 0$. Hence by induction,

$$\mathrm{E}[h_t\sqrt{h_s}z_s] = (\beta+\alpha\gamma^2)\left(-2\alpha\gamma\left(\beta+\alpha\gamma^2\right)^{t-1-s-1}\mathrm{E}[h_s]\right) = -2\alpha\gamma\left(\beta+\alpha\gamma^2\right)^{t-s-1}\mathrm{E}[h_s]$$

□

Lemma B.5. *For $t > s$ and $t,s > 0$,*

$$\begin{cases} E[r_t] = (\mu - 1/2)E[h_t], \\ E[r_t^2] = (\mu - 1/2)^2 E[h_t^2] + E[h_t] \\ E[r_t r_s] = (\mu - 1/2)^2 E[h_t h_s] - 2\alpha\gamma(\mu - 1/2)(\beta + \alpha\gamma^2)^{t-s-1} E[h_s] \end{cases}$$

Proof. Regarding the first two statements, we trivially have

$$\begin{aligned} \mathrm{E}[r_t] &= \mathrm{E}[(\mu - 1/2)h_t + \sqrt{h_t} z_t] = (\mu - 1/2)\mathrm{E}[h_t] \\ \mathrm{E}[r_t^2] &= \mathrm{E}[(\mu - 1/2)^2 h_t^2 + 2(\mu - 1/2)h_t^{3/2} z_t + h_t z_t^2] = (\mu - 1/2)^2 \mathrm{E}[h_t^2] + \mathrm{E}[h_t]. \end{aligned}$$

For the last statement,

$$\begin{aligned} \mathrm{E}[r_t r_s] &= \mathrm{E}\left[\left((\mu - 1/2)h_t + \sqrt{h_t} z_t\right)\left((\mu - 1/2)h_s + \sqrt{h_s} z_s\right)\right] \\ &= (\mu - 1/2)^2 \mathrm{E}[h_t h_s] + (\mu - 1/2)\left(\mathrm{E}[h_t \sqrt{h_s} z_s] + \mathrm{E}[h_s \sqrt{h_t} z_t]\right) + \mathrm{E}[\sqrt{h_t h_s} z_t z_s] \\ &= (\mu - 1/2)^2 \mathrm{E}[h_t h_s] + (\mu - 1/2)\mathrm{E}[h_t \sqrt{h_s} z_s] \end{aligned}$$

since $\mathrm{E}[h_s \sqrt{h_t} z_t] = \mathrm{E}[h_s \sqrt{h_t}]\mathrm{E}[z_t] = 0$ and $\mathrm{E}[\sqrt{h_t h_s} z_t z_s] = \mathrm{E}[\sqrt{h_t h_s} z_s]\mathrm{E}[z_t] = 0$. The result follows from Lemma B.4. □

We are now ready for the main result.

Proposition B.2. *For $t > 0$,*

$$\begin{aligned} Var\left[\sum_{s=1}^{t} r_s\right] &= \sum_{s=1}^{t} E[h_s] + (\mu - 1/2)^2 \sum_{s=1}^{t}\sum_{u=1}^{t} Cov[h_s, h_u] \\ &\quad - 4\alpha\gamma(\mu - 1/2) \sum_{s=1}^{t}\sum_{u=1}^{s-1} (\beta + \alpha\gamma^2)^{s-u-1} E[h_u] \end{aligned}$$

Proof. Using the well-know fact

$$\begin{aligned} \mathrm{Var}\left[\sum_{s=1}^{t} x_s\right] &= \sum_{s=1}^{t}\left(\mathrm{Var}[x_s] + 2\sum_{u=1}^{s-1} \mathrm{Cov}[x_s, x_u]\right) \\ &= \sum_{s=1}^{t}\sum_{u=1}^{t} \mathrm{Cov}[x_s, x_u], \end{aligned}$$

and return expectations from Lemma B.5

$$
\begin{aligned}
\operatorname{Var}\left[\sum_{s=1}^{t} r_s\right] &= \sum_{s=1}^{t}\left(\mathrm{E}[r_s^2]-(\mathrm{E}[r_s])^2+2\sum_{u=1}^{s-1}(\mathrm{E}[r_s r_u]-\mathrm{E}[r_s]\mathrm{E}[r_u])\right)\\
&= \sum_{s=1}^{t}\Bigg((\mu-1/2)^2\operatorname{Var}[h_s]+\mathrm{E}[h_s]\\
&\quad +2\sum_{u=1}^{s-1}\left((\mu-1/2)^2\operatorname{Cov}[h_s,h_u]-2\alpha\gamma(\mu-1/2)(\beta+\alpha\gamma^2)^{s-u-1}\mathrm{E}[h_u]\right)\Bigg)\\
&= \sum_{s=1}^{t}\mathrm{E}[h_s]+(\mu-1/2)^2\sum_{s=1}^{t}\sum_{u=1}^{t}\operatorname{Cov}[h_s,h_u]\\
&\quad -4\alpha\gamma(\mu-1/2)\sum_{s=1}^{t}\sum_{u=1}^{s-1}(\beta+\alpha\gamma^2)^{s-u-1}\mathrm{E}[h_u]
\end{aligned}
$$

where $\mathrm{E}[h_t]$ and $\operatorname{Cov}(h_t,h_s)$ are respectively given by Lemma B.1 and B.3. □

A.2 Unconditional Simulation under HN-GARCH

Algorithm B.5. *Unconditional Standardized GARCH Random Variates*

Assuming a known initial variance $h_{1,s}=h_1$, $s_{0,s}=0$ and S a given number of simulations, we do the following for $t=1,\ldots,T$ and $s=0,\ldots,S$,

1. *Draw a standard normal random variate $z_{t,s}$;*

2. *Compute $s_{t,s}$ and $h_{t+1,s}$ according to the GARCH recursion,*

$$
\begin{cases}
s_{t,s}=\mathscr{S}(s_{t-1,s},h_{t,s},z_{t,s})\\
h_{t+1,s}=\mathscr{H}(h_{t,s},z_{t,s})
\end{cases}
$$

3. *Standardize according to*

$$
\xi_{t,s}=\left(\frac{s_{t,s}}{\sqrt{Var[s_t]}},\frac{h_{t+1,s}}{\sqrt{Var[h_{t+1}]}}\right).
$$

where $Var[h_{t+1}]$ and $Var[s_t]$ are predetermined constants respectively given by (**??**) *and* (**??**).

A.3 Product Quantization under HN-GARCH

A.3.1 Log-Price Quantization

For typical GARCH specifications, $\mathscr{S}(s,h,z)$ is strictly monotonic in z such that we may map prices tiles to innovation tiles. We do so by implicitly defining $\mathscr{Z}_t^{(il)}(s)$ according to $s=\mathscr{S}(s_{t-1}^{(i)},h_t^{(l)},\mathscr{Z}_t^{(il)}(s))$ e.g. in the HN-GARCH case

$$\mathscr{Z}_t^{(il)}(s)=\frac{s-s_{t-1}^{(i)}-(\mu-1/2)h_t^{(l)}}{\sqrt{h_t^{(l)}}}.$$

Letting

$$\mathscr{I}_{tp}^{(il)}(s)=\int_{-\infty}^{\mathscr{Z}_t^{(il)}(s)}(\mathscr{S}(s_{t-1}^{(i)},h_t^{(l)},z'))^p\varphi(z')dz'$$

for $p=0,1,2$, e.g. in the HN-GARCH case

$$\begin{aligned}\mathscr{I}_{t0}^{(il)}(s)&=\Phi(\mathscr{Z}_t^{(il)}(s)),\\ \mathscr{I}_{t1}^{(il)}(s)&=\left(s_{t-1}^{(i)}+(\mu-1/2)h_t^{(l)}\right)\Phi(\mathscr{Z}_t^{(il)}(s))-\sqrt{h_t^{(l)}}\varphi(\mathscr{Z}_t^{(il)}(s))\\ \mathscr{I}_{t2}^{(il)}(s)&=\left(\left(s_{t-1}^{(i)}+(\mu-1/2)h_t^{(l)}\right)^2+h_t^{(l)}\right)\Phi(\mathscr{Z}_t^{(il)}(s))\\ &\quad-\left(\left(s_{t-1}^{(i)}+(\mu-1/2)h_t^{(l)}\right)\sqrt{h_t^{(l)}}+\mathscr{Z}_t^{(il)}(s)h_t^{(l)}\right)\varphi(\mathscr{Z}_t^{(il)}(s)),\end{aligned}$$

the distortion function to be minimized for log-prices is

$$D_{s_t}=\sum_{i=1}^{K}\sum_{l=1}^{N}p_{t-1}^{(il)}\sum_{j=1}^{K}\left[\mathscr{I}_{t2}^{(il)}-2\mathscr{I}_{t1}^{(il)}s_t^{(j)}+\mathscr{I}_{t0}^{(il)}(s_t^{(j)})^2\right]_{C_{s_t}^{(j)}}$$

where use the notation $[\mathscr{I}]_C=\mathscr{I}(b)-\mathscr{I}(a)$ for $\mathscr{I}:\mathbb{R}\mapsto\mathbb{R}$ with $C=(b,a)$ an open interval over $\mathbb{R}$ and $[\mathscr{I}]_\emptyset=0$ by convention.

Proposition B.3. *For $k=1,\dots,K$, $i=1,\dots,K$, $l=1,\dots,N$ and $p=0,1,2$,*

$$\frac{d}{ds_t^{(k)}}\sum_{j=1}^{K}\left[\mathscr{I}_{tp}^{(il)}\right]_{C_{s_t}^{(j)}}=0$$

Proof. Letting

$$\underline{s}_t^{(i)}=(s_t^{(i-1)}+s_t^{(i)})/2,\quad \bar{s}_t^{(i)}=(s_t^{(i)}+s_t^{(i+1)})/2,$$

$$\underline{h}_{t+1}^{(l)}=(h_{t+1}^{(l-1)}+h_{t+1}^{(l)})/2,\quad \bar{h}_{t+1}^{(l)}=(h_{t+1}^{(l)}+h_{t+1}^{(l+1)})/2,$$

for $i = 1, \ldots, K$ and $l = 1, \ldots, N$, where $s^{(0)} = -\infty$ and $s^{(K+1)} = \infty$ and $h_t^{(0)} = 0$ and $h_t^{(N+1)} = \infty$, we preliminarily note using the chain rule

$$\frac{d}{ds_t^{(k)}} \mathscr{I}_{tp}^{(il)} \circ \underline{s}_t^{(j)} = (\underline{s}_t^{(j)})^p \varphi(\mathscr{Z}_t^{(il)}(\underline{s}_t^{(j)})) \left(\frac{d\mathscr{Z}_t^{(il)}}{ds} \circ \underline{s}_t^{(j)} \right) \frac{d\underline{s}_t^{(j)}}{ds_t^{(k)}}.$$

where $d\underline{s}_t^{(j)} / ds_t^{(k)} = \frac{1}{2}$ for $k = j-1, j$ and 0 otherwise. We find a similar expression for the derivative of $\mathscr{I}_{tp}^{(il)} \circ \bar{s}_t^{(j)}$ with respect to $s_t^{(k)}$ which is non-null for $k = j, j+1$. Defining

$$\delta_{tp}^{(il)}(s) = \frac{1}{2} s^p \varphi(\mathscr{Z}_t^{(il)}(s)) \frac{d\mathscr{Z}_t^{(il)}}{ds}$$

for notational convenience, we hence have

$$\begin{aligned}
\frac{d}{ds_t^{(j-1)}} \left[\mathscr{I}_{tp}^{(il)} \right]_{C_{s_t}^{(j)}} &= -\delta_{tp}^{(il)}(\underline{s}_t^{(j)}), \\
\frac{d}{ds_t^{(j)}} \left[\mathscr{I}_{tp}^{(il)} \right]_{C_{s_t}^{(j)}} &= \delta_{tp}^{(il)}(\bar{s}_t^{(j)}) - \delta_{tp}^{(il)}(\underline{s}_t^{(j)}), \\
\frac{d}{ds_t^{(j+1)}} \left[\mathscr{I}_{tp}^{(il)} \right]_{C_{s_t}^{(j)}} &= \delta_{tp}^{(il)}(\bar{s}_t^{(j)}),
\end{aligned}$$

and

$$\frac{d \left[\mathscr{I}_{tp}^{(il)} \right]_{C_{s_t}^{(j)}}}{ds_t^{(k)}} = 0$$

otherwise i.e. for $k \neq j-1, j, j+1$. Finally,

$$\frac{d}{ds^{(k)}} \sum_{j=1}^{K} \left[\mathscr{I}_{tp}^{(il)} \right]_{C_{s_t}^{(j)}} = -\delta_{tp}^{(il)}(\underline{s}_t^{(k+1)}) + \delta_{tp}^{(il)}(\bar{s}_t^{(k)}) - \delta_{tp}^{(il)}(\underline{s}_t^{(k)}) + \delta_{tp}^{(il)}(\bar{s}^{(k-1)}) = 0$$

using the fact $\underline{s}^{(k+1)} = \bar{s}^{(k)}$ and $\bar{s}^{(k-1)} = \underline{s}^{(k)}$. □

As a direct consequence of Proposition B.3, the corresponding gradient is given by

$$\nabla_j D_{s_t} = \frac{dD_{s_t}}{ds_t^{(j)}} = -2 \sum_{i=1}^{K} \sum_{l=1}^{N} p_{t-1}^{(il)} \left[\mathscr{I}_{t1}^{(il)} - \mathscr{I}_{t0}^{(il)} s_t^{(j)} \right]_{C_{s_t}^{(j)}}, \tag{A.1}$$

which may be obtained at virtually no additional cost following the computation of distortions.

A.3.2 Conditional Variance Quantization

For typical GARCH models, $\mathscr{H}(h,z)$ is quadratic in innovations z and has up to two real roots. We first let $\mathscr{X}_{t\pm}^{(l)}(h)$ the real part of roots of function $h(x)=\mathscr{H}(h_t^{(l)},x)$ e.g. in the HN-GARCH case

$$\mathscr{X}_{t\pm}^{(l)}(h)=\gamma\sqrt{h_t^{(l)}}\pm\sqrt{\max\left(\frac{h-\omega-\beta h_t^{(l)}}{\alpha},0\right)}.$$

We also let

$$\mathscr{L}_{tp}^{(l)}(h)=\left(\int_{-\infty}^{\mathscr{X}_{t+}^{(l)}(h)}-\int_{-\infty}^{\mathscr{X}_{t-}^{(l)}(h)}\right)(\mathscr{H}(h_t^{(l)},z'))^p d\varphi(z')$$

for $p=0,1,2$ with indefinite integrals in the HN-GARCH case obtained by elementary integration,

$$\begin{aligned}\int\mathscr{H}(h,z')d\varphi(z')&=\int\left(\omega+(\beta+\alpha\gamma^2)h-2\alpha\gamma\sqrt{h}z+\alpha z^2\right)\varphi(z)dz\\&=\left(\omega+(\beta+\alpha\gamma^2)h+\alpha\right)\Phi(z)+\alpha\left(2\gamma\sqrt{h}-z\right)\varphi(z)\end{aligned}$$

and

$$\begin{aligned}\int(\mathscr{H}(h,z'))^2d\varphi(z')&=\int\left(\omega+(\beta+\alpha\gamma^2)h-2\alpha\gamma\sqrt{h}z+\alpha z^2\right)^2\varphi(z)dz\\&=\left(\alpha^2(\gamma^2h(\gamma^2h+6)+3)+2\alpha(\gamma^2h+1)(\beta h+\omega)+(\beta h+\omega)^2\right)\Phi(z)\\&\quad+\alpha\Bigg(\alpha\left(2\gamma^2h(2\gamma\sqrt{h}-3z)+4\gamma\sqrt{h}(z^2+2)-z(z^2+3)\right)\\&\quad+2(2\gamma\sqrt{h}-z)(\beta h+\omega)\Bigg)\varphi(z).\end{aligned}$$

The distortion to be minimized for conditional variances may hence be explicitly written as

$$D_{h_{t+1}}=\sum_{i=1}^{K}\sum_{l=1}^{N}p_{t-1}^{(il)}\sum_{m=1}^{N}\left[\mathscr{L}_{t2}^{(l)}-2\mathscr{L}_{t1}^{(l)}h_{t+1}^{(m)}+\mathscr{L}_{t0}^{(l)}(h_{t+1}^{(m)})^2\right]_{C_{h_{t+1}}^{(m)}}\tag{A.2}$$

where we use the notation for $[\mathscr{L}]_C$ introduced in Appendix A.3.1.

Proposition B.4. *For $l=1,\dots,N$, $n=1,\dots,N$ and $p=0,1,2$,*

$$\frac{d}{dh_t^{(n)}}\sum_{m=1}^{K}\left[\mathscr{L}_{tp}^{(l)}\right]_{C_{h_{t+1}}^{(m)}}=0$$

The proof is similar to the proof of Proposition B.3 and is omitted for brevity.

As a direct consequence of Proposition B.4, the gradient is given by

$$\nabla_m D_{h_{t+1}} = \frac{dD_{h_{t+1}}}{dh^{(m)}_{t+1}} = -2\sum_{i=1}^{K}\sum_{l=1}^{N} p^{(il)}_{t-1}\left[\mathscr{L}^{(l)}_{t1} - \mathscr{L}^{(l)}_{t0} h^{(m)}_{t+1}\right]_{C^{(m)}_{h_{t+1}}}, \tag{A.3}$$

which may be computed at virtually no additional cost.

A.3.3 Transition Probabilities

Using notations previously introduced in Appendices A.3.1 and A.3.2, transition probabilities may be written as

$$p_t^{(il,jm)} = [\Phi]_{\mathscr{Z}_t^{(il)}(C^{(j)}_{s_t})\cap\mathscr{X}^{(l)}_{t-}(C^{(m)}_{h_{t+1}})} + [\Phi]_{\mathscr{Z}_t^{(il)}(C^{(j)}_{s_t})\cap\mathscr{X}^{(l)}_{t+}(C^{(m)}_{h_{t+1}})}$$

where the first(second) term captures the probability of a transition arising from low(high) innovations i.e. from the left(right) branch of the parabolic support of the price-variance distribution. While only one of these terms is non-null for most (il, jm), we must solve both set intersections in practice. These set operations remain computationally negligible with respect to normal CDF evaluations.

A.4 Conditional Quantization under HN-GARCH

The variance distortion conditional on the event $\left\{\tilde{s}_t \in C^{(j)}_{s_t}\right\}$ is

$$D^{(j)}_{h_{t+1}} = \frac{1}{d^{(j)}_t}\sum_{i=1}^{K}\sum_{l=1}^{N} p^{(il)}_{t-1}\int_{-\infty}^{\infty} \mathrm{I}(\mathscr{S}(s^{(i)}_{t-1}, h^{(il)}_t, z) \in C^{(j)}_{s_t})$$
$$\sum_{m=1}^{N} \mathrm{I}(\mathscr{H}(h^{(il)}_t, z) \in C^{(jm)}_{h_{t+1}})(\mathscr{H}(h^{(il)}_t, z) - h^{(jm)}_{t+1})^2 d\varphi(z)$$

where

$$d^{(j)}_t = \sum_{i=1}^{K}\sum_{l=1}^{N} p^{(il)}_{t-1}\int_{-\infty}^{\infty} \mathrm{I}(\mathscr{S}(s^{(i)}_{t-1}, h^{(il)}_t, z) \in C^{(j)}_{s_t}) d\varphi(z)$$

is a normalization constant, intuitively interpreted as the unconditional probability of selecting the j-th conditional variance quantizer $\Gamma^{(j)}_{h_{t+1}}$.

We proceed as previously for product quantization by introducing $\mathscr{X}_{t\pm}^{(il)}(h)$ the real part of roots of $h(x) = \mathscr{H}(h_t^{(il)}, x)$. We follow the left-right branch decomposition of Appendix A.3.3 and define *definite* integrals of interest,

$$M_{tp}^{(il,jm)} = \left(\int_{\mathscr{Z}_t^{(il)}(C_{s_t}^{(j)}) \cap \mathscr{X}_{t-}^{(il)}(C_{h_{t+1}}^{(jm)})} + \int_{\mathscr{Z}_t^{(il)}(C_{s_t}^{(j)}) \cap \mathscr{X}_{t+}^{(il)}(C_{h_{t+1}}^{(jm)})} \right) (\mathscr{H}(h_t^{(il)}, z'))^p d\varphi(z');$$

see Appendix A.3.2 for explicit expressions and corresponding indefinite integrals in the HN-GARCH case. In practice, set intersections must be solved prior to evaluating integrals, which as previously noted is computationally negligible with regards to CDF evaluations. The K distortion functions to be minimized may finally be written as

$$D_{h_{t+1}}^{(j)} = \frac{1}{d_t^{(j)}} \sum_{i=1}^{K} \sum_{l=1}^{N} p_{t-1}^{(il)} \sum_{m=1}^{N} \left(M_{t2}^{(il,jm)} - 2M_{t1}^{(il,jm)} h_{t+1}^{(jm)} + M_{t0}^{(il,jm)} (h_{t+1}^{(jm)})^2 \right) \tag{A.4}$$

for $j = 1, \ldots, K$ where we may further solve

$$d_t^{(j)} = \sum_{i=1}^{K} \sum_{l=1}^{N} p_{t-1}^{(il)} \left[\mathscr{I}_{t0}^{(il)} \right]_{C_{s_t}^{(j)}}.$$

This is typically unnecessary in practice.

A result similar to Proposition B.4 holds here such that the gradient is[1]

$$\nabla_m D_{h_{t+1}}^{(j)} = \frac{dD_{h_{t+1}}^{(j)}}{dh_{t+1}^{(jm)}} = \frac{-2}{d_t^{(j)}} \sum_{i=1}^{K} \sum_{l=1}^{N} p_{t-1}^{(il)} (L_{t1}^{(il,jm)} - L_{t0}^{(il,jm)} h_{t+1}^{(jm)}) \tag{A.5}$$

Finally, transition probabilities are obtained as a by-product since we have

$$p_t^{(il,jm)} = \int_{-\infty}^{\infty} \mathrm{I}(\mathscr{S}(s_{t-1}^{(i)}, h_t^{(il)}, z) \in C_{s_t}^{(j)}) \mathrm{I}(\mathscr{H}(h_t^{(il)}, z) \in C_{h_{t+1}}^{(jm)}) d\varphi(z) = M_{t0}^{(il,jm)}.$$

A.5 Numerical Implementation

The strict ordering of quantizers is likely to be violated during a gradient descent. We are also concerned with tiles stalling over null probability areas (e.g. when conditional

[1] More precisely, we can show

$$\frac{d}{dh_t^{(jn)}} \sum_{m=1}^{K} M_{tp}^{(il,jm)} = 0$$

for $i = 1, \ldots, K$, $l = 1, \ldots, N$, $j = 1, \ldots, K$, $n = 1, \ldots, N$ and $p = 0, 1, 2$.

variance quantizers approach zero) where the distortion function is flat. We propose a change of variable that overcomes these issues.

Letting Γ either an asset price (i.e. Γ_{s_t}) or a conditional variance quantizer (i.e. either $\Gamma_{h_{t+1}}$ under product quantization or $\Gamma^{(j)}_{h_{t+1}}$ under conditional quantization). Here, P refers to the size of Γ, namely K for log-price quantizers and N for conditional variance quantizers. We optimize over $\lambda \in \mathbb{R}^P$ implicitly defined by $\Gamma = \Psi(\lambda; a, b, c)$ with m-th component given by $\psi^{(m)} = a + (b-a)\ell(c\theta^{(m)})$ for $m = 1, \ldots, P$ where $\ell(y) = (1+e^{-y})^{-1}$ is the logistic function, $\theta^{(m)} = -\frac{1}{2} + \sum_{n=1}^{m} \frac{e^{\lambda^{(n)}}}{P}$ and a, b, $c \in \mathbb{R}$ are predetermined parameters.

The proposed re-parameterization Ψ maps $\mathbb{R}^P$ to the space of valid (i.e. with increasing components) P-quantizers bounded by (a,b) i.e. $a < \psi^{(m)} < b$ for $m = 1, \ldots, P$. While quantizers are theoretically unbounded, limiting gradient descents to areas of significant probability improves convergence. For example, we let (a,b) be

$$\left(\min_{\{i,l\}} \mathscr{S}(s^{(i)}_{t-1}, h^{(il)}_t, -d), \quad \max_{\{i,l\}} \mathscr{S}(s^{(i)}_{t-1}, h^{(il)}_t, d) \right)$$

for log-price quantizers and similarly

$$\left(\min_{\{i,l\}} \min_{z \in (-d,d)} \mathscr{H}(h^{(il)}_t, z), \quad \max_{\{i,l\}} \max_{z \in (-d,d)} \mathscr{H}(h^{(il)}_t, z) \right)$$

for conditional variance quantizers (i.e. when $\Gamma = \Gamma_{h_{t+1}}$) where $d > 0$ is a truncation parameter. Under conditional quantization (i.e. when $\Gamma = \Gamma^{(j)}_{h_{t+1}}$), it makes sense to let (a,b) be

$$\left(\min_{\{i,l\}} \min_{z \in Y^{(il,j)}_t} \mathscr{H}(h^{(il)}_t, z), \quad \max_{\{i,l\}} \max_{z \in Y^{(il,j)}_t} \mathscr{H}(h^{(il)}_t, z) \right)$$

where $Y^{(il,j)}_t = Z^{(il)}_t(C^{(j)}_{s_t}) \cap (-d,d)$ and $\min_{\emptyset} = \emptyset$ and $\max_{\emptyset} = \emptyset$ by convention. Overall, we found that setting $c = 50$ and $d = 10$ provides satisfactory results.

The gradient of a distortion D (either D_{s_t}, $D_{h_{t+1}}$ or $D^{(j)}_{h_{t+1}}$) as a function of λ is given by

$$\widetilde{\nabla}_n D = \frac{dD}{d\lambda^{(n)}} = \sum_{m=1}^{P} (\nabla_m D)(\nabla_{mn} \Psi)$$

where the Jacobian of Ψ is

$$\nabla_{mn}\Psi = \frac{d\psi^{(m)}}{d\lambda^{(n)}} = \mathrm{I}(m \geq n)(b-a)\frac{ce^{\lambda^{(n)}}}{P}\ell'(c\theta^{(m)})$$

and $\nabla_n D$ corresponds to one of the gradients given in either Appendix A.3 or A.4 and $\ell'(x) = \ell(x)(1-\ell(x))$.

A.6 Delta-Hedging

Proposition B.5.

$$\partial C_t / \partial \tilde{S}_t = -E[e^{s_T - s_t} I(s_T \leq k) | \mathscr{F}_t]$$

where $C_t = S_0 e^{rt-\delta T} E[(e^k - e^{s_T})^+ | \mathscr{F}_t]$ *and* $\tilde{S}_t = e^{-\delta(T-t)} S_t$.

Proof. Since $e^{s_t} = S_t / S_0 e^{-(r-\delta)t}$ by definition,

$$\begin{aligned}\frac{\partial C_t}{\partial \tilde{S}_t} &= -S_0 e^{rt-\delta T} \frac{\partial}{\partial \tilde{S}_t} e^{s_t} \mathrm{E}[e^{s_T - s_t} \mathrm{I}(s_t \leq k) | \mathscr{F}_t] \\ &= -S_0 e^{rt-\delta T} \frac{\partial}{\partial \tilde{S}_t} \frac{S_t}{S_0} e^{-(r-\delta)t} \mathrm{E}[e^{s_T - s_t} \mathrm{I}(s_t \leq k) | \mathscr{F}_t] \\ &= -e^{-\delta(T-t)} \frac{\partial}{\partial \tilde{S}_t} \tilde{S}_t e^{\delta(T-t)} \mathrm{E}[e^{s_T - s_t} \mathrm{I}(s_t \leq k) | \mathscr{F}_t] \\ &= -\mathrm{E}[e^{s_T - s_t} \mathrm{I}(s_T \leq k) | \mathscr{F}_t]\end{aligned}$$

where the last step is justified by the *dominated convergence theorem* which relies on some technical conditions met by the HN-GARCH model and vanilla payoffs; see e.g. Section 7.2.2 of Glasserman (2003). □

Proposition B.6.

$$\Delta_t^{DH}(s_{t-1}, h_t) = -\frac{1}{2} + \frac{1}{\pi e^{s_t}} \int_0^\infty Re\left[\frac{e^{-i\phi k} f_{t-1}(i\phi + 1)}{i\phi}\right] d\phi$$

where $f_t(\phi) = E[e^{\phi s_T} | \mathscr{F}_t] = \exp(\phi s_t + A_t(\phi) + B_t(\phi) h_{t+1})$ *with* $A_t()$ *and* $B_t()$ *found in Heston and Nandi (2000) (with their r set to zero).*

Proof. From the usual decomposition $c_t = \mathrm{E}[(e^{s_T} - e^k)^+] = e^{s_t} P_1() - e^k P_2()$ of Eq. (10) of Heston and Nandi (2000), we have for *call* options

$$\mathrm{E}[e^{s_T - s_t} \mathrm{I}(s_T \geq k) | \mathscr{F}_t] = \frac{1}{2} + \frac{1}{\pi e^{s_t}} \int_0^\infty \mathrm{Re}\left[\frac{e^{-i\phi k} f_t(i\phi + 1)}{i\phi}\right] d\phi.$$

The result for *put* options follows from put-call parity $e^{s_t} - e^k = c_{t,call} - c_{t,put}$ by taking the partial derivative with respect to e^{s_t}. □

A.7 Supplementary Figures

Figure B.1: Box-plot of RMS at 21 trading days for 1,000 runs of the splitting method (Sub-Algorithm 3) followed by 10 iterations of Lloyd's method (Sub-Algorithm 2) for varying degrees of precision $\delta^\star$ with $P = 100$. Optimal quantizers are obtained using unconditional random variates computed according to Algorithm B.5. White boxes [Precision+Sampling] capture both RMS sampling errors and finite precision errors from the CLVQ algorithm, while black boxes [Sampling] capture RMS sampling errors only (for a *single* randomly selected optimal quantizer). More precisely, white boxes are obtained by first computing 1,000 optimal quantizers and then computing a single RMS per optimal quantizer. Black boxes are obtained by first randomly selecting a single optimal quantizer and then computing 1,000 RMS.

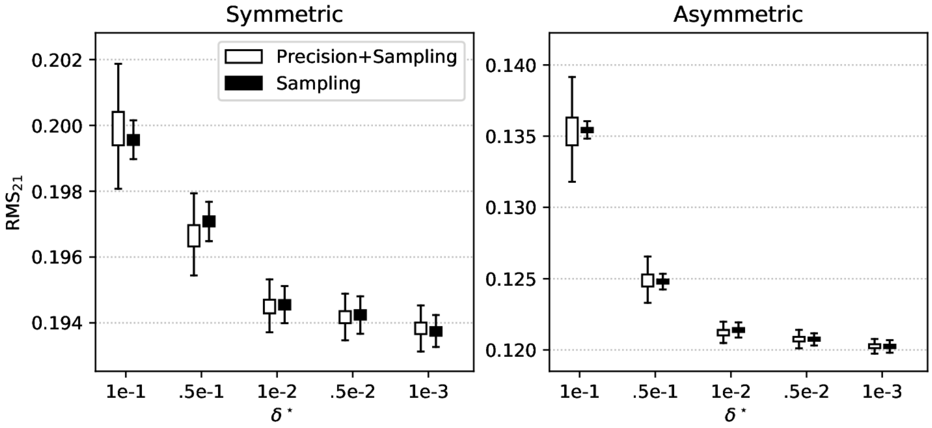


Figure B.2: RMS for optimal quantizations obtained via marginal quantization (Algorithm 1) with $P = 1681$.

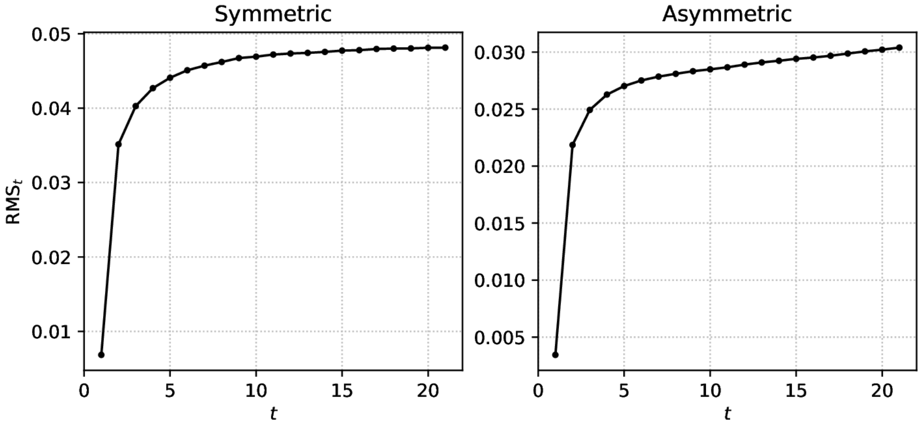


Figure B.3: Excess RMS from recursive quantization in time, i.e. $\text{RMS} - \text{RMS}^\star$ where $\text{RMS}^\star$ is a quantization error under the marginal approach (see Algorithm 1). Optimal recursive quantizations are obtained via Algorithms 2 [Markov], 3 [Product] and 4 [Conditional]. We use $P = 1681$ for both marginal and Markovian stochastic approaches, $(K = 41, N = 41)$ for both product and conditional quantization under the symmetric case and $(K = 58, N = 29)$ under the asymmetric case for conditional quantization.

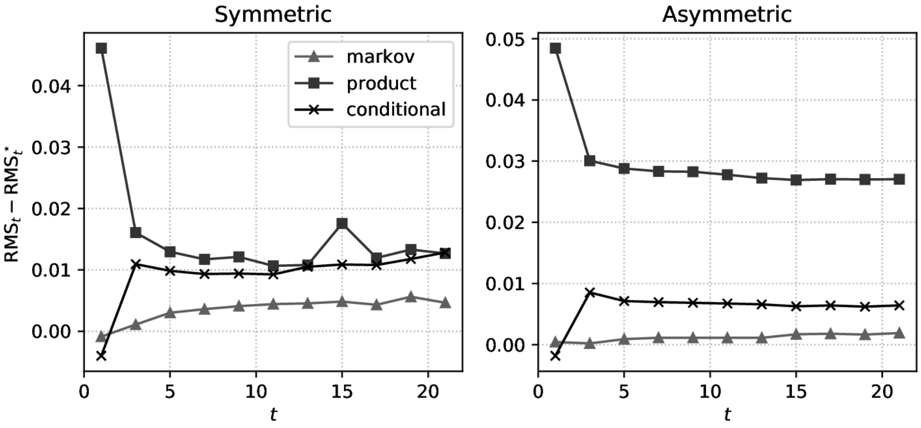


Figure B.4: Optimal quantizations at $t = 1, 5, 10, 21$ obtained via marginal quantization (Algorithm 1) with $P = 50$ and $\delta^{\star} = 1\mathrm{e}{-6}$.

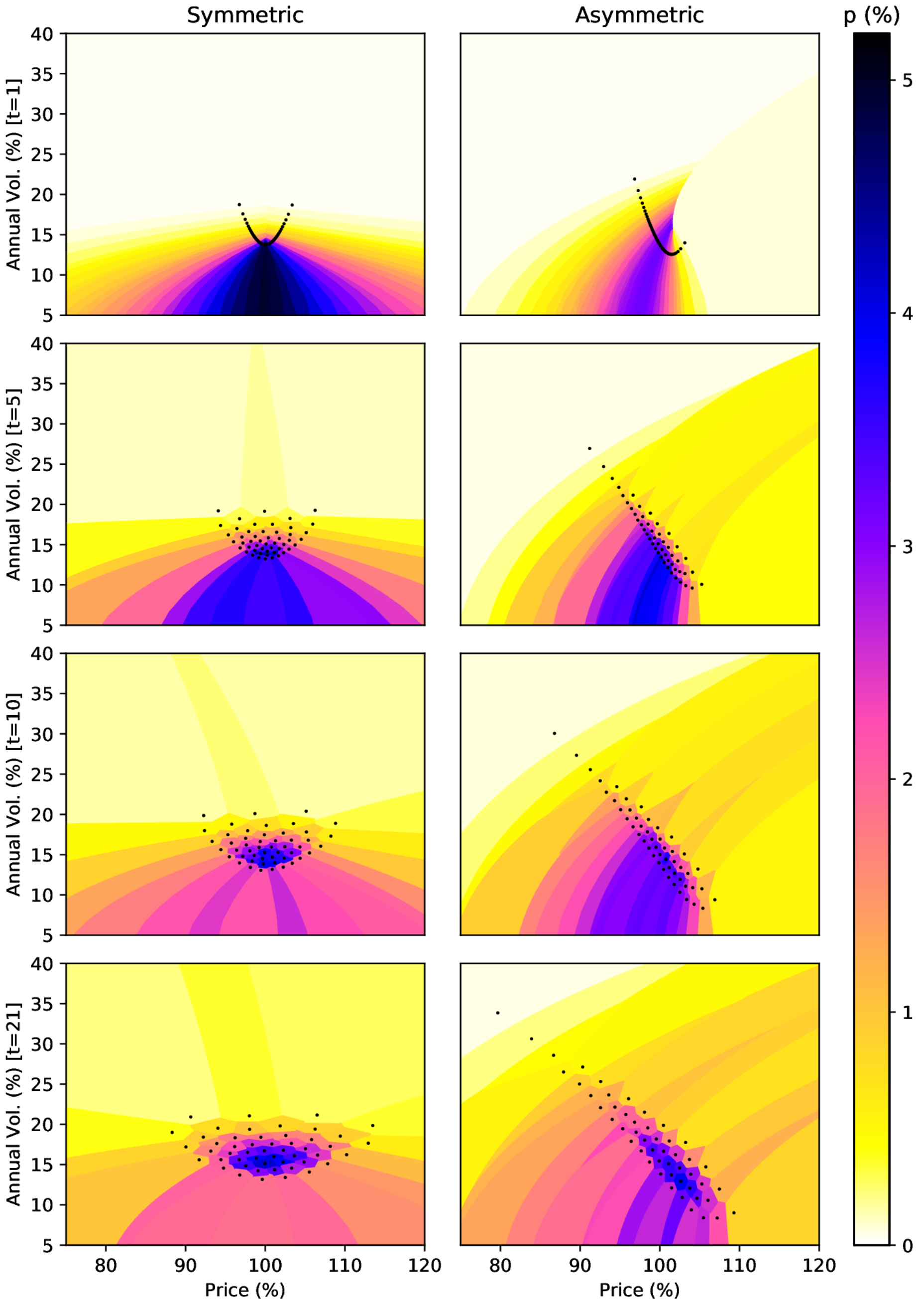
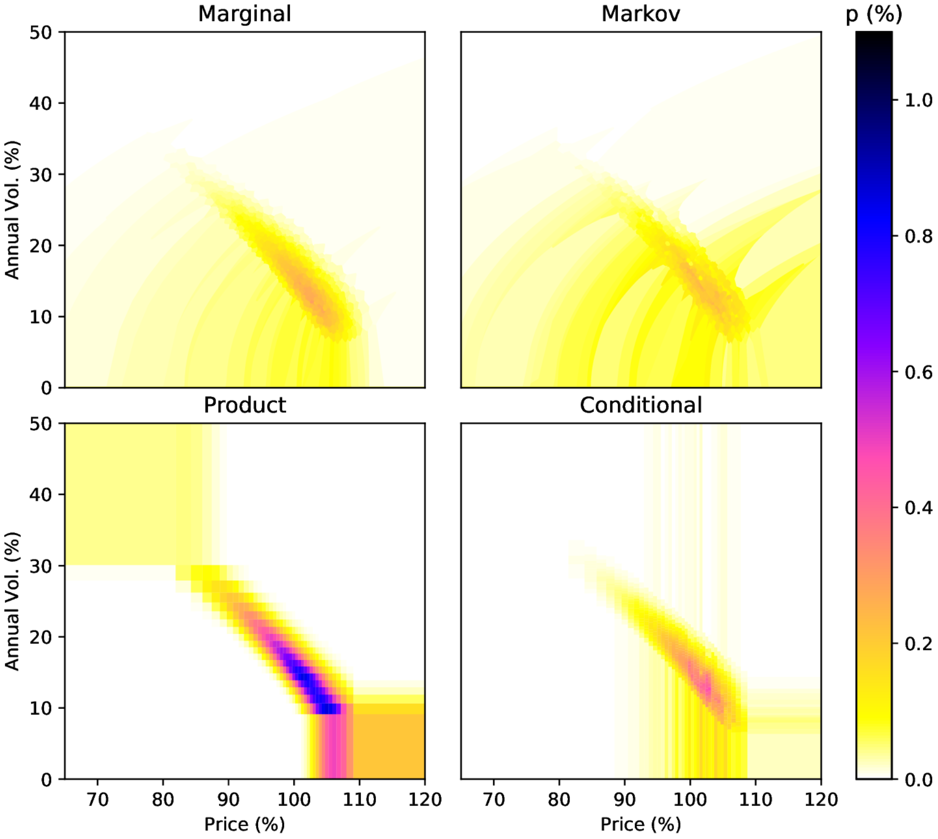


Figure B.5: Comparison of all optimal quantizations, i.e. obtained via Algorithms 1 [Marginal], 2 [Markov], 3 [Product] and 4 [Conditional] in the asymmetric case at 21 trading days for $P = 800$ and ($K = 40$, $N = 20$). Quantizer elements are not shown for clarity.



References

Christoffersen, P., Heston, S., and Jacobs, K. (2013). Capturing option anomalies with a variance-dependent pricing kernel. *Review of Financial Studies*, 26(8):1963–2006.

Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York.

Heston, S. L. and Nandi, S. (2000). A closed-form GARCH option valuation model. *Review of Financial Studies*, 13(3):585–625.