Implied volatility surface predictability: The case of commodity markets

Fearghal Kearney [a,*], Han Lin Shang [b], Lisa Sheenan [a]

[a] Queen's Management School, Queen's University Belfast, BT9 5EE, UK
[b] Research School of Finance, Actuarial Studies and Statistics, Australian National University, Canberra ACT 2601, Australia

ARTICLE INFO

ABSTRACT

Recent literature seek to forecast implied volatility derived from equity, index, foreign exchange, and interest rate options using latent factor and parametric frameworks. Motivated by increased public attention borne out of the financialization of futures markets in the early 2000s, we investigate if these extant models can uncover predictable patterns in the implied volatility surfaces of the most actively traded commodity options between 2006 and 2016. Adopting a rolling out-of-sample forecasting framework that addresses the common multiple comparisons problem, we establish that, for energy and precious metals options, explicitly modeling the term structure of implied volatility using the Nelson-Siegel factors produces the most accurate forecasts.

© 2019 Published by Elsevier B.V.

1. Introduction

When using the Black and Scholes (1973) model to price options, the only variable not known with certainty is volatility. The estimated future volatility backed out of these option prices is referred to as implied volatility (IV). This can be plotted against both moneyness and time-to-maturity to produce an implied volatility surface (IVS). While Black and Scholes (1973) assume that the IVS is flat, this is not observed empirically, as option contracts of varying maturity and moneyness levels tend to be priced according to different levels of IV. The shape of these IVSs evolves, and so accurate modeling of their inter-temporal dynamics is required (Deuskar et al., 2008; Konstantinidi et al., 2008; Bedendo and Hodges, 2009). Bedendo and Hodges (2009) state that producing reliable forecasts for the evolution of IV is essential to indicate prevailing market conditions as well as to facilitate efficient option portfolio risk management. Research to date focuses on modeling and forecasting equity, index, interest rate, and foreign exchange options, while uncovering predictable dynamics in commodity IV

has yet to be explored in the literature. We aim to address this research gap.

Irwin and Sanders (2011) report that flows into commodity investments increased from $15 billion in 2003 to $250 billion in 2009. Fattouh and Mahadeva (2014) cite this *financialization* of futures markets as the starting point for increased public interest in commodities.[1] Adams and Glück (2015) conclude that the entrance of these new types of investors significantly influenced the behavior of commodities in financial markets and how commodities linked to other assets. More specifically, as commodities became a significant part of financial investors' portfolios, they were treated as a new category within the universe of stocks, becoming part of a more general equity style. A similar view is put forward by Cheng and Xiong (2014), who postulate that commodity futures are treated as an asset class just like stocks and bonds.[2] Given this interconnectedness between commodities and other asset classes, we investigate how readily extendable to commodity markets the

---

* Corresponding author.
  E-mail addresses: f.kearney@qub.ac.uk, fkearney@fkearney.ie (F. Kearney), hanlin.shang@anu.edu.au (H.L. Shang), l.sheenan@qub.ac.uk (L. Sheenan).

[1] Financialization is the term commonly used for the phenomenon in the early 2000s whereby large inflows into commodity investments occurred. Domanski and Heath (2007) attribute this to institutional investors who were historically not engaged in commodity investing on such a large scale.

[2] This recent shift in how commodities are considered represents a stark departure from the description of commodity returns as being weakly correlated with the stock market (Bessembinder and Chan, 1992; Gorton and Rouwenhorst, 2006).

methods popular for modeling and forecasting the IVSs of other assets are. Specifically, we determine how accurately the existing frameworks characterize the shape of the commodity IVS and how well they uncover predictable patterns and dynamics unique to these underlying assets.

Related literature highlights a multitude of reasons why forecasting IV is important for asset pricing and risk management. As well as being a transformation of the option price, and a key parameter in many asset pricing formulae (Giot, 2003), IV is of interest because of its use in forecasting realized volatility (see Corrado and Miller, 2006; Taylor et al., 2010; Muzzioli, 2010; Garvey and Gallagher, 2012). Giot (2003) also demonstrate the high informational content of IV by successfully adopting it in a value-at-risk modeling framework for agricultural futures. Bernales and Guidolin (2014) posit that understanding the dynamics of the IVS can inform traders regarding the design of speculative or hedging strategies. Knowledge of the dynamic process of the IVS is also relevant for investment decisions in other markets because options are commonly used to obtain forward-looking market information. Such forward-looking analyses rely on the assumption that option prices reveal agents' expectations about prospective economic scenarios, where investor forecast horizons correspond to the expiry dates of traded options contracts.

A variety of frameworks has been proposed to model and forecast IV. We split them into four broad classes: general equilibrium, principal component (PC), machine learning, and parametric modeling. Turning first to general equilibrium models David and Veronesi (2000); Guidolin and Timmermann (2003); Garcia et al. (2003); Hibbert et al. (2008) and Bernales and Guidolin (2015) develop rational asset pricing models that theoretically calibrate to observed asymmetric IVS shape and its evolution over time. The underlying idea is that investors' uncertainty about economic fundamentals (e.g., dividends) affects stochastic volatility and leverage behind the IVS. This uncertainty is said to evolve. However, as outlined by Bedendo and Hodges (2009), these general equilibrium models suffer from tractability issues and are not appropriate for risk management or forecasting purposes.

An alternative approach is to use a purely data-driven method that describes the empirical shape of the IVS through the use of unobservable latent statistical factors, as in Skiadopoulos et al. (2000); Alexander (2001); Chalamandaris and Tsekrekos (2010). Chalamandaris and Tsekrekos (2010) exploit PC factors further by employing them to produce out-of-sample forecasts of the IVS. However, these forecasts are produced using a regular-sized daily IVS dataset with precisely the same moneyness and maturity observations available on each observation date. Such an approach is not possible using our trade-level commodity dataset in which contracts with the same moneyness and maturity are not traded every day.

A third related methodology is that of machine learning and non-parametric techniques to model the IVSs, as in Cont and Da Fonseca (2002); Fengler et al. (2003, 2007); Audrino and Colangelo (2010); Fengler and Hin (2015). The primary intuition behind these models is that, instead of estimating a highly parameterized model, it is better to use non- or semi-parametric techniques with the amount of complexity controlled to avoid overfitting. However, of the studies cited above, only Audrino and Colangelo (2010) produce out-of-sample forecasts of the IVS in line with the aim of our paper. They utilize regression trees informed by a cross-validation strategy to produce forecasts of the surface, an approach that we also follow in our out-of-sample analysis.

A fourth popular methodology involves the use of deterministic parametric specifications based on the cross-section of options available at any one point in time. The specifications link the maturity and moneyness of option contracts to the shape of the IVS, with Dumas et al. (1998) and Peña et al. (1999) the most

popular examples. Goncalves and Guidolin (2006) use a two-stage framework to forecast Dumas et al. (1998) cross-sectional maturity and moneyness coefficients. Further, Chalamandaris and Tsekrekos (2011) extend their work by explicitly modeling the IVS term structure component by incorporating Nelson and Siegel (1987) factors.

These forecasting frameworks have been applied across a wide number of markets including equities, foreign exchange, and interest rates. For example, Goncalves and Guidolin (2006) describe the daily IVS of S&P500 Index options using Dumas et al. (1998) parametric specifications. They interpret the daily estimated parameter coefficients as proxies for factors that drive the evolution of the IVS and empirically demonstrate that modeling the time series dynamics can identify predictable IVS components. Bedendo and Hodges (2009) also focus on S&P500 Index options, instead of implementing parameter estimation using a Kalman filter. However, they focus solely on information from an isolated IV smile and do not incorporate information from the full surface as we seek to do. Fengler et al. (2007), also model equity index options, namely the German stock index (DAX) using their semi-parametric factor model. After selecting an appropriate model size and bandwidth, it is proposed that the estimated factors series could follow a vector autoregressive (VAR) model. However, no direct out-of-sample forecasts of the IV surface are produced. Chalamandaris and Tsekrekos (2010, 2011, 2014) examine foreign exchange markets. For instance, Chalamandaris and Tsekrekos (2014) use an extensive time series of over-the-counter (OTC) options for eight currencies versus the Euro. They find that in medium- to long-term forecasts the shape and dynamics of the IVS can be forecast successfully using the aforementioned latent factor and parametric models. Finally, Deuskar et al. (2008) analyze the interest rate options market, finding that the shape of the IV smile for Euro interest rate caps and floors is affected dynamically by the yield curve and by future uncertainty in interest rate markets.[3][4]

There is a growing body of evidence for predictable IVS movements. Predicting the entire IVS, and thus all future option prices, would directly contradict the efficient market hypothesis. However, uncovering predictability in isolated segments of a commodity IVS may signal the existence of important pockets of market inefficiency. Chalamandaris and Tsekrekos (2014) conclude that non-uniform trading across the IVS may lead to some segments exhibiting more predictability than others as they adjust to information at different speeds. Further, Chalamandaris and Tsekrekos (2011) argue that over longer horizons, the shape of the IVS term structure contains predictive information, whereby forward IVs may be utilized to forecast future implied volatilities of nearer expiries. If the efficient market hypothesis is imposed, this IVS predictability may be traced back to either micro-structural imperfections or unobservable and hard-to-estimate time-varying risk premium.

As in Chalamandaris and Tsekrekos (2014) we choose popular models from the literature: the two-step approach from Goncalves and Guidolin (2006), and the Diebold and Li (2006) based approach of Chalamandaris and Tsekrekos (2011). We apply these frameworks along with a regression tree benchmark model from Audrino and Colangelo (2010), to a representative sample of the most actively traded commodity options contracts: Cocoa, Corn,

---

[3] Deuskar et al. (2008) also highlight that conclusions from equity markets do not necessarily transfer to interest rate options as they are predominantly traded OTC by asking side institutional investors. However, such constraints do not apply to commodity options as futures contracts are the most active avenue via which investors take positions in the market.

[4] In a related study of IV indices by Badshah et al. (2013), casual contemporaneous bidirectional spill-over between gold and exchange rates are uncovered. This is a tentative indication that frameworks adopted to model the dynamics of foreign exchange markets may also apply to Precious Metal commodities.

Table 1
Data description.

Ticker	Commodity (Exchange)	Mean (%)	Median (%)	Std. Dev. (%)	1st Quart. (%)	3rd Quart. (%)	Skewness	Longest Mat.	Avg. Convenience Yield Slope
CC	Cocoa (ICE)	26.76	24.38	6.79	21.66	31.74	80.43	6 Months	1.08
C	Corn (CBOT)	29.35	28.08	7.12	23.89	34.22	57.16	12 Months	0.31
CT	Cotton (ICE)	27.24	24.15	9.15	20.93	30.00	166.07	12 Months	-0.55
S	Soybeans (CBOT)	23.56	21.87	6.63	19.33	25.75	146.45	12 Months	-5.21
BO	Soybean Oil (CBOT)	22.30	21.05	5.06	18.98	23.72	164.42	6 Months	0.15
SB	Sugar (ICE)	32.36	30.39	10.57	24.52	39.56	58.35	6 Months	-0.05
W	Wheat (CBOT)	30.59	28.87	7.54	24.87	35.80	79.51	12 Months	3.71
CL	Crude Oil (NYMEX)	31.48	30.24	10.70	24.99	36.96	121.20	24 Months	0.12
HO	Heating Oil (NYMEX)	27.71	28.09	8.20	20.61	33.12	37.02	6 Months	0.01
NG	Natural Gas (NYMEX)	36.65	34.53	9.72	29.77	41.67	123.95	24 Months	0.06
GC	Gold (COMEX)	19.03	17.81	5.27	15.61	21.03	199.60	12 Months	0.99
SI	Silver (COMEX)	31.82	30.33	8.04	26.21	35.71	124.25	12 Months	0.02

For each of the 12 commodity options in our sample, this table reports descriptive statistics for the implied volatilities, essentially the market's expectation of future volatility between now and the option expiry date. It reports the ticker symbol used, the exchange the commodity trades on, the longest option maturity considered after filtering criteria is applied, and the average convenience yield slope of the underlying futures contract. The sample period is January 2006-December 2016.

Cotton, Soybean, Soybean Oil, Sugar, Wheat, Crude Oil, Heating Oil, Natural Gas, Gold, and Silver futures. We seek to model the shape of the observed IVS in the context of this distinct market. We also employ a rolling out-of-sample framework to identify statistically significant forecasting performance. Through an assessment of these models in a multiple comparisons setting, we formally establish the set of superior approaches. We find that in a cross-model comparison of parametric and machine learning approaches no one class of model systematically outperforms. The most striking observation is that the Goncalves and Guidolin (2006) model performs poorly overall, indicating that using smile shape factors only and a linear approximation of the term structure dimension is not sufficient to characterize the dynamics observed in commodity options. However, the results for different classes of commodities are more illuminating, in that for Energy and Precious Metals, the Chalamandaris and Tsekrekos (2011) framework produces the most accurate forecasts. Conversely the theory-free machine learning regression tree approach exhibits the most promising results for Agricultural options, in particular those models with flatter underlying futures convenience yields.

The paper proceeds as follows: Section 2 presents the data used in the study. Section 3 outlines the models used to characterize and forecast the IVS. Section 4 provides an empirical assessment of out-of-sample testing. Finally, Section 5 concludes and suggests avenues for potential future research.

2. The data

The dataset is obtained from ivolatility.com and contains IV quotes for a representative sample of the most actively traded commodity options, covering Agricultural, Energy and Precious Metals. The primary data are from ACTIV Financial Systems, Inc., which is an Options Price Reporting Authority-approved vendor, with data incorporating traded-level information including volume. Specifically, it includes options data for Cocoa, Corn, Cotton, Soybean, Soybean Oil, Sugar, Wheat, Crude Oil, Heating Oil, Natural Gas, Gold, Copper, and Silver futures. The commodities can be grouped into three broad classes: Agricultural (Cocoa, Corn, Cotton, Soybean, Soybean Oil, Sugar, and Wheat); Energy (Crude Oil, Heating Oil, and Natural Gas); and Precious Metals (Copper, Gold, and Silver). The sampling frequency is daily and the sample period runs from January 2006 to December 2016. We consider both call and put options but only include contracts that have at least one transaction on a given day, that is volume > 0. As one would expect we do not have the same number of traded contracts on every observation date. To enhance data quality and to ensure that

all segments of the surface are actively traded, we first consider only options in the 90–110% moneyness range and with maturities of between one month and two years. Second, to eliminate any remaining bias arising from thinly traded expiration dates, we concentrate on the most liquid maturities by imposing a filtering rule. This rule requires at least 15,000 quotes within each maturity group of 1–6 months, 6–12 months, 12–18 months, and 18–24 months, for that maturity group to be included in our dataset. This filtering criterion ensures that we have at least five IV quotes for each day for each maturity group. However, the vast majority of days across the commodities in our sample contain many multiples of this. Unfortunately, after filtering the data, Copper options have an insufficient number of contracts traded and are therefore dropped from our dataset. The results of this filtering exercise, as well as the descriptive statistics for the dataset, are reported in Table 1. Mean IV levels lie in the 19.03–36.65% range with distributions displaying a predominantly positive skew, in line with the commodity options literature (Mandelbrot, 1963; Fama, 1965).

Further, we include an indicator of the term structure of the underlying futures curves for each of our commodities through the inclusion of an average convenience yield slope. The convenience yield slope provides an indication of both the cost of carrying, in our case most often storage and insurance, and also the market expectation of future price trajectory. We can see how this varies across commodities as strong futures curves dynamics are observed for Soybeans and Wheat, with average slopes of -5.21 and 3.71, respectively.

Fig. 1 presents the average IVS for four heavily traded options (Corn, Gold, Crude Oil, and Natural Gas) over the in-sample 2006–2014 period. The average IVS for Corn is presented in the upper left quadrant. We observe a positive skew in the moneyness dimension, whereby out-of-the-money options are more expensive than their in-the-money counterparts, implying fear of future supply-side disruption, as outlined by Askari and Krichene (2008); Liu and Tang (2011) and Kearney et al. (2015). There is also evidence of a premium for short-term maturity options. This suggests that the market anticipates the impact of such supply-side disruptions to even out over a longer time-frame.

Precious Metals are often viewed as a separate asset in their own right, given that they display unique attributes in comparison to other commodities. Such bespoke properties can be observed here when focusing on the average Gold IVS. Despite being predominantly positively skewed the IV smile displays a complex structure. The plot also reveals the non-linear shape of the term structure component. As the option maturities increase up to 12 months, the associated IVs also increase. This is in contrast to
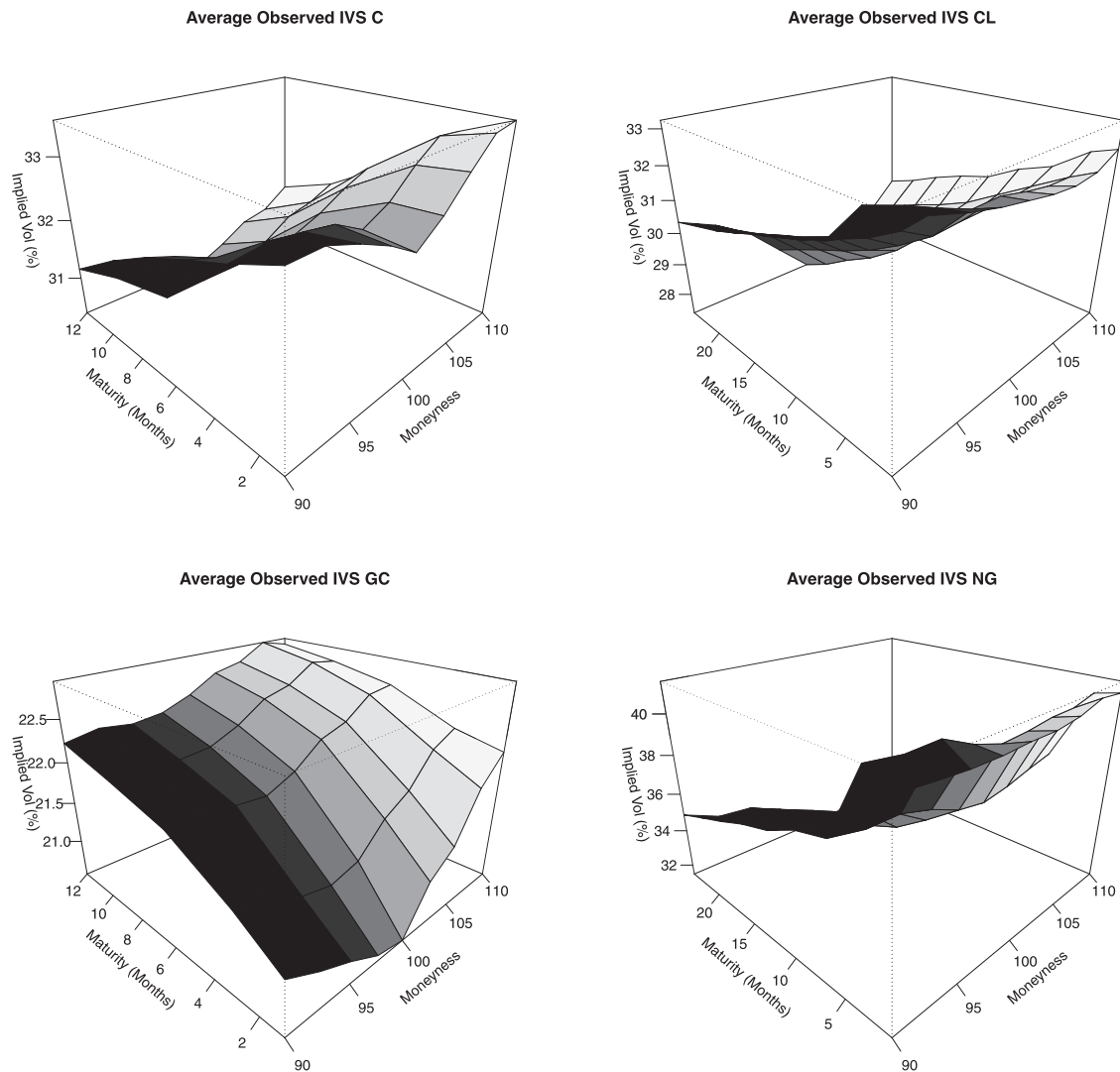
Fig. 1. Average Implied Volatility Surfaces This figure presents the average IVS plotted for each class of commodity over the January 2006-December 2014 in-sample period. In clockwise order: Corn, Crude Oil, Gold, and Natural Gas. The IVS is a three-dimensional plot where the *x*-axis is the time-to-maturity, the *z*-axis is the moneyness, and the *y*-axis is the implied volatility derived from the option contracts.

the dynamics of the other commodity classes. The dynamic corresponds to a relatively high average underlying futures convenience yield curve slope, at 0.99. This yield curve slope indicates that futures on gold contracts further in the future are more expensive than those futures that are closer to maturity.

Finally, we return to the Energy class, with the average IVS for Natural Gas and Oil options presented in the top right and lower right quadrants of Fig. 1. Interestingly, despite being Energy commodities where one would expect fear of supply-side disruptions to dominate, in line with Doran and Ronn (2008), the primary driver during our in-sample period is fear of demand-side shock. This is inferred by the slight negative moneyness skew observed and could be due to the weak economic environment caused by the global financial crisis and its aftermath depressing demand for fossil fuel use. Furthermore, there is visually significant curvature across the maturities in line with Doran and Ronn (2008), who highlight the presence of such curvature in Energy markets. In the next section, we outline models to capture the interesting IVS dynamics presented here.

While useful for understanding the predominant dynamics observed within an individual surface, the above figures do not show us how the surfaces evolve. In contrast, Fig. 2 plots the

evolution of different segments of the IVS over time. It is presented for Gold option contracts broken down first by moneyness and second by maturity. The groups are as follows: out-of-the-money (OTM) (<95%), at-the-money (ATM) (97.5–102.5%), and in-the-money (ITM) (>105%), Short- (1 to 3 months), Medium- (3 to 6 months), and Long-term (<6 months) contracts. We illustrate the evolution of IV for Gold commodities in the year 2010 in isolation to ensure an appropriate graphical resolution for the dynamics.[5] From these plots, we can see that the average level of IV evolves greatly over a single-year period and that the relative relationships between the different moneyness and maturity groups do not stay constant. For instance, the January-March period at the beginning of the plot shows a noticeably defined order in terms of long-term maturities being associated with high IV, relative to shorter-term maturities. However, this relationship does not hold in the later October-November period. The complexity of these dynamics, even on an aggregated basis such as that presented here, motivates the need to explicitly model the evolution of the full IVS.

---

[5] Plots for other time periods and commodity options are available from the authors upon request.
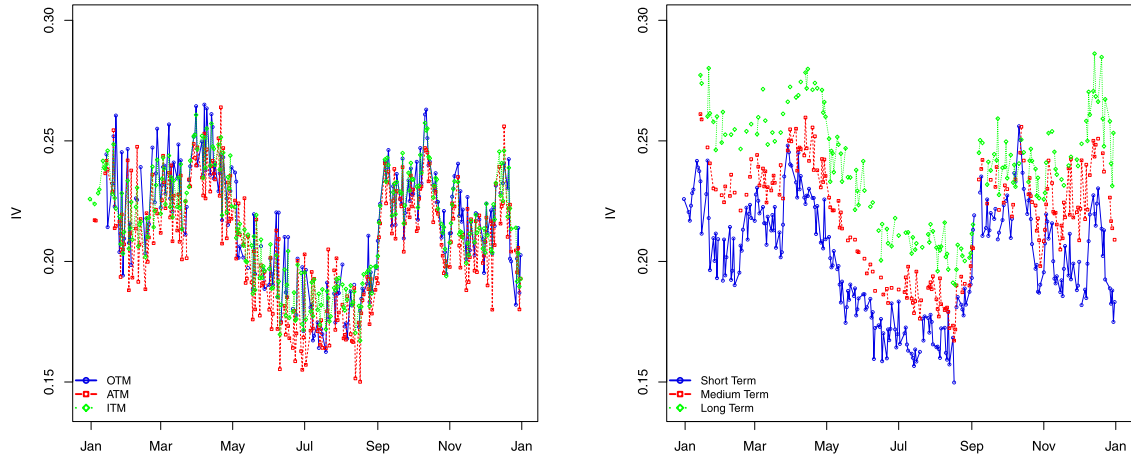
Fig. 2. Gold Implied Volatility Evolution 2010 This figure presents the evolution of the average IV for gold commodity options over the year 2010. It is split by both moneyness and maturity: Out-of-the-money (OTM) (<95%), At-the-money (ATM) (97.5% to 102.5%) and In-the-money (ITM) (>105%); Short- (one to three months), Medium- (three to six months) and Long-term (> six months) contracts.

3. Methodology

3.1. Modeling and forecasting the surface

We now outline three general frameworks to describe and forecast the IVS. First, Goncalves and Guidolin (2006) use a parametric approach, in that they model daily IV surfaces using Dumas et al. (1998) specifications based on moneyness and time-to-maturity. Second, Chalamandaris and Tsekrekos (2011) extend Dumas et al. (1998) by concluding that the linear approximation of maturity is insufficient. They instead propose the use of Nelson-Siegel term structure factors in the spirit of Diebold and Li (2006) to produce fitted IV surfaces. Forecasts from both these parametric models are produced in a two-step framework, as outlined in Section 3.3, by assuming that the estimated coefficients evolve according to a specified time series process. This allows us to construct a forecasted IVS using these predicted coefficients. Finally, we employ the regression tree benchmark model from Audrino and Colangelo (2010). We now present each specification in more detail.

Goncalves and Guidolin (2006) propose a parametric specification based on moneyness and time-to-maturity to characterize the IVS. More precisely, they employ the following model, based on Dumas et al. (1998):

$$\sigma_{i,t} = \alpha_{0,t} + \alpha_{1,t}\Delta_i + \alpha_{2,t}\Delta_i^2 + \alpha_{3,t}\tau_i + \alpha_{4,t}(\Delta_i \times \tau_i) + \epsilon_{i,t},$$

where Δ_i is the option moneyness, τ_i is the option time-to-maturity, and $\epsilon_{i,t}$ is a random error term. The estimates, $\widehat{\alpha}_t = [\widehat{\alpha}_{0,t}, \widehat{\alpha}_{1,t}, \ldots, \widehat{\alpha}_{4,t}]^\top$ are interpreted as proxies for latent factors that drive the daily evolution of the IVS. Goncalves and Guidolin (2006) subsequently proceed to show that modeling the dynamics of $\widehat{\alpha}_t$ can provide superior out-of-sample forecasts of IV.

One drawback of Goncalves and Guidolin (2006) specification is the assumed symmetry in the moneyness dimension of the IVS. A second drawback is that the model linearly approximates the term structure of the IVS. In the context of our study, commodity markets have been shown to exhibit both IV skews and non-linear term structure. To address these two issues, Chalamandaris and Tsekrekos (2011) employ an augmented version of Dumas et al. (1998) parametric specification, namely:

$$\sigma_{i,t} = \beta_{0,t} + \beta_{1,t}\mathbf{1}_{\{\Delta_i>0\}}\Delta_i^2 + \beta_{2,t}\mathbf{1}_{\{\Delta_i<0\}}\Delta_i^2 + \beta_{3,t}\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i}$$

$$+ \beta_{4,t}\left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}\right) + \beta_{5,t}\mathbf{1}_{\{\Delta_i>0\}}\Delta_i\tau_i$$

$$+ \beta_{6,t}\mathbf{1}_{\{\Delta_i<0\}}\Delta_i\tau_i + \epsilon_{i,t},$$

where $\mathbf{1}_{\{x\}}$ is an indicator function that takes the value one if condition x is true, and zero otherwise. The terms $\frac{(1-e^{-\lambda\tau_i})}{\lambda\tau_i}$ and $\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}$ represent the term structure of the IVS and are based on the Nelson-Siegel factors as successfully employed by Diebold and Li (2006) to describe and forecast the yield curve. The λ parameter determines the exponential decay rate of the term structure. In line with Diebold and Li (2006), we first utilize a non-linear estimation of all parameters, including λ, by minimizing the daily sum of squared errors. Following this we fix λ to be equal to the median estimated value from the first step, and re-estimate the model using ordinary least squares. Coefficients $\beta_{1,t}$ and $\beta_{2,t}$ capture the IV skew by modeling in-the-money and out-of-the-money options separately. Similarly, the coefficients $\beta_{5,t}$ and $\beta_{6,t}$ capture the attenuation of the smile with time-to-maturity. Chalamandaris and Tsekrekos (2011) show that modeling the dynamics of $\widehat{\boldsymbol{\beta}}_t = [\widehat{\beta}_{0,t}, \widehat{\beta}_{1,t}, \ldots, \widehat{\beta}_{6,t}]^\top$ leads to accurate forecasts of the entire IVS out-of-sample. The methodology for forecasting these coefficients is outlined Section 3.2.

We also consider a regression tree approach. A regression tree is a set of logical conditions that recursively creates a binary partition of the predictor space. The algorithm has three main components: (1) a way to select a split rule; (2) a rule to determine when a tree node is terminal; and (3) a standard for assigning a value to each terminal node. Its ability to split variables to construct a regression tree with end nodes is of great importance.

In Audrino and Colangelo (2010), the IV is regressed on a vector of predictors $\boldsymbol{x}^{\text{pred}}$ through an unspecified function $f_{m,\tau}$ such that

$$\sigma_{\Delta,\tau}^{\text{IV}} = f_{\Delta,\tau}\left(\boldsymbol{x}^{\text{pred}}\right) + \epsilon_{\Delta,\tau}$$

$$= f_{\Delta,\tau}(\Delta,\tau) + \epsilon_{\Delta,\tau},$$

where Δ denotes option moneyness and τ denotes maturity. $\mathrm{E}[\epsilon_{\Delta,\tau}] = 0$ and $\mathrm{E}[\epsilon_{\Delta,\tau}^2] < \infty$ for each Δ, $\tau > 0$. The functions $f_{\Delta,\tau}(\cdot)$ are estimated by minimizing the expectation of a given loss function in a regression tree framework. However, when a decision tree is formed in this manner, the classification algorithm can generate some unwanted rules as it grows deeper, referred to as overfitting. To avoid this, we prune our regression tree using cost-complexity pruning as proposed by Breiman et al. (1984) (see Esposito et al., 1997, for further technical details).

Following the approach of Audrino and Colangelo (2010), we specify our regression function $f_{\Delta,\tau}$ to be a linear additive expansion of regression trees. More specifically, we use Gini's diversity index to choose the splits up to a maximum of ten end nodes. The complexity parameter is selected using the minimum

Table 2
Average estimated coefficients.

Commodity	Goncalves and Guidolin (2006) estimation					Chalamandaris and Tsekrekos (2011) estimation						
	α_0	α_1	α_2	α_3	α_4	β_0	β_1	β_2	β_3	β_4	β_5	β_6
Cocoa	2.247	−44.508	0.024	269.184	−0.498	0.301	0.515	1560.918	−0.011	0.154	−0.074	5.326
Corn	0.330	0.022	−0.003	0.332	−0.001	0.254	0.856	0.728	0.063	0.123	−0.008	0.014
Cotton	0.308	0.037	−0.006	0.404	−0.006	0.183	0.855	12.216	2.025	−2.217	−0.010	0.055
Soybean	0.268	0.054	−0.001	0.601	−0.007	0.237	1.655	0.819	53.163	−53.236	−0.011	0.008
Soybean Oil	0.204	0.315	0.015	1.313	−0.179	0.242	2.246	−193.687	−7.619	10.879	−0.044	0.069
Sugar	0.345	0.075	−0.006	0.851	−0.025	0.298	1.576	−7.398	0.009	0.149	−0.017	−0.061
Wheat	0.344	0.049	−0.005	0.521	−0.005	0.286	1.006	0.898	75.931	−78.678	−0.006	0.003
Crude Oil	0.337	−0.014	−0.004	0.910	−0.004	0.218	0.845	1.506	0.118	0.102	−0.007	0.003
Heating Oil	0.269	0.097	0.002	0.599	−0.024	0.283	0.562	−210.339	0.001	−0.047	0.017	0.244
Natural Gas	0.464	−0.031	−0.023	2.348	0.001	0.290	−7.135	3.013	0.237	0.167	−0.014	0.122
Gold	0.199	0.047	0.003	1.523	−0.012	0.252	2.162	3.595	−0.057	−0.045	−0.017	0.041
Silver	0.345	0.030	0.000	1.558	−0.006	0.359	2.683	−66.433	−0.007	−0.044	−0.012	−0.336

For each of the 12 different commodity options in our sample, this table presents the average daily estimated coefficients for the Goncalves and Guidolin (2006) and Chalamandaris and Tsekrekos (2011) models fitted to our implied volatility data set during the January 2006-December 2014 in-sample period.

cross-validated error calculated across the full in-sample period 2006–2014. This selected complexity parameter is then used as a constant throughout our forecasting exercise. Computationally, the regression trees are implemented using the *rpart* package (Therneau and Atkinson, 2018) in R (R Core Team, 2018). Once the regression function $\widehat{f}_{\Delta,\tau}$ is estimated, the forecast $\sigma_{\Delta,\tau}^{IV}$ is obtained by providing it with the new set of moneyness and maturity variables. This forecasting model is henceforth referred to as RT.

3.2. Time series properties of parameters

First, we estimate the models of Goncalves and Guidolin (2006) (GG) and Chalamandaris and Tsekrekos (2011) (CT) on a daily basis. The average estimated coefficients over the in-sample period presented in Table 2. Focusing first on the CT model, the skews in the moneyness dimension observed here are exhibited through the different coefficients for the left and right smiles. CT also estimates coefficients for both short- and medium-term structure variables, with negative estimates calculated for Gold options for instance. This relates to a dynamic that options with longer-dated expiries have higher IV than those with short-term maturity, and corresponds to the dynamics observed in Fig. 1. It should be noted here, however, that for Cocoa options the majority of days exhibits insignificant parameter estimates.

Estimating GG and CT models on a daily basis, as we do, produces sets of dynamic factors, $\widehat{\boldsymbol{\alpha}}_t$ and $\widehat{\boldsymbol{\beta}}_t$, respectively. Figs. 3 and 4 show the evolution of the coefficients estimated throughout 2010. They demonstrate that the coefficients are not constant over time, with some coefficients even alternating between positive and negative states throughout a single year. We now empirically investigate the time series properties of these parameters. This, in turn, informs our choice of which time series processes to adopt to model the evolution of the factors. In order to establish the time series properties we present the cross-correlation matrices for GG and CT separately in Tables 3 and 4, and the autocorrelation functions of each coefficient in Figs. 5 and 6.

First, we focus on the autocorrelation function plots to identify temporal dependencies within each parameter series. There is a pattern of high autocorrelation function (ACF) values observed across the majority of the series, motivating the adoption of a simple autoregressive (AR) process to model this evolution. Second, parameters such as Beta 2 and Beta 6 of the CT model may be well modeled by a moving average model, whereby the ACF values decay dramatically at low lag orders. We incorporate this dynamic along with the widespread autoregressive dependency established previously through the Autoregressive Integrated Moving Average (ARIMA). In the ARIMA, d is selected based on

Table 3
Correlation matrix of GG-coefficients.

	α_0	α_1	α_2	α_3	α_4
α_0	1	−0.262***	−0.536***	−0.362***	0.375***
α_1	−0.262***	1	0.208***	−0.211***	−0.743***
α_2	−0.536***	0.208***	1	0.082***	−0.265***
α_3	−0.362***	−0.211***	0.082***	1	−0.301***
α_4	0.375***	−0.743***	−0.265***	−0.301***	1

This table presents the calculated correlation matrix between the five coefficient values of the Goncalves and Guidolin (2006) model when estimated for the Gold commodity option during the in-sample January 2006-December 2014 period. The alpha coefficients, α_0, α_1, α_2, α_3 and α_4, refer to the variables of intercept, moneyness, maturity, moneyness multiplied by maturity, and moneyness squared, respectively. *, **, and ***, indicate significance at 10%, 5%, and 1%, respectively.

successive Kwiatkowski–Phillips–Schmidt–Shin (KPSS) unit root tests (Kwiatkowski et al., 1992).[6] We test the original data for a unit root; if the test result is significant, then we test the differenced data for a unit root. The procedure continues until we obtain our first insignificant result. Third, based on the possible exponential decay patterns observed across GG Alpha 1, CT Beta 3, and CT Beta 4, we consider an exponential smoothing time series approach with optimal trend and seasonal components selected using a corrected Akaike information criterion. Finally, we turn our attention to Tables 3 and 4, where we learn from the cross-correlation matrices that significant linear relationships exist between the coefficient estimates produced by both models. This motivates the use of a tractable multivariate time series approach to capture these interacting dynamics simultaneously; for this reason, we also specify a VAR model. We now provide more details of these time series models.

In line with Chalamandaris and Tsekrekos (2014) we hypothesize that the shape of the IVS at time $t + h$ depends on factors $\widehat{\boldsymbol{\alpha}}_{t+h}$ and $\widehat{\boldsymbol{\beta}}_{t+h}$. To model the time evolution of $\widehat{\boldsymbol{\alpha}}_t$, and $\widehat{\boldsymbol{\beta}}_t$ we employ various time series models, namely AR, VAR, ARIMA, and Exponential Smoothing (ETS), to produce forecasts of their coefficients. Finally, we also adopt a no change driftless random walk (RW) model to act as an intuitive benchmark.

More specifically, forecasts are produced using:

$$\widehat{\sigma}_{i,t+h} = \widehat{\alpha}_{0,t+h} + \widehat{\alpha}_{1,t+h}\Delta_i + \widehat{\alpha}_{2,t+h}\Delta_i^2 + \widehat{\alpha}_{3,t+h}\tau_i + \widehat{\alpha}_{4,t+h}(\Delta_i \times \tau_i) + \widehat{\epsilon}_{i,t+h},$$

---

[6] KPSS tests are primarily used for testing the null hypothesis that an observable time series is stationary around a deterministic trend, although KPSS tests can also detect the presence of a long-memory process.
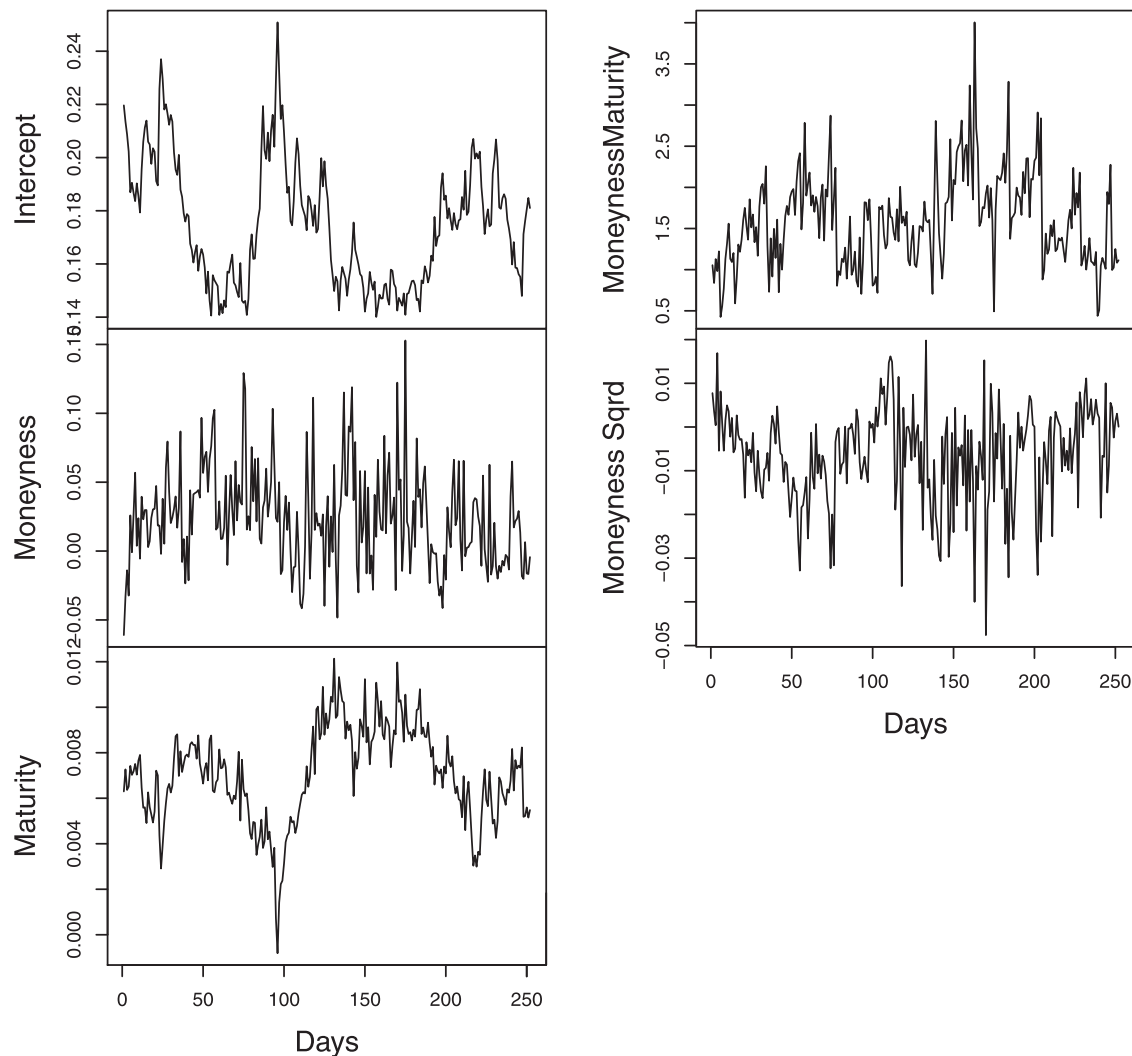
GC GG Coefficients 2010

Fig. 3. Gold Commodity Options; Estimated GG-Coefficients (2010) This figure presents the time variation of the five coefficient values of the Goncalves and Guidolin (2006) model when estimated for Gold commodity options during the year 2010. The coefficients presented refer to the variables of intercept, moneyness, maturity, moneyness multiplied by maturity, and moneyness squared.

and

$$\widehat{\sigma}_{i,t+h} = \widehat{\beta}_{0,t+h} + \widehat{\beta}_{1,t+h}\mathbf{1}_{\{\Delta_i > 0\}}\Delta_i^2 + \widehat{\beta}_{2,t+h}\mathbf{1}_{\{\Delta_i < 0\}}\Delta_i^2$$
$$+ \widehat{\beta}_{3,t+h}\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} + \widehat{\beta}_{4,t+h}\left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}\right)$$
$$+ \widehat{\beta}_{5,t+h}\mathbf{1}_{\{\Delta_i > 0\}}\Delta_i\tau_i + \widehat{\beta}_{6,t+h}\mathbf{1}_{\{\Delta_i < 0\}}\Delta_i^2 + \widehat{\epsilon}_{i,t+h},$$

where $\widehat{\alpha}_{t+h} = [\widehat{\alpha}_{0,t+h}, \ldots, \widehat{\alpha}_{4,t+h}]^\top$ and $\widehat{\beta}_{t+h} = [\widehat{\beta}_{0,t+h}, \ldots, \widehat{\beta}_{6,t+h}]^\top$ are the result of time-series specifications:

- AR (henceforth referred to as GG-AR and CT-AR, for the Goncalves and Guidolin (2006) and Chalamandaris and Tsekrekos (2011) models, respectively)

Table 4
Correlation matrix of CT-Coefficients.

	β_0	β_1	β_2	β_3	β_4	β_5	β_6
β_0	1	−0.227***	0.064***	−0.169***	−0.115***	0.058**	0.056**
β_1	−0.227***	1	0.036	−0.125***	−0.350***	−0.864***	0.029
β_2	0.064***	0.036	1	−0.032	−0.045**	−0.057**	0.984***
β_3	−0.169***	−0.125***	−0.032	1	−0.13***	0.123***	−0.026
β_4	−0.115***	−0.350***	−0.045**	−0.13***	1	0.453***	−0.037
β_5	0.058**	−0.864***	−0.057**	0.123***	0.453***	1	−0.043*
β_6	0.056**	0.029	0.984***	−0.026	−0.037	−0.043*	1

This table presents the calculated correlation matrix between the seven coefficient values of the Chalamandaris and Tsekrekos (2011) model when estimated for the Gold commodity option during the in-sample January 2006-December 2014 period. The beta coefficients, β_0, β_1, β_2, β_3, β_4, β_5 and β_6, refer to the variables of implied volatility level, right smile, left smile, short-term structure, medium-term structure, right smile attenuation, and left smile attenuation, respectively. *, **, and ***, indicate significance at 10%, 5%, and 1%, respectively.
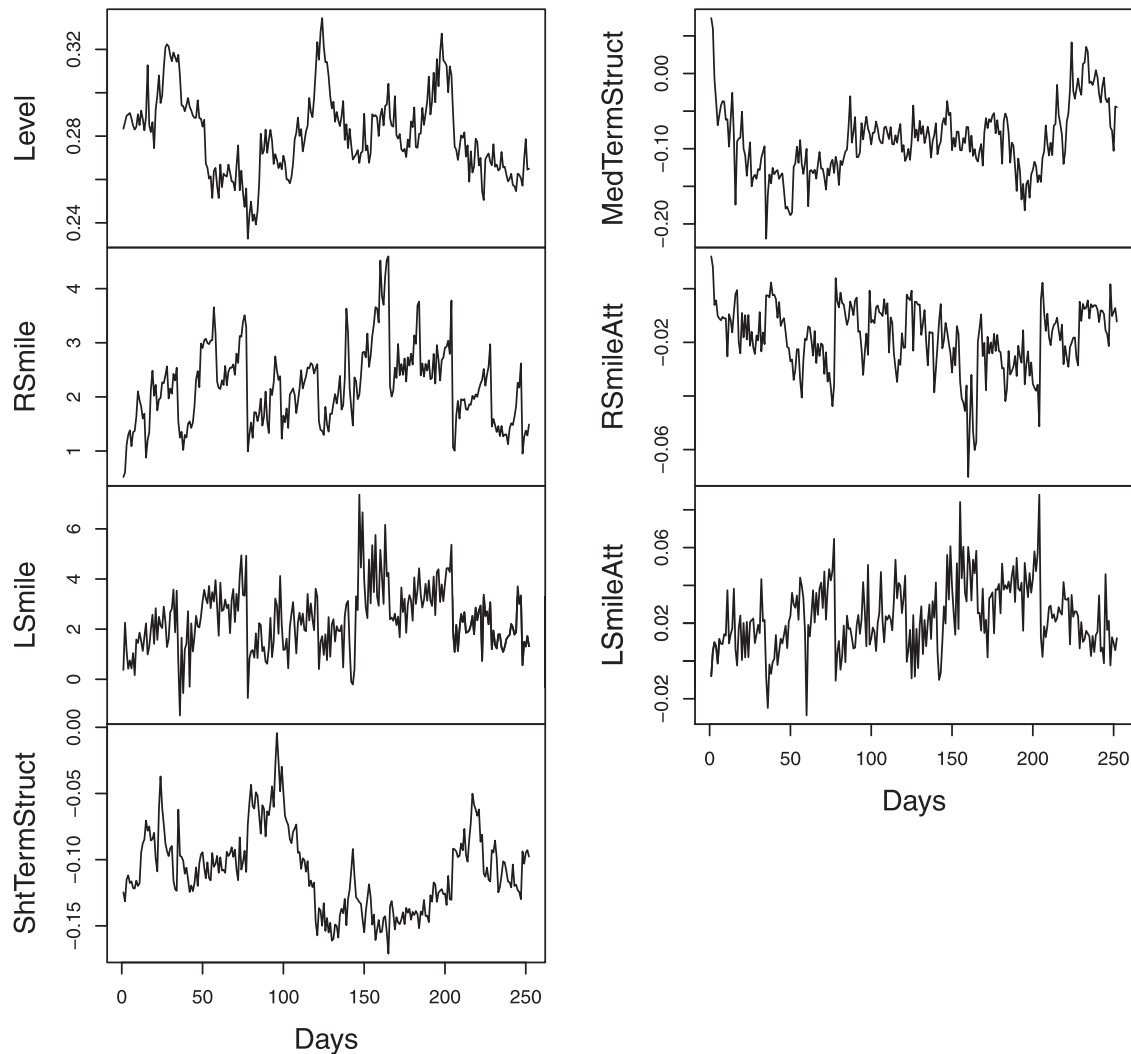
GC CT Coefficients 2010

Fig. 4. Gold Commodity Options; Estimated CT-Coefficients (2010) This figure presents the time variation of the seven coefficient values of the Chalamandaris and Tsekrekos (2011) model when estimated for Gold commodity options during the year 2010. The coefficients presented refer to the variables of implied volatility level, right smile, left smile, short-term structure, medium-term structure, right smile attenuation and left smile attenuation.

- ARIMA (henceforth referred to as GG-ARIMA and CT-ARIMA, for the Goncalves and Guidolin (2006) and Chalamandaris and Tsekrekos (2011) models, respectively)
- ETS (henceforth referred to as GG-ETS and CT-ETS, for the Goncalves and Guidolin (2006) and Chalamandaris and Tsekrekos (2011) models, respectively)
- VAR (henceforth referred to as GG-VAR and CT-VAR, for the Goncalves and Guidolin (2006) and Chalamandaris and Tsekrekos (2011) models, respectively).
- As a no change benchmark we also specify the RW (henceforth referred to as GG-RW and CT-RW, for the Goncalves and Guidolin (2006) and Chalamandaris and Tsekrekos (2011) models, respectively). This essentially amounts to keeping the parameters constant.

Forecasting horizons of 1, 2, 5, 10, and 30 days ahead are considered. For forecasts of longer than one-day ahead we implement direct multi-day ahead forecasts. Our choice of direct forecasts is based on Cox (1961), who concludes that they are more efficient than iterated forecasts in the context of exponential smoothing, and Klein (1968), who suggest employing direct multi-

period estimation of dynamic forecasting models. Further, Bhansali (1999) and Ing (2003) conclude that the robustness of the direct approach to model misspecification makes it a more attractive procedure than the bias-prone indirect approach.

3.3. Out-of-sample forecasting

We assess the forecast performance of each model using the following measures:

1. Root mean squared error (RMSE) is a measure of the differences between the realized values and the values predicted by a model. It is defined as the square root of the mean squared forecast error and serves to aggregate the errors into a single measure of predictive power:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\sigma_{i,t+h} - \widehat{\sigma}_{i,t+h})^2}{n}},$$

where $\sigma_{i,t+h}$ are the observed values and $\widehat{\sigma}_{i,t+h}$ are the values predicted from the model.
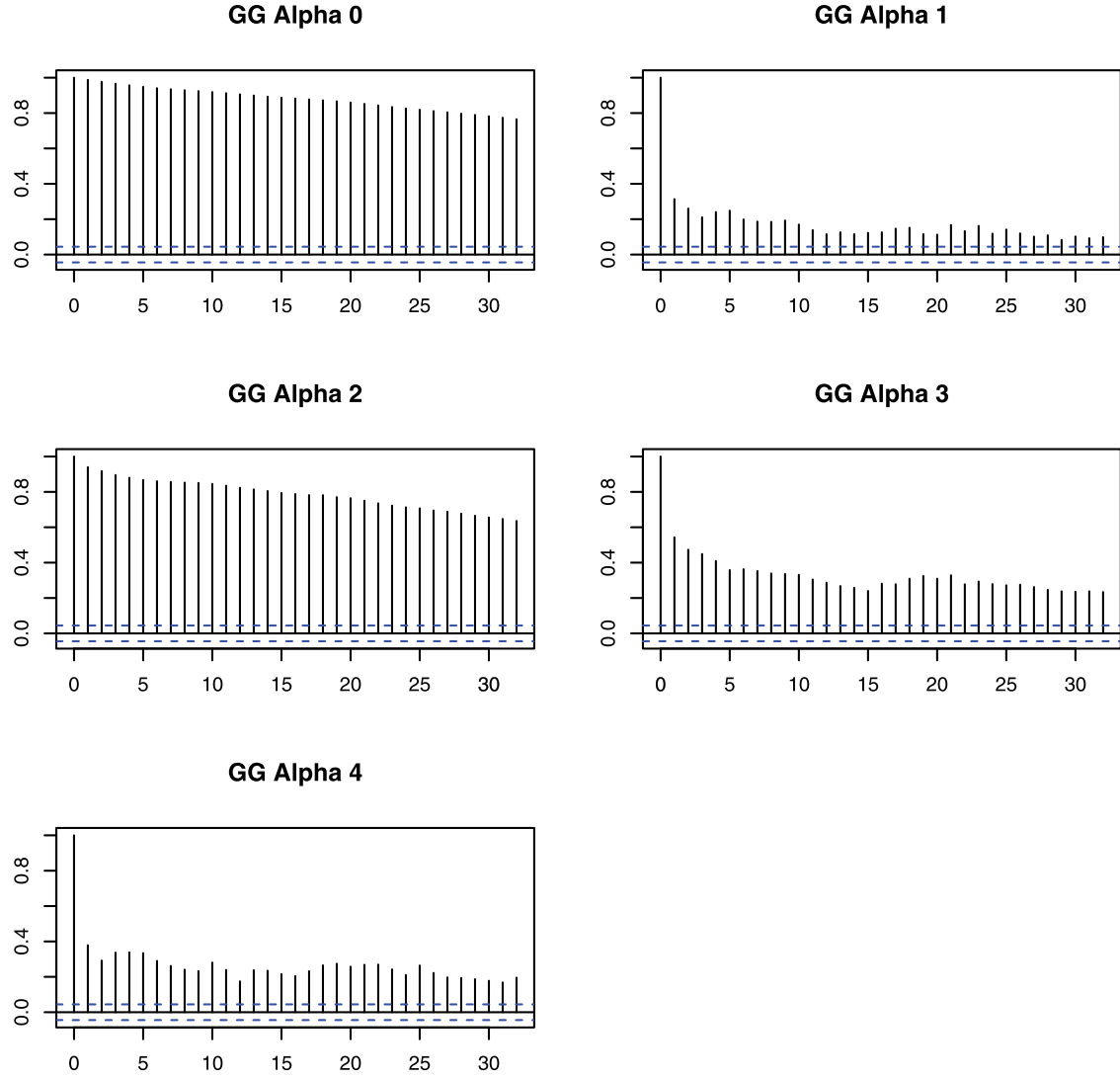
Fig. 5. Autocorrelation Function of GG-Coefficients for Gold Commodity Options This figure presents the calculated autocorrelation function for the five coefficient values of the Goncalves and Guidolin (2006) model when estimated for Gold commodity options during the in-sample January 2006-December 2014 period. The *x*-axis represents the lags in days. The Alpha coefficients presented refer to the variables of intercept, moneyness, maturity, moneyness multiplied by maturity, and moneyness squared.

2. Root mean squared percentage error (RMSPE) is also a measure of the differences between the realized values and the values predicted by a model, however it is scaled by the realized observations and is expressed in percentage terms, as follows:

$$\text{RMSPE} = \sqrt{\frac{\sum_{i=1}^{n}\left[\frac{100(\sigma_{i,t+h}-\widehat{\sigma}_{i,t+h})}{\sigma_{i,t+h}}\right]^2}{n}},$$

where $\sigma_{i,t+h}$ are the observed values and $\widehat{\sigma}_{i,t+h}$ are the values predicted from the model.

3. Mean absolute percentage error (MAPE) is the average of the absolute differences between the realized values and the values predicted by the model, again scaled by the realized observations and expressed in percentage terms:

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\sigma_{i,t+h}-\widehat{\sigma}_{i,t+h}}{\sigma_{i,t+h}}\right|,$$

where $\sigma_{i,t+h}$ are the observed values and $\widehat{\sigma}_{i,t+h}$ are the values predicted from the model.

4. The sign success ratio (SSR) is the percentage of predictions for which the change in the values predicted by the model, $\widehat{\sigma}_{i,t+h}$, have the same sign as the corresponding change in the realized

values, $\sigma_{i,t+h}$. The SSR measures how well the model can predict the direction of movement, regardless of error magnitude.

We implement a rolling window approach to produce forecasts for each of our 500 out-of-sample days. Using a size-constant rolling window of 1167 days, we produce daily forecasts. The forecasts we produce are based only on values available at the date on which the forecast is made. Further, we never expand the window by adding new observations. Instead, we add the new observation and remove the oldest one when rolling over. This rolling window approach has the advantage of mitigating the impact of structural breaks. As a further sensibility check, we constrain all time-series IV forecasts to reside in the [0%, 100%] range. We formally test if any outperformance uncovered is sample specific or if we can draw inferences regarding the entire population. To this end, we utilize the Model Confidence Set outlined in the next section to establish which model(s) are statistically superior.

3.4. Model confidence set

To examine statistical significance among our multiple approaches, we employ the model confidence set (MCS) procedure. The MCS procedure proposed by Hansen et al. (2011) consists of
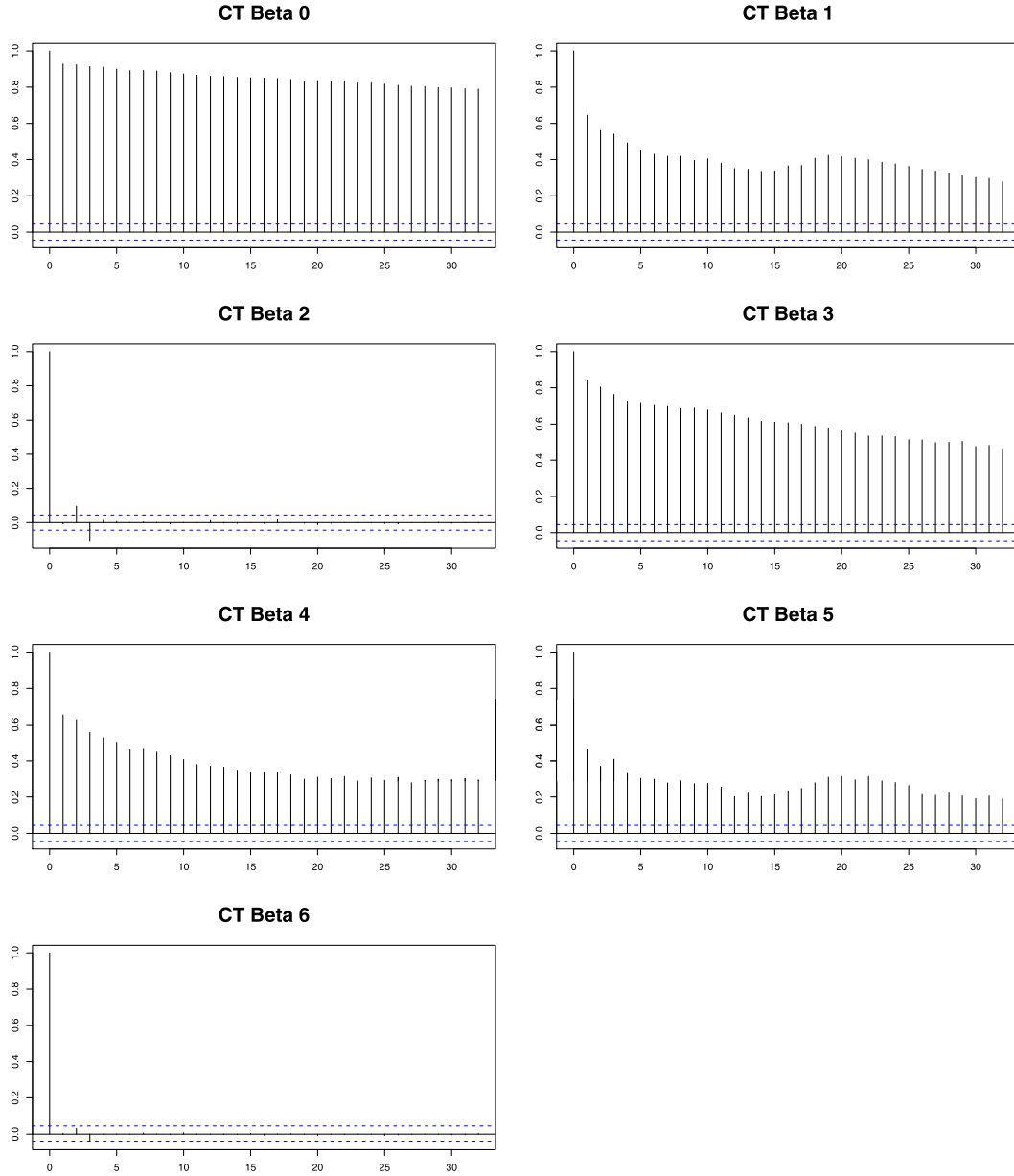
Fig. 6. Autocorrelation Function of CT-Coefficients for Gold Commodity Options This figure presents the calculated autocorrelation function for the seven coefficient values of the Chalamandaris and Tsekrekos (2011) model when estimated for Gold commodity options during the in-sample January 2006-December 2014 period. The x-axis represents the lags in days. The Beta coefficients presented refer to the variables of implied volatility level, right smile, left smile, short-term structure, medium-term structure, right smile attenuation, and left smile attenuation.

a sequence of tests that permit construction of a set of "superior" models, where the null hypothesis of equal predictive ability (EPA) is not rejected at a specified confidence level. As the EPA test statistic can be evaluated for any loss function, we follow Hansen et al. (2011) and adopt the popular squared error measure at a 75% confidence level.

The procedure begins with our initial set of models of dimension $m = 11$ encompassing all the IVS forecasting models considered, $M_0 = \{M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8, M_9, M_{10}, M_{11}\}$. For a given confidence level, the superior set of models $\widehat{M}^*_{1-\alpha}$, is determined where $m^* \leq m$. The best scenario is when the final set consists of a single model that is, $m = 1$. First, let $d_{ij,t}$ denote the loss differential between models i and j, that is:

$$d_{ij,t} = l_{i,t} - l_{j,t}, \qquad i, j = 1, \ldots, m, \quad t = 1, \ldots, n.$$

The EPA hypothesis for a given set of M candidate models can subsequently be formulated by:

$$H_{0,M} : c_{ij} = 0, \qquad \text{for all} \quad i, j = 1, 2, \ldots, m$$
$$H_{A,M} : c_{ij} \neq 0, \qquad \text{for some} \quad i, j = 1, 2, \ldots, m, \qquad (1)$$

where $c_{ij} = E(d_{ij})$ is assumed to be finite and time independent. Based on c_{ij} we construct a hypothesis test as follows:

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{\text{Var}}(\bar{d}_{ij})}}, \qquad (2)$$

where $\bar{d}_{ij} = \frac{1}{n} \sum_{t=1}^{n} d_{ij,t}$ measures the relative sample loss between the ith and j^{th} models. Note that $\widehat{\text{Var}}(\bar{d}_{ij})$ are the bootstrapped estimates of $\text{Var}(\bar{d}_{ij})$. To produce these we perform a block bootstrap

Table 5
Out-of-sample RMSPE.

Forecasting	Agricultural							Energy			Metals	
Horizon (h)	Cocoa	Corn	Cotton	Soybean	Soybean Oil	Sugar	Wheat	Crude Oil	Heating Oil	Natural Gas	Gold	Silver
h = 1												
RT	**20.30**	25.55	**29.65**	23.77	**10.81**	19.08	16.36	28.72	30.57	13.32	14.14	15.11
GG-AR	335.31	103.10	54.06	128.90	110.08	55.05	69.96	55.39	39.33	59.34	167.43	78.49
CT-AR	103.33	6.49	94.29	7.14	28.15	10.00	**6.48**	8.19	40.86	11.45	5.88	6.35
GG-VAR	284.40	105.84	60.50	135.69	113.86	48.21	72.69	47.59	49.13	57.13	170.22	87.21
CT-VAR	110.98	**6.44**	112.33	**6.96**	27.22	12.73	6.78	**5.92**	48.38	**6.54**	**5.69**	4.97
GG-ARIMA	100.00	112.70	54.10	134.60	100.24	41.43	82.18	48.81	**24.39**	58.68	164.10	79.81
CT-ARIMA	216.75	6.49	147.31	7.11	34.62	**8.51**	6.71	6.71	46.51	8.75	5.77	5.07
GG-ETS	342.83	112.84	55.47	136.52	110.04	41.39	83.05	49.32	42.63	61.37	162.78	80.38
CT-ETS	152.35	6.45	116.03	7.07	26.31	8.71	6.69	6.73	36.09	8.75	5.70	**4.90**
h = 5												
RT	**20.30**	25.55	**29.65**	23.77	**10.81**	19.08	16.36	28.72	**30.57**	13.32	14.14	15.11
GG-AR	329.22	102.04	82.06	116.94	86.19	81.86	88.93	94.98	89.06	93.10	132.89	78.93
CT-AR	103.38	9.34	94.43	10.92	28.35	10.12	8.97	9.74	40.14	9.01	**8.06**	6.77
GG-VAR	249.33	104.53	56.96	134.23	113.38	50.72	67.47	46.36	51.10	57.76	173.97	85.25
CT-VAR	112.95	9.27	111.05	**9.67**	31.88	10.57	**8.84**	**9.04**	45.75	**7.73**	8.35	7.01
GG-ARIMA	100.00	107.71	81.51	121.75	84.19	87.87	90.78	96.11	91.47	93.85	132.54	79.34
CT-ARIMA	214.30	9.63	147.38	10.88	34.73	9.71	9.27	9.64	45.63	9.64	8.30	**6.42**
GG-ETS	312.13	107.98	81.31	123.53	85.77	87.95	91.26	95.94	90.99	94.56	132.24	79.56
CT-ETS	154.46	**9.22**	114.52	9.69	27.04	**9.28**	8.96	9.56	35.41	9.61	8.16	6.56
h = 30												
RT	**20.30**	25.55	**29.65**	23.77	**10.81**	19.08	16.36	28.72	**30.57**	13.32	14.14	15.11
GG-AR	329.22	99.05	82.14	112.00	86.18	81.83	88.29	94.21	89.64	93.31	134.64	78.90
CT-AR	103.70	20.34	93.54	21.84	28.62	12.82	15.48	**16.21**	38.91	**11.37**	**11.23**	**9.42**
GG-VAR	239.82	89.18	70.64	126.52	106.13	59.96	62.41	53.19	66.30	60.75	177.72	85.95
CT-VAR	146.90	**17.46**	150.34	**15.88**	62.49	22.70	18.98	21.75	72.19	12.90	14.11	11.30
GG-ARIMA	100.00	103.78	81.66	115.78	84.61	86.89	90.15	95.18	91.35	94.81	134.74	79.50
CT-ARIMA	207.98	21.85	146.53	22.64	36.24	15.31	16.73	18.60	40.17	15.02	12.53	9.62
GG-ETS	312.09	104.06	81.43	117.56	85.38	87.14	90.71	94.94	91.90	95.39	134.58	79.68
CT-ETS	168.61	18.47	132.03	19.82	28.37	**12.48**	**12.85**	18.46	33.13	13.46	12.86	9.95

This table presents the Root Mean Squared Percentage Error (RMSPE) metric aggregated across all moneyness and term structure data points available for the 12 commodity options in our sample during the 500-day out-of-sample period (December 2014–December 2016). Results are presented for 1-, 5-, and 30-day-ahead forecasts. The lowest forecast errors for each commodity and forecast horizon are shown in bold.

procedure with 5000 bootstrap samples, based on Hansen et al. (2011) and Bernardi and Catania (2015), where the block length p is given by the maximum number of significant parameters obtained by fitting an AR(p) process to all the d_{ij} terms. For the hypothesis in (1), we utilize the test statistic:

$$T_{R,M} = \max_{i,j \in M} |t_{ij}|, \qquad (3)$$

where t_{ij} is defined in (2).

MCS is a sequential testing procedure that eliminates the worst model at each step until the hypothesis of EPA is accepted for all the models belonging to a set of superior models. The selection of the worst model is determined by an elimination rule that is consistent with the test statistic:

$$e_{R,M} = \underset{i \in M}{\mathrm{argmax}} \left\{ \sup_{j \in M} \frac{\bar{d}_{ij}}{\sqrt{\widehat{\mathrm{Var}}(\bar{d}_{ij})}} \right\}.$$

To summarize, the MCS procedure to obtain a superior set of models consists of the following steps:

(1) Set $M = M_0$.
(2) If the null hypothesis is accepted, then $M^* = M$; otherwise use the elimination rules defined in (3) to determine the worst model.
(3) Remove the worst model, and go to Step 2).

4. Out-of-sample results

We now move to an out-of-sample forecasting environment. We employ the RMSPE and MAPE measures outlined in Section 3.3 of

the full IVS for all 12 commodities in our sample. Results are presented in Tables 5 and 6, respectively. The first notable item is the large error size of the GG framework in this out-of-sample setting, a dynamic that is broadly consistent across the various commodity options and forecasting horizons. When we focus on the relative performances of the RT and CT models we note that the RT model leads to lower errors than the CT models for a number of Agricultural commodities; Cocoa, Cotton, and Soybean Oil. This observed dynamic holds at both short- and long-term forecasting horizons. In contrast, the Precious Metals in our sample exhibit lower forecasting errors using CT models; for example, the RMSPE for one-day-ahead Gold options forecasts of 14.14 and 5.69 for RT and CT-VAR, respectively.

In modeling the parameter estimates produced by the CT framework, a dynamic we observe is that, for a number of commodities, the VAR specification leads to lower RMSPEs and MAPEs at short forecasting horizons, with the simpler univariate AR specification producing lower errors at longer forecasting horizons. Examples of this are Natural Gas, Gold, and Silver commodity options. Focusing on Natural Gas for instance, we observe CT-AR and CT-VAR RMSPEs of 11.45 and 6.54, respectively, at a one-day-ahead forecasting horizon, in contrast with 11.37 and 12.90, respectively, at the 30-day-ahead horizon.[7]

Table 7 presents the SSR metric, which is the percentage of times the model correctly predicts the direction of the forecast. The metric is given across all moneyness and maturity

---

[7] Sensitivity to other forecasting horizons is explored with broadly similar results uncovered for h=2- and 10-day-ahead forecasting horizons. For brevity these results are excluded.

Table 6
Out-of-sample MAPE.

Forecasting	Agricultural							Energy			Metals	
Horizon (h)	Cocoa	Corn	Cotton	Soybean	Soybean Oil	Sugar	Wheat	Crude Oil	Heating Oil	Natural Gas	Gold	Silver
h = 1												
RT	**15.18**	19.37	**24.93**	19.08	**8.57**	16.83	12.49	26.21	28.58	10.65	11.39	12.17
GG-AR	319.33	68.12	36.25	82.07	78.28	39.10	43.32	30.26	22.49	31.35	120.15	56.76
CT-AR	65.14	4.72	68.68	4.98	15.47	8.08	4.86	6.41	34.18	8.42	4.47	4.66
GG-VAR	255.82	70.15	38.86	86.65	80.63	31.38	45.59	23.74	23.33	30.13	122.45	62.22
CT-VAR	66.56	**4.71**	82.79	**4.66**	11.15	**5.71**	**4.85**	**4.06**	27.73	**4.56**	**4.29**	**3.58**
GG-ARIMA	100.00	75.54	36.10	85.94	70.87	25.86	52.20	24.94	**16.65**	30.46	115.38	57.69
CT-ARIMA	173.39	4.77	143.36	4.79	17.78	6.13	5.05	4.74	31.72	6.50	4.39	3.70
GG-ETS	330.38	75.54	37.02	87.39	78.11	25.64	52.71	25.69	24.84	32.09	113.47	58.13
CT-ETS	96.73	**4.71**	86.07	4.74	12.98	6.34	5.00	4.76	24.53	6.53	4.32	3.77
h = 5												
RT	**15.18**	19.37	**24.93**	19.08	**8.57**	16.83	12.49	26.21	28.58	10.65	11.39	12.17
GG-AR	310.90	87.30	77.45	91.95	71.53	76.92	81.72	90.74	86.24	89.78	94.33	70.19
CT-AR	65.27	6.81	68.75	8.36	15.58	8.06	6.88	7.64	33.81	7.07	6.26	5.35
GG-VAR	215.00	70.13	37.27	85.89	80.07	34.09	44.23	23.99	25.64	30.44	126.86	61.85
CT-VAR	68.15	6.74	83.36	6.98	13.88	**6.61**	**6.66**	**6.75**	**23.65**	**5.90**	6.39	5.30
GG-ARIMA	100.00	90.22	76.71	94.69	71.84	84.74	81.01	92.90	90.25	90.58	94.81	70.28
CT-ARIMA	170.58	7.13	143.45	7.99	18.85	7.27	7.11	7.24	31.04	7.39	6.42	**5.02**
GG-ETS	287.44	90.45	76.41	95.59	71.33	84.89	81.34	92.64	88.06	90.62	94.96	70.33
CT-ETS	98.07	**6.72**	85.45	**6.85**	13.92	6.91	6.78	7.12	24.10	7.40	**6.22**	5.13
h = 30												
RT	**15.18**	19.37	**24.93**	19.08	**8.57**	16.83	12.49	26.21	28.58	10.65	11.39	12.17
GG-AR	310.89	85.36	77.77	89.22	71.32	76.78	81.39	90.54	87.34	90.08	95.13	70.25
CT-AR	64.88	15.52	67.93	17.38	16.54	10.61	12.08	**13.45**	31.35	**9.07**	**8.73**	**7.60**
GG-VAR	202.47	61.13	45.55	84.14	70.97	41.76	42.36	34.12	41.92	34.92	131.67	63.10
CT-VAR	89.31	**12.93**	110.02	**11.90**	25.74	13.62	11.72	16.55	41.51	9.91	11.36	9.00
GG-ARIMA	100.00	87.52	77.15	91.22	72.28	83.19	80.56	92.96	90.13	91.54	95.70	70.38
CT-ARIMA	159.50	17.00	141.90	17.32	21.31	11.71	13.89	14.03	27.41	12.21	9.83	7.90
GG-ETS	287.44	87.72	76.80	92.16	71.29	83.49	80.90	92.72	89.59	91.50	95.96	70.39
CT-ETS	105.94	14.05	92.63	14.56	16.60	**9.59**	**10.22**	14.02	**23.15**	10.76	10.16	8.14

This table presents the Mean Absolute Percentage Error (MAPE) metric across all moneyness and term structure data points available for the 12 commodity options in our sample during the 500-day out-of-sample period (December 2014-December 2016). Results are presented for 1-, 5-, and 30-day-ahead forecasts. The lowest forecast errors for each commodity and forecast horizon are shown in bold.

Table 7
Out-of-sample SSR.

Forecasting	Agricultural							Energy			Metals	
Horizon (h)	Cocoa	Corn	Cotton	Soybean	Soybean Oil	Sugar	Wheat	Crude Oil	Heating Oil	Natural Gas	Gold	Silver
h = 1												
RT	**55%**	54%	53%	52%	56%	52%	58%	52%	51%	53%	59%	57%
GG-AR	51%	54%	54%	53%	51%	51%	56%	59%	55%	58%	50%	52%
CT-AR	52%	94%	52%	84%	56%	59%	88%	69%	51%	64%	88%	74%
GG-VAR	52%	54%	**55%**	53%	51%	53%	55%	63%	59%	57%	50%	52%
CT-VAR	54%	**95%**	51%	91%	76%	78%	92%	94%	61%	85%	95%	87%
GG-ARIMA	50%	54%	**55%**	53%	51%	55%	55%	61%	56%	58%	51%	52%
CT-ARIMA	53%	93%	51%	87%	60%	65%	85%	81%	52%	71%	89%	77%
GG-ETS	50%	54%	**55%**	53%	51%	55%	55%	61%	56%	58%	51%	52%
CT-ETS	52%	94%	52%	90%	61%	63%	85%	81%	52%	71%	91%	77%
h = 30												
RT	57%	63%	56%	67%	70%	56%	70%	44%	49%	57%	73%	65%
GG-AR	49%	51%	50%	50%	54%	62%	49%	43%	50%	46%	48%	47%
CT-AR	49%	75%	56%	71%	71%	74%	80%	64%	51%	75%	86%	84%
GG-VAR	58%	57%	**59%**	59%	52%	56%	60%	72%	69%	69%	59%	58%
CT-VAR	**60%**	69%	54%	74%	**78%**	78%	76%	81%	**73%**	70%	82%	83%
GG-ARIMA	53%	50%	51%	49%	55%	61%	49%	42%	49%	46%	47%	47%
CT-ARIMA	55%	92%	50%	82%	74%	78%	83%	89%	53%	84%	89%	87%
GG-ETS	50%	50%	51%	49%	55%	61%	49%	42%	50%	46%	47%	47%
CT-ETS	56%	**98%**	57%	**96%**	76%	**81%**	**94%**	**94%**	54%	**86%**	**96%**	**91%**

This table presents the Sign Success Ratio (SSR) metric across all moneyness and term structure data points available for the 12 commodity options in our sample during the 500-day out-of-sample period (December 2014-December 2016). The SSR metric shows us what percentage of the time each model correctly predicts the direction of implied volatility change. Results are presented for 1-, and 30-day-ahead forecasts. All models are benchmarked versus the Chalamandaris and Tsekrekos (2011) fitted implied volatility surface from the previous period. What percentage of the time does it correctly predict the direction of implied volatility. The highest SSR for each commodity and forecast horizon are shown in bold.

Table 8
RMSE ratio by moneyness and maturity for CT-RW and GG-RW.

	Moneynesss						Maturity					
	OTM		ATM		ITM		Short Term		Medium Term		Long Term	
	CT-RW	GG-RW	CT-RW	GG-RW	CT-RW	GG-RW	CT-RW	GG-RW	CT-RW	GG-RW	CT-RW	GG-RW
$h = 1$												
RT	**0.19**	**0.25**	0.10	0.46	0.11	0.06	0.13	0.09	0.05	0.15	2.37	0.07
GG-AR	0.46	0.58	0.16	0.73	0.57	0.30	0.28	0.21	0.22	0.63	18.00	0.56
CT-AR	0.24	0.30	**0.07**	**0.33**	**0.07**	**0.04**	**0.11**	**0.08**	**0.04**	**0.11**	1.16	0.04
GG-VAR	0.44	0.56	0.15	0.67	0.57	0.30	0.27	0.20	0.22	0.63	17.88	0.55
CT-VAR	0.25	0.32	0.08	0.37	0.08	**0.04**	0.12	0.09	**0.04**	0.13	**1.12**	**0.03**
GG-ARIMA	0.44	0.56	0.11	0.50	0.54	0.29	0.22	0.16	0.21	0.59	17.96	0.56
CT-ARIMA	0.23	0.30	0.12	0.56	0.14	0.07	0.16	0.12	0.08	0.22	2.35	0.07
GG-ETS	0.45	0.58	0.17	0.76	0.57	0.30	0.28	0.21	0.23	0.64	18.07	0.56
CT-ETS	0.22	0.28	0.09	0.41	0.10	0.05	0.13	0.10	0.05	0.15	1.32	0.04
$h = 30$												
RT	**0.02**	**0.23**	**0.02**	0.18	0.03	0.11	**0.02**	**0.14**	0.55	0.12	1.47	0.09
GG-AR	0.09	1.01	0.10	0.72	0.13	0.51	0.08	0.51	2.42	0.53	9.85	0.58
CT-AR	0.03	0.32	**0.02**	**0.16**	**0.02**	**0.09**	**0.02**	**0.14**	**0.49**	**0.11**	**1.02**	**0.06**
GG-VAR	0.05	0.58	0.04	0.27	0.14	0.56	0.05	0.28	2.23	0.49	10.91	0.65
CT-VAR	0.04	0.42	0.03	0.23	0.03	0.13	0.03	0.19	0.69	0.15	1.44	0.09
GG-ARIMA	0.09	1.01	0.09	0.70	0.13	0.50	0.08	0.48	2.35	0.52	10.11	0.60
CT-ARIMA	0.03	0.33	0.03	0.24	0.04	0.15	0.03	0.19	0.83	0.18	1.69	0.10
GG-ETS	0.09	1.00	0.10	0.71	0.13	0.53	0.08	0.51	2.47	0.54	10.15	0.60
CT-ETS	0.03	0.31	0.03	0.20	0.03	0.13	0.03	0.17	0.63	0.14	1.39	0.08

This table presents the Root Mean Squared Error (RMSE) ratio of each of the time-series models listed to the benchmark Chalamandaris and Tsekrekos (2011) and Goncalves and Guidolin (2006) frameworks with no change random walk parameters specified, CT-RW and GG-RW, respectively. The ratios are averaged across all 12 commodity contracts (where applicable), and broken down by level of moneyness and maturity. For table brevity, results are presented for 1- and 30-day-ahead forecasts only, split into OTM ($<=95\%$), ATM (97.5% to 102.5%), ITM ($>=105\%$), Short- (1 to 3 months), Medium- (3 to 6 months) and Long-term ($>= 6$ months) contracts. The tests are conducted during the 500-day out-of-sample period (December 2014-December 2016) with results presented for 1- and 30-day ahead forecasts. Please note that there is not an equal number of implied volatility observations in each group. A value of less than 1 indicates that the respective model performs better than the CT-RW or GG-RW benchmark, and a value of greater than 1 indicates that the model performs worse than the CT-RW or GG-RW benchmark, based on the RMSE metric. The best performing model broken down by moneyness, maturity, forecast horizon and benchmark, is shown in bold.

observations for each commodity. The directional accuracy measure has an intuitive 50% random chance benchmark. It does not necessarily translate into superior magnitude metrics but could be used to help determine signals for trading strategies. Each model is benchmarked versus the no change CT-RW, with the under performing GG-RW benchmark excluded for brevity. Overall, we find that in the majority of instances, the time-series models directionally outperform the random walk benchmark. Time-series modeling of parameters leads to directional advantages with a large number of instances of CT models achieving over 80% SSRs. It is again observed that GG models do not perform as well as RT and CT, with SSRs as low as 42% observed when forecasting crude oil options using a GG-ETS at a 30-day-ahead horizon. We also see in general that it is more difficult to predict the direction of change at longer horizons. There are 22 instances of SSRs falling below the intuitive 50% benchmark at the 30-day-ahead forecasting horizon, including one CT model; the CT-AR forecast for Cocoa options with its 49% SSR. We can also see distinction in terms of the time-series model specified, with the multivariate CT-VAR working well at the one-day-ahead forecasting horizon but the univariate CT-ETS showing improved directional forecasts at the 30-day-ahead horizon.

To further identify the source of the forecasting performance of the various models we break the results down into moneyness and maturity segments of the IVS. First, we define the OTM options group as those with moneyness levels of less than or equal to 95%, ATM options group as those with moneyness between 97.5% and 102.5%, and ITM options group as those with moneyness of greater than or equal to 105%. Second, for maturities, we define short-term as between one and three months, medium-term as between three months and six months, and long-term (where applicable) as greater than or equal to 6 months. Using these moneyness and expiration splits we calculate errors for each group and present the results in Table 8. The errors are expressed as the ratio of the RMSE

for each time-series model to the RMSE of the benchmark random walk models aggregated across all 12 commodity options in our sample. A value of less than 1 indicates that a model outperforms the CT-RW or GG-RW benchmark, with a value of greater than 1 indicating that the model underperforms the benchmark. Overall, we see that the majority of the ratios are less than 1; indicating that the models outperform the benchmarks and that time-series modeling of the coefficients is effective.[8] We also see that the improvement over the benchmark is generally more pronounced at the longer 30-day-ahead forecasting horizon, with ratios as low as 0.02.

When we group the results by moneyness it is difficult to see any strong dynamics at play with, for example, the RT model showing relatively similar performance across all moneyness levels. More illuminating, however, is the maturity split. We find that while the time-series models perform well at predicting short- and medium- term maturity options, the CT-RW benchmark shows greater forecasting accuracy for long-term maturity options, indicated by all RMSE ratios being greater than 1. This highlights the difficulty of beating the random walk benchmark across the entire IVS and also indicates that time-series modeling of parameters is not as relevant for long-term maturity models as the explicit modeling of the term structure component.

Having focused on a direct comparison of forecasting errors to date, we now formally investigate cross-model superiority using

---

[8] The groups presented in Table 8 are very diverse in terms of their composition. This complicates a direct comparison of the ratio levels in a cross group evaluation. For instance, the groups range from short-term, that include 1 month maturity options, to long-term, that include 2 year maturity options, and are comprised of differing numbers of commodities and IV observations. For example, the short-term group has nine times more observations than the OTM group and the long-term group does not include four commodities from its calculation (as outlined in Table 1).

Table 9
Model confidence set results.

$T_{R,M}$	Agricultural							Energy			Metals	
Test Statistic	Cocoa	Corn	Cotton	Soybean	Soybean Oil	Sugar	Wheat	Crude Oil	Heating Oil	Natural Gas	Gold	Silver
$h = 1$												
RT	†		†		†							
CT-AR		†					†					
CT-VAR		†		†				†		†	†	†
CT-ETS		†				†					†	†
CT-ARIMA						†						†
CT-RW				†	†	†	†					†
GG-AR												
GG-VAR												
GG-ETS												
GG-ARIMA									†			
GG-RW												
$h = 5$												
RT	†		†		†			†				
CT-AR		†									†	
CT-VAR		†		†			†	†		†	†	
CT-ETS						†					†	
CT-ARIMA						†						†
CT-RW						†		†				
GG-AR												
GG-VAR												
GG-ETS												
GG-ARIMA												
GG-RW												
$h = 30$												
RT	†		†		†			†				
CT-AR								†		†	†	†
CT-VAR		†		†								
CT-ETS						†	†	†	†			
CT-ARIMA								†				
CT-RW						†		†	†			
GG-AR												
GG-VAR												
GG-ETS												
GG-ARIMA												
GG-RW										†		

This table presents the results of the model confidence set test of Hansen et al. (2011) based on the out-of-sample forecasts produced by various models across the full implied volatility surface for each of the 12 commodity options in our sample. It is assessed during the 500-day out-of-sample period (December 2014-December 2016), with results being presented for 1-, 5- and 30-day-ahead forecasts. The † symbol is used to indicate that a particular model resides in the superior set of models for a given commodity and forecast horizon.

the MCS of Hansen et al. (2011). The goal is to identify what model(s) reside in the superior set of models to establish if any statistically outperform the others. It also addresses the issue of possible false discoveries arising from the multiple comparisons problem. Overall, the results in Table 9 indicate that all three frameworks perform well in terms of identifying predictability, with even the GG-ARIMA residing in the superior set of models for one-day-ahead forecasts of Heating Oil.

Despite no random walk benchmark solely residing in the superior set of models for any commodity or forecasting horizon, the simultaneous appearance of the benchmarks alongside other models highlights how difficult they are to beat. More specifically, for the short one-day-ahead forecasting horizon the random walk benchmarks are statistically indistinguishable from the time-series process specified models across five commodities, Soybean, Soybean Oil, Sugar, Wheat and Silver. However, for longer forecasting horizons the strength of this dynamic subsides, with the benchmarks only appearing in the superior set of models for two commodities at the 5-day ahead horizon. This indicates that it takes a significant passing of time to see a formal benefit from the second step of our model construction in which we explicitly specify parameter evolution.

When we turn our attention to the commodity group split it can be seen that for Precious Metals the CT models are the only models to exhibit formal outperformance. Conversely, the RT model performs best when focusing on Agricultural options, despite results here being more mixed, with the RT model appearing in the superior set of models for Cocoa, Cotton, and Soybean Oil across all horizons. The CT models reside in the superior set of models for Wheat and Soybean, commodities associated with extreme average convenience yield slopes, indicating that those commodities with steeper convenience yield curves are more accurately forecasted using the explicit Nelson-Siegel CT framework.

5. Conclusion

Inter-temporal predictability in the IVSs of equity, index, foreign exchange, and interest rate options has been uncovered using a small number of models. Motivated by the growing popularity of commodities and the increasing correlation between commodities and other markets, this paper employs these existing frameworks to ascertain if they can both characterize and forecast the IVSs observed for popular commodity options. IV smile asymmetry is observed across commodities with fear relating to a supply-side shock being the most prevalent characteristic for agricultural options. In a cross-model comparison of the multiple parametric factor and machine learning models considered, no model systematically outperforms its competitors. For this reason, we draw inferences from what models demonstrate outperformance for specific groups of commodities.

Focusing on out-of-sample testing the most striking conclusion is that when modeling Energy and Precious Metals options the CT specification produces the most accurate forecasts. This aligns with Doran and Ronn (2008) who highlight that there is a significant term structure of volatility present in Energy markets even at short-term maturities. The finding is at odds with what Chalamandaris and Tsekrekos (2011) conclude for foreign exchange options; that a linear approximation of the term structure is sufficient up to an option expiry of 12 months ahead. Precious Metals, however, is a class of commodities that exhibits relatively low levels of IV for longer maturity dates, in contrast to what is observed for other commodities. We hypothesize that this converse maturity dynamic is caused by safe haven properties associated with Precious Metals; properties not commonly attributed to Agricultural or Energy commodities.[9] Again, the explicit modeling of the term structure using the Nelson-Siegel factors leads to CT exhibiting improved predictability for long-term maturities. We also find that cross-model outperformance is less stark for Agricultural options. However, the RT model exhibits the most promising results, in particular for commodities with flatter underlying convenience yield curves. Further, we conclude that the simpler specification of the univariate models over the multivariate VAR model leads, on average, to, superior forecasting results at longer forecasting horizons, as also established for foreign exchange options in Chalamandaris and Tsekrekos (2014).

The mapping between the approaches we consider, and the optimizing behavior of a market participant who experiences incomplete information in a rational asset pricing framework, is complex. Despite this, the existence of a precisely estimable dynamic relationship could still be explained by general equilibrium models, whereby a learning process followed by investors leads to observed inter-temporal IVS dependence, as hypothesized by Bernales and Guidolin (2015). Additionally, despite not being the focus of this performance paper, the pockets of predictability uncovered could be exploited economically in future work. However, this would involve the implementation of bespoke hedging or trading strategies dependent on each market participant's specific option focus and portfolio requirements.[10] Finally, as a potential avenue for future research, a forecast combination approach along the lines of Bates and Granger (1969) and Gogolin and Kearney (2016) may yield a further refinement of forecast accuracy using the superior approaches identified here.

Acknowledgements

---

[9] Baur and McDermott (2010) empirically demonstrate the short-term safe haven attribute associated with gold.

[10] Another reason we refrain from presenting a trading strategy is the inability to realistically simulate a live market environment. Previous studies rely on idiosyncratic market assumptions, and ignore implementation issues such as liquidity, strategy drawdowns, margin calls, bid-ask spread considerations, and microstructure effects that might distort any calculated profits.

References

Adams, Z., Glück, T., 2015. Financialization in commodity markets: a passing trend or the new normal? J. Banking Finance 60, 93–111.

Alexander, C., 2001. Market Models: A Guide to Financial Data Analysis. John Wiley & Sons, New York.

Askari, H., Krichene, N., 2008. Oil price dynamics (2002–2006). Energy Econ. 30 (5), 2134–2153.

Audrino, F., Colangelo, D., 2010. Semi-parametric forecasts of the implied volatility surface using regression trees. Stat. Comput. 20 (4), 421–434.

Badshah, I.U., Frijns, B., Tourani-Rad, A., 2013. Contemporaneous spill-over among equity, gold, and exchange rate implied volatility indices. J. Futures Markets 33 (6), 555–572.

Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. Oper. Res. Q. 20 (4), 451–468.

Baur, D.G., McDermott, T.K., 2010. Is gold a safe haven? International evidence. J. Banking Finance 34 (8), 1886–1898.

Bedendo, M., Hodges, S.D., 2009. The dynamics of the volatility skew: A Kalman filter approach. J. Banking Finance 33 (6), 1156–1165.

Bernales, A., Guidolin, M., 2014. Can we forecast the implied volatility surface dynamics of equity options? predictability and economic value tests. J. Bank. Finance 46, 326–342.

Bernales, A., Guidolin, M., 2015. Learning to smile: can rational learning explain predictable dynamics in the implied volatility surface? J. Financ. Markets 26, 1–37.

Bernardi, M., Catania, L., 2015. The model confidence set package for R. Working paper. University of Rome "Tor Vergata".

Bessembinder, H., Chan, K., 1992. Time-varying risk premia and forecastable returns in futures markets. J. Financ. Econ. 32 (2), 169–193.

Bhansali, R., 1999. for multistep prediction of a time series: A review. Asymptotics, nonparametrics, and time series 201.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. J.Political Econ. 81 (3), 637–654.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA.

Chalamandaris, G., Tsekrekos, A.E., 2010. Predictable dynamics in implied volatility surfaces from OTC currency options. J. Bank. Finance 34 (6), 1175–1188.

Chalamandaris, G., Tsekrekos, A.E., 2011. How important is the term structure in implied volatility surface modeling? evidence from foreign exchange options. J. Int. Money Finance 30 (4), 623–640.

Chalamandaris, G., Tsekrekos, A.E., 2014. Predictability in implied volatility surfaces: Evidence from the Euro OTC FX market. Eur. J. Finance 20 (1), 33–58.

Cheng, I.-H., Xiong, W., 2014. Financialization of commodity markets. Annu. Rev. Financ. Econ. 6 (1), 419–441.

Cont, R., Da Fonseca, J., 2002. Dynamics of implied volatility surfaces. Quant. Finance 2 (1), 45–60.

Corrado, C.J., Miller, T.W., 2006. Estimating expected excess returns using historical and option-implied volatility. J. Financ. Res. 29 (1), 95–112.

Cox, D.R., 1961. Prediction by exponentially weighted moving averages and related methods. J. R. Stat. Soc. 23 (2), 414–422.

David, A., Veronesi, P., 2000. Option prices with uncertain fundamentals: Theory and evidence on the dynamics of implied volatilities. CRSP Working Paper. University of Calgary.

Deuskar, P., Gupta, A., Subrahmanyam, M.G., 2008. The economic determinants of interest rate option smiles. J. Bank. Finance 32 (5), 714–728.

Diebold, F.X., Li, C., 2006. Forecasting the term structure of government bond yields. J. Econom. 130 (2), 337–364.

Domanski, D., Heath, A., 2007. Financial investors and commodity markets. BIS Q. Rev. March, 53–67.

Doran, J.S., Ronn, E.I., 2008. Computing the market price of volatility risk in the energy commodity markets. J. Banking Finance 32 (12), 2541–2552.

Dumas, B., Fleming, J., Whaley, R.E., 1998. Implied volatility functions: empirical tests. J. Finance 53 (6), 2059–2106.

Esposito, F., Malerba, D., Semeraro, G., 1997. A comparative analysis of methods for pruning decision trees. IEEE Trans. Pattern Anal. Mach. Intell. 19 (5), 476–491.

Fama, E.F., 1965. The behavior of stock-market prices. J. Bus. 38 (1), 34–105.

Fattouh, B., Mahadeva, L., 2014. Causes and implications of shifts in financial participation in commodity markets. J. Futures Markets 34 (8), 757–787.

Fengler, M.R., Härdle, W.K., Mammen, E., 2007. A semiparametric factor model for implied volatility surface dynamics. J. Financ. Econom. 5 (2), 189–218.

Fengler, M.R., Härdle, W.K., Villa, C., 2003. The dynamics of implied volatilities: a common principal components approach. Rev. Derivatives Res. 6 (3), 179–202.

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

Garcia, R., Luger, R., Renault, E., 2003. Empirical assessment of an intertemporal option pricing model with latent variables. J. Econom. 116 (1), 49–83.

Garvey, J.F., Gallagher, L.A., 2012. The realised-implied volatility relationship: recent empirical evidence from FTSE-100 stocks. J. Forecast 31 (7), 639–660.

Giot, P., 2003. The information content of implied volatility in agricultural commodity markets. J. Futures Markets 23 (5), 441–454.

Gogolin, F., Kearney, F., 2016. Does speculation impact what factors determine oil futures prices? Econ. Lett. 144, 119–122.

Goncalves, S., Guidolin, M., 2006. Predictable dynamics in the S&P 500 index options implied volatility surface. J. Bus. 79 (3), 1591–1635.

Gorton, G., Rouwenhorst, K.G., 2006. Facts and fantasies about commodity futures (digest summary). Financ. Anal. J. 62 (2), 47–68.

Guidolin, M., Timmermann, A., 2003. Option prices under bayesian learning: implied volatility dynamics and predictive densities. J. Econ. Dyn. Control 27 (5), 717–769.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.

Hibbert, A.M., Daigler, R.T., Dupoyet, B., 2008. A behavioral explanation for the negative asymmetric return–volatility relation. J. Banking Finance 32 (10), 2254–2266.

Ing, C.-K., 2003. Multistep prediction in autoregressive processes. Econ. Theory 19 (2), 254–279.

Irwin, S.H., Sanders, D.R., 2011. Index funds, financialization, and commodity futures markets. Appl. Econ. Perspect. Policy 33 (1), 1–31.

Kearney, F., Murphy, F., Cummins, M., 2015. An analysis of implied volatility jump dynamics: novel functional data representation in crude oil markets. North Am. J. Econ. Finance 33, 199–216.

Klein, L., 1968. An essay on the theory of economic prediction yrjo jahnssonin saa-tio.

Konstantinidi, E., Skiadopoulos, G., Tzagkaraki, E., 2008. Can the evolution of implied volatility be forecasted? evidence from european and US implied volatility indices. J. Banking Finance 32 (11), 2401–2411.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. J. Econom. 54 (1–3), 159–178.

Liu, P., Tang, K., 2011. The stochastic behavior of commodity prices with heteroskedasticity in the convenience yield. J. Empirical Finance 18 (2), 211–224.

Mandelbrot, B., 1963. New methods in statistical economics. J. Political Econ. 71 (5), 421–440.

Muzzioli, S., 2010. Option-based forecasts of volatility: an empirical study in the DAX-index options market. Eur. J. Finance 16 (6), 561–586.

Nelson, C.R., Siegel, A.F., 1987. Parsimonious modeling of yield curves. J. Business 60 (4). 473–89.

Peña, I., Rubio, G., Serna, G., 1999. Why do we smile? on the determinants of the implied volatility function. J. Banking Finance 23 (8), 1151–1179.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Skiadopoulos, G., Hodges, S., Clewlow, L., 2000. The dynamics of the S&P 500 implied volatility surface. Rev. Derivatives Res. 3 (3), 263–282.

Taylor, S.J., Yadav, P.K., Zhang, Y., 2010. The information content of implied volatilities and model-free volatility expectations: evidence from options written on individual stocks. J. Banking Finance 34 (4), 871–881.

Therneau, T., Atkinson, B., 2018. rpart: Recursive Partitioning and Regression Trees. R package version 4.1–13.