Predictable Dynamics in the S&P 500 Index Options Implied Volatility Surface

Author(s): Sílvia Gonçalves and Massimo Guidolin

Source: *The Journal of Business*, Vol. 79, No. 3 (May 2006), pp. 1591-1635
Published by: The University of Chicago Press
Stable URL: https://www.jstor.org/stable/10.1086/500686

Sílvia Gonçalves
Department of Economics, CIREQ and CIRANO, Université de Montréal

Massimo Guidolin
Federal Reserve Bank of St. Louis

Predictable Dynamics in the S&P 500 Index Options Implied Volatility Surface*

Recent evidence suggests that the parameters characterizing the implied volatility surface (IVS) in option prices are unstable. We study whether the resulting predictability patterns may be exploited. In a first stage we model the surface along cross-sectional moneyness and maturity dimensions. In a second stage we model the dynamics of the first-stage coefficients. We find that the movements of the S&P 500 IVS are highly predictable. Whereas profitable delta-hedged positions can be set up under selective trading rules, profits disappear when we increase transaction costs and trade on wide segments of the IVS.

I. Introduction

Volatilities implicit in observed option prices are often used to gain information on expected market volatility (see, e.g., Poterba and Summers 1986; Jorion 1995; Christensen and Prabhala 1998; Fleming 1998). Therefore, accurate forecasts of implied volatilities may be valuable in many situations. For instance, in derivative pricing applications, volatility characterizes the beliefs of market participants and hence is intimately related to the fundamental pricing measure. Implied volatilities are commonly used by practitioners for option pricing purposes and risk management.

Implied volatilities are typically found by first equating observed option prices to Black-Scholes (1973) theoretical prices and then solving for the unknown volatility parameter, given data on the option contracts and the underlying asset prices. Contrary to the Black-Scholes assumption of constant volatility,

* We would like to thank Peter Christoffersen, Steven Clark, Patrick Dennis, Kris Jacobs, and seminar participants at the 2003 Midwest Finance Association meetings for helpful comments. We are especially grateful to an anonymous referee, René Garcia, Rob Engle, and Hal White for their comments and suggestions at various stages of this project, which greatly improved the paper. Gonçalves acknowledges financial support from the Institut de Finance Mathématique de Montréal. Contact the corresponding author, Massimo Guidolin, at Massimo.Guidolin@stls.frb.org.

[*Journal of Business*, 2006, vol. 79, no. 3]

0021-9398/2006/7903-0019$10.00

implied volatilities tend to systematically vary with the options strike price and date of expiration, giving rise to an *implied volatility surface* (IVS). For instance, Canina and Figlewski (1993) and Rubinstein (1994) show that when plotted against moneyness (the ratio between the strike price and the underlying spot price), implied volatilities describe either an asymmetric smile or a smirk. Campa and Chang (1995) show that implied volatilities are a function of time to expiration. Furthermore, the IVS is known to dynamically change over time, in response to news affecting investors' beliefs and portfolios.

Practitioners have long tried to exploit the predictability in the IVS. The usual approach consists of fitting linear models linking implied volatility to time to maturity and moneyness, for each available cross section of option contracts at a point in time. The empirical evidence suggests that the estimated parameters of such models are highly unstable over time. For instance, Dumas, Fleming, and Whaley (1998) propose a model in which implied volatilities are a function of the strike price and time to maturity. They observe that the coefficients estimated on weekly cross sections of S&P 500 option prices are highly unstable. Christoffersen and Jacobs (2004) report identical results. Similarly, Heston and Nandi (2000) estimate a moving window nonlinear GARCH(1, 1) (generalized autoregressive conditional heteroskedasticity) and show that some of the coefficients are unstable. To explain the superior performance of their GARCH pricing model, Heston and Nandi stress the ability of the GARCH framework to exploit the information on path dependency in volatility contained in the spot S&P 500 index. Thus time variation of the S&P 500 IVS matters for option pricing purposes.

In this paper we propose a modeling approach for the time-series properties of the S&P 500 index options IVS. Our approach delivers easy-to-compute forecasts of implied volatilities for any strike price or maturity level. This is in contrast to the existing literature, which has focused on either modeling the cross section of the implied volatilities, ignoring the time-series dimension, or modeling the time-series properties of an arbitrarily chosen point on the IVS, that is, the volatility implicit in contracts with a given moneyness and/or time to expiration. To the best of our knowledge, we are the first to jointly model the cross-sectional features and the dynamics of the IVS for stock index options.

We ask the following questions: Given the evidence of time variation in the IVS, is there any gain from explicitly modeling its time-series properties? In particular, can such an effort improve our ability to forecast volatility and hence option prices? To answer these questions, we combine a cross-sectional approach to fitting the IVS similar to Dumas et al. (1998) with the application of vector autoregression (VAR) models to the (multivariate) time series of estimated cross-sectional coefficients. Therefore, our approach is a simple extension of the Dumas et al. approach in which modeling occurs in two distinct stages. In a first stage, we fit daily cross-sectional models that describe implied volatilities as a function of moneyness and time to maturity. Consistently with the previous literature, we report evidence of structure in the S&P

500 IVS and find that a simple model linear in the coefficients and nonlinear in moneyness and time to maturity achieves an excellent fit. The documented instability of the estimated cross-sectional coefficients motivates our second step: we fit time-series models of a VAR type to capture the presence of time variation in the first-stage estimated coefficients. We find that the fit provided by this class of models is remarkable and describes a law of motion for the IVS that conforms to a number of stylized facts.

To assess the performance of the proposed IVS modeling approach, we use both statistical and economic criteria. First, we study its ability to correctly predict the level and the direction of change of one-day-ahead implied volatility. We find that our models achieve good accuracy, both in absolute terms and relatively to a few natural benchmarks, such as random walks for implied volatilities and Heston and Nandi's (2000) NGARCH(1, 1). Second, we evaluate the ability of our forecasts to support portfolio decisions. We find that the performance of our two-stage dynamic IVS models at predicting one-step-ahead option prices is satisfactory. We then simulate out-of-sample delta-hedged trading strategies based on deviations of volatilities implicit in observed option prices from model-based predicted volatilities with a constant, fixed investment of $1,000 per day. The simulated strategies that rely on two-stage IVS models generate positive and statistically significant out-of-sample returns when low to moderate transaction costs are imputed on all traded (option and stock) contracts. These profits are abnormal as signaled by Sharpe ratios in excess of benchmarks such as buying and holding the S&P 500 index; that is, they are hardly rationalizable in the light of the risk absorbed. Importantly, our finding of abnormal profitability appears to be fairly robust to the adoption of performance measures that take into account nonnormalities of the empirical distribution of profits and to imputing transaction costs that account for the presence of bid-ask spreads. In particular, our approach is most accurate (hence profitable) on specific segments of the IVS, mainly out-of-the-money and short- to medium-term contracts.

These results turn mixed when higher transaction costs and/or trading strategies that imply trades on large numbers of contracts along the entire IVS are employed in calculating profits. We conclude that predictability in the structure of the S&P 500 IVS is strong in statistical terms and ought to be taken into account to improve both volatility forecasting and portfolio decisions. However, such predictability patterns hardly represent outright rejections of the tenet that deep and sophisticated capital markets such as the S&P 500 index options market are informationally efficient. In particular, even when filters are applied to make our trading rules rather selective in terms of the ex ante expected profits per trade, we find that as soon as transaction costs are raised to the levels that are likely to be faced by small (retail) speculators, all profits disappear.

The option pricing literature has devoted many efforts to propose pricing models consistent with the stylized facts derived in the empirical literature, of which the implied volatility surface is probably the best-known example.

Models featuring stochastic volatility, jumps in returns and volatility, and the existence of leverage effects (i.e., a nonzero covariance between returns and volatility) are popular approaches (see Garcia, Ghysels, and Renault [2005] for a review of the literature). More recently, several papers have proposed models relying on a general equilibrium framework to investigate the economics of these stylized facts.[1] For instance, David and Veronesi (2002) propose a dynamic asset pricing model in which the drift of the dividend growth rate follows a regime-switching process. Investors' uncertainty about the current state of the economy endogenously creates stochastic volatility and leverage, thus giving rise to an IVS. Because investors' uncertainty evolves over time and is persistent, this model induces predictability in the IVS. Similarly, Guidolin and Timmermann (2003) propose a general equilibrium model in which dividends evolve on a binomial lattice. Investors' learning is found to generate asymmetric skews and systematic patterns in the IVS. The changing beliefs of investors within a rational learning scheme imply dynamic restrictions on how the IVS evolves over time. Finally, in Garcia, Luger, and Renault's (2003) utility-based option pricing model, investors learn about the drift and volatility regime of the joint process describing returns and the stochastic discount factor, modeled as a bivariate regime-switching model. Under their assumptions, the IVS depends on an unobservable latent variable characterizing the regime of the economy. Persistence of the process describing this latent variable implies predictability of the IVS. These models are examples of equilibrium-based models that generate time-varying implied volatility patterns consistent with those observed in the data. We view our approach as a reduced-form approach to model the time variation in the IVS that could have been generated by any of these models. As is often the case in forecasting, a simple reduced-form approach such as ours is able to efficiently exploit the predictability generated by more sophisticated models.

A few existing papers are closely related to ours. Harvey and Whaley (1992) study the time variation in volatility implied by the S&P 100 index option prices for short-term, nearest at-the-money contracts. They test the hypothesis that volatility changes are unpredictable on the basis of regressions of the changes in implied volatility on information variables that include day-of-the-week dummy variables, lagged implied volatilities, interest rate measures, and the lagged index return. They conclude that one-day-ahead volatility forecasts are statistically quite precise but do not help devising profitable trading strategies once transaction costs are taken into account. We depart from Harvey and Whaley's analysis in several ways. First, we look at European-style S&P

1. Bakshi and Chen (1997) derive option pricing results in a general equilibrium model with a representative agent. In equilibrium, both interest rates and stock returns are stochastic, with the latter having a systematic and an idiosyncratic volatility component. They show that this model is able to reproduce various shapes of the smile, although the dynamic properties of the IVS are left unexplored.

500 index options. Second, we do not reduce the IVS to a single point (at-the-money, short-term) and instead model the dynamics of the entire surface.

Noh, Engle, and Kane (1994) compare mean daily trading profits for two alternative forecasting models of the S&P 500 volatility: a GARCH(1, 1) model (with calendar adjustments) and a regression model applied to daily changes in weighted implied volatilities. Trading strategies employ closest-at-the-money, short-term straddles. They report the superior performance of GARCH one-day-ahead volatility forecasts at delivering profitable trading strategies, even after accounting for transaction costs of magnitude similar to those assumed in our paper. Although Noh et al.'s implied volatility–based model has a time-series dimension, a generalized least squares (GLS) procedure (Day and Lewis 1988) is applied to compress the entire daily IVS in a single, volume-weighted volatility index, so that the rich cross-sectional nature of the IVS is lost. Instead, we evaluate our dynamic models over the entire IVS and thus consider trading in option contracts of several alternative moneyness levels and expiration dates. We also adopt a GARCH-type model as a benchmark but estimate it on options data (cf. Heston and Nandi 2000), whereas Noh et al. (1994) obtain quasi–maximum likelihood estimates from stock returns data.

Diebold and Li (2006) use a two-step approach similar to ours in an unrelated application to modeling and forecasting the yield curve. In a first step, they apply a variation of the Nelson-Siegel exponential component framework to model the yield curve derived from U.S. government bond prices at the cross-sectional level. In a second step, they propose autoregressive integrated moving average–type models for the coefficients estimated in the first step. Finally, Rosenberg and Engle (2002) propose a flexible method to estimate the pricing kernel. Their empirical results suggest that the shape of the pricing kernel changes over time. To model this time variation, Rosenberg and Engle postulate a VAR model for the parameters that enter the pricing kernel at each point in time. Using hedging performance as an indicator of accuracy, they show that their time-varying model of the pricing kernel outperforms a time-invariant model, and they thus conclude that time variation in the pricing kernel is economically important.

The plan of the paper is as follows. Section II describes the data and a few stylized facts concerning the time variation of the S&P 500 IVS. We estimate a cross-sectional model of the IVS and discuss the estimation results. In Section III, we propose and estimate VAR-type models for the estimated parameters obtained in the first stage. Section IV is devoted to out-of-sample statistical measures of prediction accuracy, and Section V examines performance in terms of simulated trading profits under a variety of assumptions concerning the structure of transaction costs. Section VI discusses some robustness checks that help us qualify the extent of the IVS predictability previously isolated. Section VII presents conclusions.

II. The Implied Volatility Surface

A. The Data

We use a sample of daily closing prices for S&P 500 index options (calls and puts) from the Chicago Board Options Exchange covering the period January 3, 1992–June 28, 1996. S&P 500 index options are European-style and expire the third Friday of each calendar month. Each day up to six contracts are traded, with a maximum expiration of one year. We use trading days to calculate days to expiration (DTE) throughout. Given maturity, prices for a number of strikes are available. The data set is completed by observations on the underlying index (S) and T-bill yields (r), interpolated to match the maturity of each option contract, proxying for the risk-free rate.

For European options, the spot price of the underlying bond must be adjusted for the payment of discrete dividends by the stocks in the S&P 500 basket. As in Bakshi, Cao, and Chen (1997) and Dumas et al. (1998), we assume these cash flows to be perfectly anticipated by market participants. For each contract traded on day t with days to expiration DTE, we first calculate the present value D_t of all dividends paid on S&P 500 stocks between t and t + DTE. We then subtract D_t from the time t synchronous observation on the spot index to obtain the dividend-adjusted stock price. Data on S&P 500 cash dividends are collected from the S&P 500 *Information Bulletin*.

Five exclusionary criteria are applied. First, we exclude thinly traded options, with an arbitrary cutoff chosen at 100 contracts per day. Second, we exclude all options that violate at least one of a number of basic no-arbitrage conditions. Violations of these conditions presumably arise from misrecordings and are unlikely to derive from thick trading. Third, we discard data for contracts with fewer than six trading days to maturity since their prices are noisy,[2] possibly containing liquidity-related biases, and because they contain very little information on the time dimension of the IVS. We also exclude all contracts with more than one year to maturity. Fourth, we follow Dumas et al. (1998) and Heston and Nandi (2000) by excluding options with absolute moneyness in excess of 10%, with moneyness defined as $m \equiv$ (strike price/forward price) $-$ 1.[3] Fifth and finally, as in Bakshi et al. (1997), we exclude contracts with price lower than \$3/8 to mitigate the impact of price discreteness on the IVS structure. The filtered data correspond to a total of 48,192 observations, of which 20,615 refer to call contracts and 27,577 to puts. The average number of options per day is 41 with a minimum of five and a maximum of 63.

Table 1 reports summary statistics for implied volatilities computed by the Black-Scholes formula adjusted for dividend payments. We divide the data into several categories according to moneyness and time to maturity. A put

2. See Sec. VI and Hentschel (2003) for measurement error issues related to the calculation (estimation) of implied volatilities.
3. The forward price is defined as $\exp(r\tau)S$, where τ is time to maturity measured as a fraction of the year.

TABLE 1 Summary Statistics for Implied Volatilities by Maturity and Moneyness

	Short-Term		Medium-Term		Long-Term			
	Call	Put	Call	Put	Call	Put	Total	%
DOTM:								
Observations	146	2,550	771	2,423	442	825	7,157	14.85
Average IV	.124	.185	.109	.164	.117	.156	.163	
Standard deviation IV	.014	.027	.015	.018	.015	.015	.032	
OTM:								
Observations	4,608	7,366	3,105	4,515	606	1,233	21,433	44.47
Average IV	.105	.145	.109	.139	.126	.143	.129	
Standard deviation IV	.018	.025	.018	.019	.017	.015	.027	
ATM:								
Observations	3,187	3,186	1,804	1,774	290	310	10,551	21.89
Average IV	.113	.117	.122	.124	.135	.130	.118	
Standard deviation IV	.019	.022	.018	.018	.015	.018	.020	
ITM:								
Observations	3,162	1,896	1,474	815	388	379	8,114	16.84
Average IV	.135	.121	.132	.117	.146	.122	.129	
Standard deviation IV	.036	.035	.022	.023	.019	.018	.031	
DITM:								
Observations	312	71	218	137	102	97	937	1.94
Average IV	.220	.213	.162	.116	.160	.104	.171	
Standard deviation IV	.078	.086	.035	.028	.022	.018	.068	
Total:								
Observations	26,484 (55.0%)		17,036 (35.4%)		4,672 (9.6%)		48,192	100
Average IV	.131		.133		.139		.132	
Standard deviation IV	.035		.023		.020		.032	

NOTE.—The sample period is January 3, 1992–June 28, 1996, for a total of 48,192 observations (after applying exclusionary criteria). Moneyness (m) is defined as the ratio of the contract strike to forward spot price minus one. DOTM denotes deep-out-of-the-money ($m < -0.06$ for puts and $m > 0.06$ for calls); OTM, out-of-the money ($-0.06 < m \leq -0.01$ for puts and $0.01 < m \leq 0.06$ for calls); ATM, at-the-money ($-0.01 \leq m \leq 0.01$); ITM, in-the-money ($0.01 \leq m < 0.06$ for puts and $-0.06 < m \leq -0.01$ for calls); and DITM, deep-in-the-money ($m > 0.06$ for puts and $m < -0.06$ for calls). Short-term contracts have less then 60 (trading) days to maturity, medium-term contracts time to maturity in the interval [60, 180] days, and long-term contracts have more than 180 days to expiration.

contract is said to be *deep in the money* (DITM) if $m > 0.06$, *in the money* (ITM) if $0.06 \geq m > 0.01$, *at the money* (ATM) if $0.01 \geq m \geq -0.01$, *out of the money* (OTM) if $-0.01 > m \geq -0.06$, and *deep out of the money* (DOTM) if $-0.06 > m$. Equivalent definitions apply to calls, with identical bounds but with m replaced with $-m$ in the inequalities. The classification based on time to expiration follows Bakshi et al. (1997): an option contract is short-term if DTE < 60 days, medium-term if $60 \leq$ DTE ≤ 180, and long-term if DTE $>$ 180 days. Roughly 61% of the data is represented by short- and medium-term OTM and ATM contracts. DITM and long-term contracts are grossly underrepresented.

Table 1 provides evidence on the heterogeneity characterizing S&P 500 implied volatilities as a function of moneyness and time to expiration. For call options, implied volatilities describe an asymmetric smile for short-term contracts and perfect skews (i.e., volatilities increase moving from DOTM to DITM) for medium- and long-term contracts. Similar patterns are observed for puts, with the difference that volatilities decrease when moving from

DOTM to DITM: protective (DOTM) puts yield higher prices and thus higher volatilities. Table 1 also shows that the smile is influenced by time to maturity: implicit volatilities are increasing in DTE for ATM contracts (calls and puts), whereas they are decreasing in DTE for DOTM puts and DITM calls.

B. *Fitting the Implied Volatility Surface*

In this subsection, we fit an implied volatility model to each cross section of options available each day in our sample. Given the evidence presented above, two factors seem determinant in modeling the implied volatilities for each daily cross section of option contracts: moneyness and time to expiration. In a second stage, we will model and forecast the estimated volatility function coefficients.

Let σ_i denote the Black-Scholes implied volatility for contract i, with time to maturity τ_i (measured as a fraction of the year, i.e. $\tau_i \equiv \mathrm{DTE}_i/252$) and strike price K_i. We consider the following time-adjusted measure of moneyness:[4]

$$M_i \equiv \frac{\ln [K_i/\exp(r\tau_i)S]}{\sqrt{\tau_i}}.$$

The term M_i is positive for OTM calls (ITM puts) and negative for ITM calls (OTM puts).

Each day we estimate the following cross-sectional model for the IVS by ordinary least squares (OLS):

$$\ln \sigma_i = \beta_0 + \beta_1 M_i + \beta_2 M_i^2 + \beta_3 \tau_i + \beta_4 (M_i \times \tau_i) + \varepsilon_i, \tag{1}$$

where ε_i is the random error term, $i = 1, \ldots, N$, and N is the number of options available in each daily cross section. We use log implied volatility as the dependent variable. This has the advantage of always producing nonnegative implied volatilities. We estimated a variety of other specifications (see Peña, Rubio, and Serna 1999). They included models in which the IVS was a function only of moneyness (either a linear or a quadratic function, or a stepwise linear function of moneyness) and models using both the moneyness and time to expiration variables, included in the regression in the logarithmic or quadratic form, without any interaction term. We omit the estimation outputs to save space and because these alternative models showed a worse fit (as measured by their adjusted R^2's) than (1).

For each day in our sample, we estimate $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)'$ by OLS and obtain a vector $\hat{\boldsymbol{\beta}}$ of daily estimates.[5] To assess the in-sample fit of our

4. Gross and Waltner (1995) and Tompkins (2001) also use a similar measure of moneyness. According to this measure, the longer the time to maturity of an option, the larger the difference should be between the strike price and the forward stock price in order for it to achieve the same normalized moneyness as a short-term option.

5. As recently remarked by Hentschel (2003), measurement errors may introduce heteroskedasticity and autocorrelation in ε_i, making the OLS estimator inefficient. In Sec. VI, we apply Hentschel's feasible GLS estimator as a robustness check.

cross-sectional model, we present in table 2 summary statistics for the adjusted R^2 as well as for the roor mean squared error (RMSE) of implied volatilities. On average, the value of $\bar{R}^2$ is equal to 81%, with a minimum value of 1.1% and a maximum value of 99%. The time series of the daily values of the adjusted R^2 and RMSE of implied volatilities (not reported) show that there is considerable time variation in the explanatory power of equation (1). The functional form implied by this model is nevertheless capable of replicating various IVS shapes, including skews and smiles as well as nonmonotone shapes with respect to time to expiration. In the upper panel of figure 1 we plot the implied "average" fitted IVS model (i.e., the fitted model evaluated at the mean values of the estimated coefficients obtained from table 2) as a function of moneyness and time to maturity. For comparison, in the lower panel of the same figure we present the average actual implied volatilities for each of the 15 categories in table 1; that is, we plot the average volatility in correspondence to the midpoint moneyness and time to maturity characterizing each of the table's cells. The two plots show close agreement between raw and fitted implied volatilities.

Figure 2 plots the time series of the daily estimates $\hat{\boldsymbol{\beta}}$. Figure 2 shows that the shape of the S&P 500 IVS is highly unstable over time, both in the moneyness and in the time to maturity dimensions. Table 2 and figure 3 contain some descriptive statistics for the estimated coefficients. In particular, the Ljung-Box (LB) statistics at lags 1 and 10 indicate that there is significant autocorrelation for all coefficients (one exception is $\hat{\beta}_4$), in both levels and squares, suggesting that some structure exists in the dynamics of the estimated coefficients. Figure 3 plots the auto- and cross-correlations for the time series of OLS estimates. The cross-correlograms between pairs of estimated coefficients show strong association between them, at both leads and lags as well as contemporaneously. This suggests the appropriateness of multivariate models for the set of estimated cross-sectional coefficients, whose specification and estimation we will consider next.[6]

III. Modeling the Dynamics of the Implied Volatility Surface

A. The Model

In this subsection we model the time variation of the IVS as captured by the dynamics of the OLS coefficients entering the cross-sectional model analyzed previously. More specifically, we fit VAR models to the time series of OLS

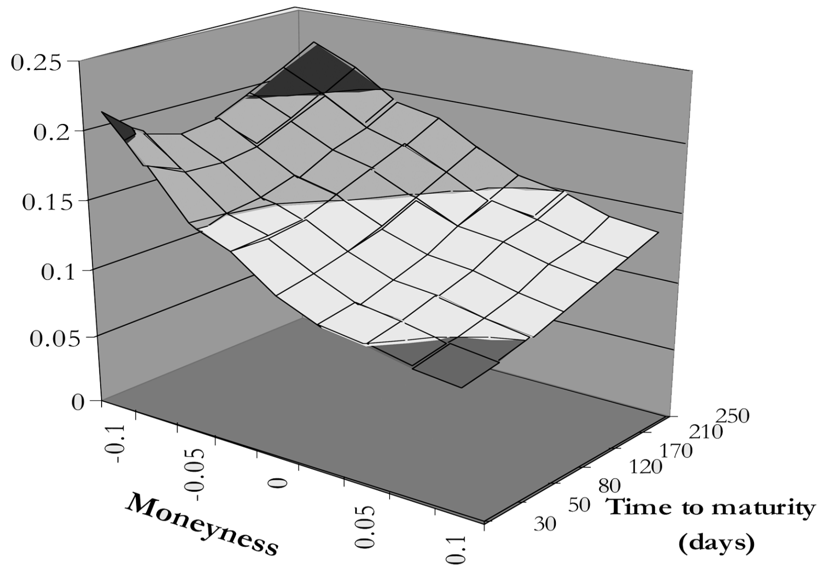
6. Although the mapping between the persistence of the cross-sectional coefficients and the persistence of (log) implied volatilities is a complicated one, for ATM contracts the mean reversion speed is well approximated by the autocorrelation function of β_0 and appears to be consistent with an AR(1) model with an autoregressive coefficient of 0.9. This estimate is lower than the volatility mean reversion parameter reported, e.g., by Heston and Nandi (2000). However, we note that Heston and Nandi study the volatility of the underlying (in levels), not implied, volatilities. Christensen and Prabhala (1998) study log-implied volatilities and find an autoregressive coefficient of 0.7.

TABLE 2 Summary Statistics for the Parameter Estimates of the Cross-Sectional Model Equation (1)

Coefficient	Mean	Standard Deviation	Minimum	Maximum	Skew	Kurtosis	LB(1)	LB(10)	LB(1) Squares	LB(10) Squares
					A. OLS Estimates					
$\hat{\beta}_0$	−2.186	.164	−2.658	−1.618	.368	2.582	927.0**	6,550**	922.7**	6,516**
$\hat{\beta}_1$	−1.265	.690	−8.854	1.518	−.985	15.75	116.4**	855.3**	23.28**	202.0**
$\hat{\beta}_2$	1.689	2.107	−8.601	14.33	1.090	6.052	56.29**	288.9**	7.23**	116.5**
$\hat{\beta}_3$	.292	.246	−.558	2.993	1.471	16.65	341.4**	2,026**	18.79**	174.7**
$\hat{\beta}_4$	−1.140	2.466	−22.30	39.09	2.840	70.34	14.74**	95.08**	.028	1.353
$\bar{R}^2$	.810	.133	.011	.990	−1.373	5.518	28.81**	112.3**	33.70**	128.0**
RMSE	.010	.005	.001	.044	1.701	7.100	55.41**	114.6**	54.62**	77.22**
					B. GLS Estimates					
$\hat{\beta}_0$	−2.144	.165	−3.040	−1.589	.074	3.117	756.8**	5,700**	727.6**	5,488**
$\hat{\beta}_1$	−1.597	.855	−3.394	48.21	−3.394	48.21	65.31**	584.1**	1.783	20.07
$\hat{\beta}_2$	.147	2.648	−29.67	19.49	−1.146	24.36	32.53**	96.79**	20.89**	53.49**
$\hat{\beta}_3$	.224	.246	−.995	4.737	5.750	103.9	131.0**	845.3**	.272	5.508
$\hat{\beta}_4$	−.379	2.816	−18.70	65.42	10.51	268.1	.015	30.65**	.004	6.632
$\bar{R}^2$	.717	.284	.001	.989	−4.497	42.92	6.956**	37.50**	5.376	35.43**
RMSE	.012	.007	.002	.056	2.100	9.595	50.86**	115.6**	45.36**	75.99**

NOTE.—For each trading day, estimation is constrained by the availability of a sufficient number of observations. Panel A concerns OLS estimates, and panel B reports GLS estimates that adjust for the effects of measurement error involving option prices and the underlying index. The data cover the period January 3, 1992–June 28, 1996, for a total number of daily estimated vector coefficients equal to 1,136. $\bar{R}^2$ denotes the adjusted R^2, and LB(j) denotes the Ljung-Box statistics testing for the absence of autocorrelation up to lag j. RMSE denotes the RMSE of (log) implied volatilities.

** Significantly different from zero at the 1% level.



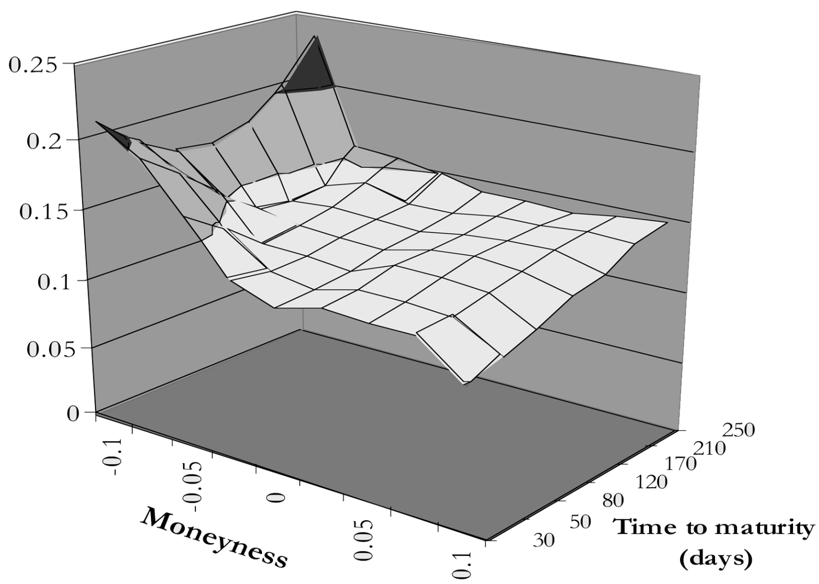


Fig. 1.—Fitted (top) and actual (bottom) S&P 500 IVS: average over January 3, 1992–June 28, 1996.

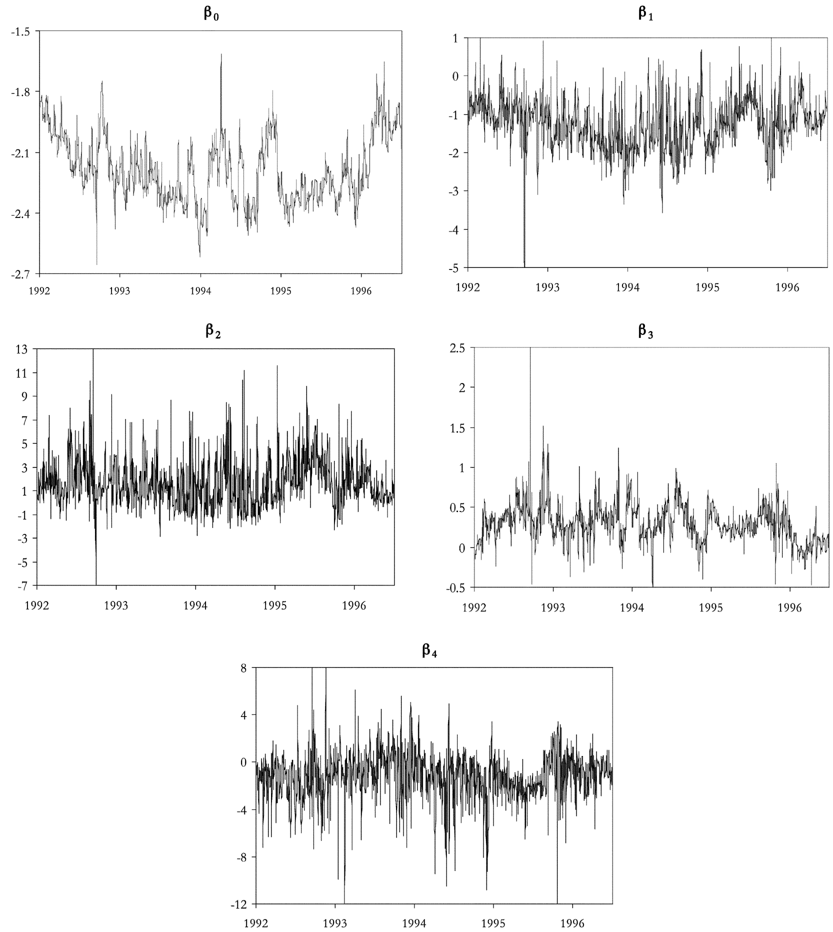


FIG. 2.—Time variation in the OLS estimates for the cross-sectional model, eq. (1): January 3, 1992–June 28, 1996.

estimates $\{\hat{\boldsymbol{\beta}}_t\}$ implied by equation (1), where $\hat{\boldsymbol{\beta}}_t$ denotes day t's coefficient estimates. Our approach is a reduced-form approach to modeling the time variation in the IVS that results from more structural models such as the investors' learning models of option prices. In particular, if the state variables that control the dynamics underlying the fundamentals in these models are persistent and follow a regime-switching model (such as in David and Veronesi [2002] or Garcia, Luger, and Renault [2003]), a VAR model appears to be a reasonable reduced-form approach to model the predictability in the IVS.

We consider the following multivariate model for the vector of estimated coefficients $\hat{\boldsymbol{\beta}}_t$:

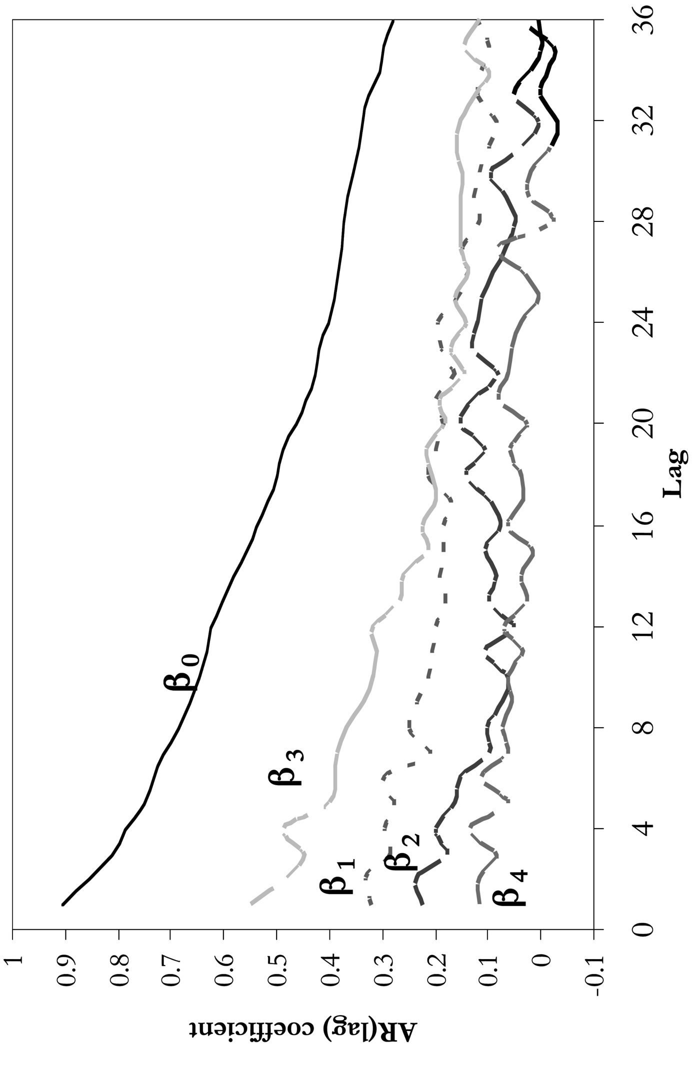
$$\hat{\boldsymbol{\beta}}_t = \boldsymbol{\mu} + \sum_{j=1}^{p} \boldsymbol{\Phi}_j \hat{\boldsymbol{\beta}}_{t-j} + \mathbf{u}_t, \tag{2}$$

where $\mathbf{u}_t \sim$ i.i.d. $N(\mathbf{0}, \boldsymbol{\Omega})$.

For later reference, let $\boldsymbol{\pi}$ denote the vector containing all parameters (including the elements of $\boldsymbol{\Omega}$) entering (2). Equations (1) and (2) describe our two-stage, dynamic IVS model. We select p using the Bayesian information criterion (BIC), starting with a maximum value of $p = 12$. This is our main model (which we label model 1).[7] For comparison purposes, we consider Dumas et al.'s (1998) ad hoc straw man, which has proved to be hard to beat in out-of-sample horse races. Christoffersen and Jacobs (2004) have recently employed this benchmark to show that once the in-sample and out-of-sample loss functions used in estimation and prediction are correctly "aligned," this practitioners' Black-Scholes model is hard to outperform even using state-of-the-art structural models. This model (henceforth model 2) is a special case of equation (2) with $\mu = 0$, $p = 1$, $\Phi_1 = I_5$, a 5×5 identity matrix, $\Phi_j = \mathbf{0}$ for $j = 2, \ldots, p$, and $\boldsymbol{\Omega}$ a diagonal matrix. It is a random walk model in which $\hat{\boldsymbol{\beta}}_t = \hat{\boldsymbol{\beta}}_{t-1}$ plus an independently and identically distributed (i.i.d.) random noise vector; that is, the best forecast of tomorrow's IVS parameters is today's set of (estimated) coefficients.

We estimate model 1 by applying OLS equation by equation. For comparison purposes, we also estimate on our options data a third structural model, Heston and Nandi's (2000) NGARCH(1, 1). Heston and Nandi report the superior performance (in- and out-of-sample) of this model over Dumas et al. ad hoc straw man when estimated on weekly S&P 500 options data for the period 1992–94. In contrast to the dynamic IVS models considered here, the NGARCH(1, 1) model does not allow for time-varying coefficients (although it implies time-varying risk-neutral densities). Thus it seems sensible to require that model 1 be able to perform at least as well as Heston and Nandi's NGARCH. We estimate Heston and Nandi's model by minimizing the sum of the squared deviations of the Black-Scholes implied volatilities from the Black-Scholes implied volatilities derived by "inverting" the

7. Equation (2) allows for a variety of dynamic specifications of the IVS (as described by the cross-sectional coefficient estimates $\hat{\boldsymbol{\beta}}_t$), depending on the choice of p and on the restrictions imposed on its coefficients. In an earlier version of this paper, we considered two further model specifications: one in which the lag order was selected by a sequential likelihood ratio testing algorithm and one in which exogenous information in the form of lagged returns on the S&P 500 index entered the VAR model. Since the out-of-sample performance of these models turned out to be inferior to model 1, we omit related results (see Gonçalves and Guidolin [2003] for details).


β0
β3
β1
β2
β4
Lag
AR(lag) coefficient
0
4
8
12
16
20
24
28
32
36
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1

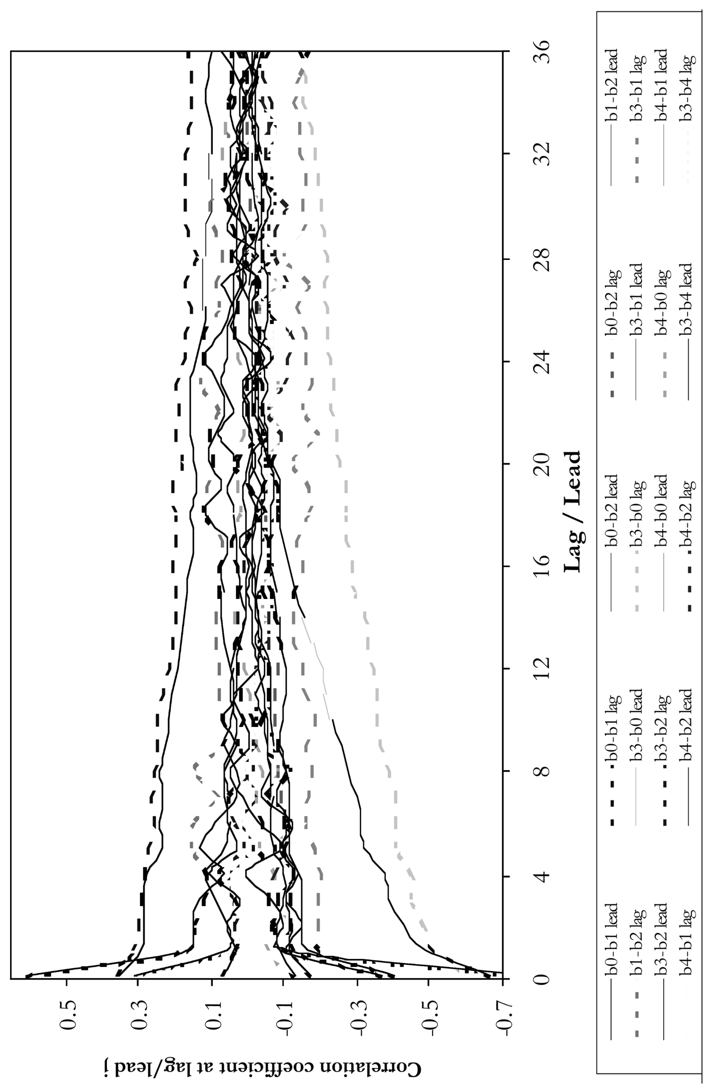


FIG. 3.—Autocorrelations (top) and cross-correlations (bottom) of the OLS estimates forthe cross-sectional model, eq. (1): January 3, 1992–June 28, 1996.

TABLE 3 Estimation Results for VAR Models of Cross-Sectional OLS Estimates

Model	Log Likelihood	BIC	RMSE $\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Model 1	−583.714	.710	.064	.600	1.98	.183	2.40
Model 2	−2,203.256	2.002	.161	.692	2.12	.245	2.48

NOTE.—Model 1 corresponds to eq. (2) in the text, with p selected by the BIC criterion (starting with a maximum value of $p = 12$). Model 2 is the Dumas et al. (1998) ad hoc straw man. All results pertain to the period January 3, 1992–June 28, 1996, for a total of 1,136 daily observations.

NGARCH(1, 1) option prices.[8] This is in contrast to Heston and Nandi, who apply a nonlinear least squares (NLS) method to option prices directly. By estimating Heston and Nandi's model in the implied volatility space, we preserve the consistency with the dynamic IVS models.[9]

B. Estimation Results

Table 3 reports estimation results for models 1 and 2, fitted to the parameter estimates from the cross-sectional model described by equation (1). Model 1 outperforms the more parsimonious model 2 in-sample, as signaled by its high value for the log likelihood function and the smallest RMSE values for the first-step parameter estimates $\hat{\boldsymbol{\beta}}_t$. We will evaluate the two models out of sample to account for the possibility that the superior performance of model 1 is due to overfitting the data.

In order to obtain an idea of the predictions implied by our two-stage IVS model, figure 4 plots the sequence of IVS snapshots over the period January 3, 1992–June 28, 1996, implied by model 1's estimates. In particular, in the top row we plot fitted implied volatilities against time and moneyness, given two distinct maturities (DTE = 30 and DTE = 120), whereas in the bottom row we plot fitted implied volatilities against time and maturity, given two distinct moneyness levels ($m = 0$ and $m = 0.05$, i.e., ATM and ITM puts [and ATM and OTM calls]). Figure 4 shows that model 1 is capable of generating considerable heterogeneity in the IVS, consistent with well-known stylized facts: skews for short-term contracts; relatively higher implied vol-

8. We obtained the following estimates:

$$r_t = r^f - \tfrac{1}{2}\sqrt{h_t} + \sqrt{h_t}z_t^*,$$

with

$$h_t = (0.83 \times 10^{-6}) + (0.67 \times 10^{-6})[z_{t-1}^* + (\tfrac{1}{2} + 316.5 + 2.45)\sqrt{h_{t-1}}]^2 + 0.91h_{t-1},$$

where we use notation similar to that used by Heston and Nandi (2000). The implied nonlinear GARCH process has high persistence ($\beta + \alpha\xi^2 = 0.98$), as typically found in the literature (Heston and Nandi found persistence levels of roughly 0.9–0.95 on their S&P 500 index options weekly data). Also, the estimate of the risk premium is standard (Heston and Nandi's estimates are between 0.5 and 2). The NGARCH(1, 1) model reaches an average implied volatility RMSE of 2.01%, which is quite impressive considering that the model specifies only five parameters.

9. For an example of NLS estimation based on a distance metric based on Black-Scholes implied volatilities, see Jackwerth (2000).

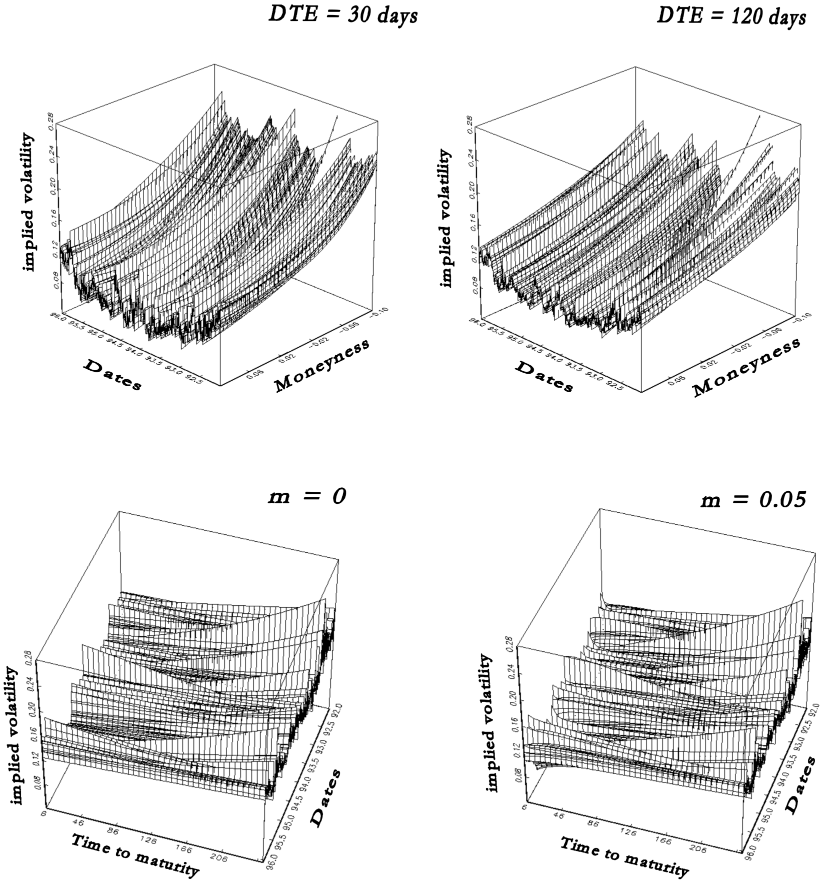


FIG. 4.—Model 1: fitted S&P 500 IVS

atilities in 1992, early 1994, and in the spring of 1996; less accentuated skews, which become asymmetric smiles when higher implied volatilities are observed; and so forth. For medium-term contracts, model 1 implies instead a flatter and practically linear IVS; skews dominate.

The bottom row of plots in figure 4 shows that some heterogeneity affects also the fitted IVS in the term structure dimension. Although positively sloping shapes dominate, flat and even downward-sloping schedules occasionally appear. For instance, between the end of 1992 and early 1993, the fitted term structure is steeply upward sloping, implying volatilities on the order of almost 30% for ATM, long-term contracts (vs. 10% for short-term ones); on the opposite, early 1995 is characterized by flat term structures. For ITM puts (OTM calls), we find flatter schedules on average, although substantial heterogeneity remains. Interestingly, in this case many schedules are actually nonmonotone; that is, they are at first decreasing (for very short maturities, less than one month) and then slowly increasing in time to expiration. We interpret figure 4 as evidence of the possibility of accurately modeling not only the cross-sectional structure of the S&P 500 IVS but also its dynamics. The conceptually simple VAR model 1 provides a very good fit and produces IVSs that are plausible both in their static structure and in their evolution.

IV. Statistical Measures of Predictability

Our approach to modeling the IVS dynamics proves successful in-sample, as previous results show. Nevertheless, a good model of the IVS should not only fit well in-sample but also provide good out-of-sample predictions. The main goal of this section is thus to analyze the out-of-sample forecasting performance of models 1 and 2 at forecasting one-step-ahead, daily implied volatilities (and option prices). For comparison purposes, we include Heston and Nandi's (2000) NGARCH(1, 1) model, as well as a random walk model for daily implied volatilities (henceforth called the "random walk model"). According to this random walk model, today's implied volatility for a given option contract is the best forecast of tomorrow's implied volatility for that same contract. Harvey and Whaley (1992, 53) comment that "while the random walk model might appear naive, discussions with practictioners reveal that this model is widely used in trading index options."

We estimate each of the models using data for the periods January 1, 1992–December 31, 1992; January 1, 1992–December 31, 1993; and so on, up to January 1, 1992–December 31, 1995. This yields four distinct (and expanding) estimation windows. For each day in a given estimation window, we estimate the cross-sectional IVS parameters $\boldsymbol{\beta}_t$ by OLS. We obtain a time series $\{\hat{\boldsymbol{\beta}}_t\}$, which we then use as raw data to obtain estimates of $\boldsymbol{\pi}$, the parameters of the multivariate models described by (2). We allow the model's specification (e.g., the number of lags p) to change in each estimation window. For the NGARCH(1, 1) benchmark, we follow Heston and Nandi's (2000) approach and estimate its parameters (which we also denote by $\boldsymbol{\pi}$ to simplify notation) by NLS, except that our objective function is defined in the IVS. Let $\hat{\boldsymbol{\pi}}$ denote the parameter estimates for each of these models and for a given estimation window. We then hold $\hat{\boldsymbol{\pi}}$ constant for the following six months—that is, January 1, 1993–June 30, 1993; January 1, 1994–June 30, 1994; and so forth up to January 1, 1996–June 28, 1996—and produce daily one-step-ahead forecasts of the estimated coefficients $\hat{\boldsymbol{\beta}}$. Because the IVS on day $t + 1$ depends on $\hat{\boldsymbol{\beta}}_{t+1}$, forecasting $\hat{\boldsymbol{\beta}}_{t+1}$ allows us to forecast implied volatilities (and option prices) for each of these four six-month prediction windows, given moneyness levels and time to maturity. Importantly, nonoverlapping estimation and prediction windows guarantee that only past information on the dynamic properties of the S&P 500 IVS are used for prediction purposes.

To assess the out-of-sample performance of the fitted models for the second half of each of the four years under consideration, for each day in a given prediction window we compute the following six measures for each model:

i. The root mean squared prediction error in implied volatilities (RMSE-V) is the square root of the average squared deviations of Black-Scholes implied volatilities (obtained using actual option prices) from the model's forecast implied volatilities, averaged over the number of options traded.
ii. The mean absolute prediction error in implied volatilities (MAE-V) is the average of the absolute differences between the Black-Scholes implied volatility and the model's forecast implied volatility across traded options.
iii. The mean correct prediction of the direction of change in implied volatility (MCP-V) is the average frequency (percentage of observations) for which the change in implied volatility predicted by the model has the same sign as the realized change in implied volatility.[10]
iv. The root mean squared prediction error in option prices (RMSE-P) is computed as in measure i but with reference to option prices.
v. The mean absolute prediction error in option prices (MAE-P) is computed as in measure ii but with reference to option prices.
vi. The mean correct prediction of the direction of change of option prices (MCP-P) is computed as in measure iii but with reference to option prices.

In computing measures iv–vi above, we compare actual option prices with the model's forecast of option prices. We use the Black-Scholes formula to compute the model's forecast of option price, using the corresponding implied volatility forecast as an input (conditional on the current values of the remaining inputs such as index value, interest rate, and the contract's features). Our use of the Black-Scholes model is obviously inconsistent with the volatility being a function of moneyness and/or time to maturity. Nevertheless, such a pricing scheme is often used by market makers (cf. Heston and Nandi 2000). It is our goal here to see whether a theoretically inconsistent but otherwise flexible approach can deliver statistically and economically significant forecasts. We follow Harvey and Whaley (1992) and view our IVS models as a "black box," which is first used to obtain implied volatilities from option prices for forecasting purposes and then transforms implied volatilities back into prices.[11]

Panel A of table 4 contains the average values of the out-of-sample daily

10. When computing this measure, we consider only contracts that are traded for two consecutive days.

11. The forecasting exercises underlying our computation of the performance measures iv–vi are subject to Christoffersen and Jacobs' (2004) critique that the loss function used in estimation (based on implied volatility matching) differs from the out-of-sample loss function (based on Black-Scholes option prices). Since the Black-Scholes formula is nonlinear in implied volatility, severe biases may be introduced. On the basis of the results of Christoffersen and Jacobs, we expect that the use of the "correct loss" function in estimation will reduce the values of the out-of-sample statistics in table 4 for our approach.

TABLE 4 Out-of-Sample Average Prediction Errors and Forecast Accuracy Tests

	RMSE-V	MAE-V	RMSE-P	MAE-P	MCP-V	MCP-P
			A. Prediction Error Measures			
OLS estimates:						
Model 1	1.429	.971	1.00	.64	62.23	51.61
Model 2	2.305	1.947	1.75	1.33	55.78	46.02
GLS estimates:						
Model 1	1.516	1.048	.93	.65	61.07	49.60
Model 2	2.386	2.051	1.68	1.35	55.19	45.51
Benchmarks:						
NGARCH(1, 1)	2.074	1.721	1.71	1.36	54.51	49.68
Random walk	1.490	1.041	1.64	1.27	NA	NA
			B. Forecast Accuracy Tests (against Model 1)			
OLS estimates:						
Model 2	−20.212***	−14.205***	−11.591***	−14.455***	6.594***	6.400***
NGARCH(1, 1)	−6.770***	−10.286***	−8.455***	−16.990***	8.759***	2.652***
Random walk	−1.947*	−3.411***	−7.620***	−13.492***	NA	NA
GLS estimates:						
Model 2	−11.265***	−14.363***	−12.474***	−14.745***	6.653***	5.420**
NGARCH(1, 1)	−6.063***	−9.103***	−9.026***	−16.016***	7.825***	−.099
Random walk	.288	.277	−8.037***	−12.813***	NA	NA

NOTE.—Model 1 corresponds to eq. (2) in the text, with p selected by the BIC criterion (starting with a maximum value of $p = 12$). Model 2 is the Dumas et al. (1998) ad hoc straw man. NGARCH(1, 1) is Heston and Nandi's (2000) model, estimated in the IVS. The random walk model sets tomorrow's implied volatility forecast equal to today's value. Each model is estimated using data in four expanding estimation windows (January 1, 1992–December 31, 1992, up to January 1, 1992–December 31, 1995), and then used to forecast implied volatilities and option prices in the second half of each year 1992–96. RMSE-V (RMSE-P) is the root mean squared error in implied volatilities (option prices) averaged across all days in the four prediction windows. MAE-V (MAE-P) is the mean absolute error between Black-Scholes implied volatilities (observed option prices) and forecast implied volatilities (forecast option prices using Black-Scholes, given forecast-implied volatilities) across all days in the out-of-sample period. MCP-V (MCP-P) is the mean percentage of correct predictions of changes in implied volatilities (option prices) across all days. The forecast accuracy tests are based on Diebold and Mariano (1995).

* Statistically significant at the 10% level.

** Statistically significant at the 5% level

*** Statistically significant at the 1% level.

performance measures i–vi aggregated across all four out-of-sample periods.[12] The aggregated out-of-sample RMSE in annualized implied volatilities is 1.43%, 2.30%, 2.07%, and 1.49% for models 1 and 2, the NGARCH(1, 1) model, and the random walk model, respectively. The values for the out-of-sample measures related to forecasting option prices are \$1.00, \$1.75, \$1.71, and \$1.64, respectively. The best-performing model according to these measures is model 1, the VAR model for $\hat{\boldsymbol{\beta}}_t$. Similar results are obtained in terms of average percentage of correct predictions for the sign of the change of volatilities between two consecutive trading days: the best performance is provided by model 1 (62.2%), followed by model 2 (55.8%). Modeling the dynamics of the IVS offers real advantages over a simpler, static Dumas et al. type specification (model 2) in which the structure of the IVS is predicted not to change from one day to the next. Model 1 also compares favorably with the two benchmarks considered, outperforming both the NGARCH(1, 1) model and the practitioners' random walk model for implied volatilities. Similarly to Heston and Nandi (2000), we find that the NGARCH(1, 1) model outperforms model 2.[13]

To formally assess the statistical significance of the difference in out-of-sample performance of model 1 compared to each of the remaining models, we employ the equal predictive ability test proposed by Diebold and Mariano (1995). We consider three types of performance indicators: the difference in squared forecast errors (corresponding to measures i and iv), the difference in absolute forecast errors (corresponding to measures ii and v), and the difference between two indicator functions, where each indicator function takes the value one if the realized change in the variable being predicted (e.g., the implied volatility) has the same sign as the predicted change (i.e., the forecast error), and zero otherwise. This last performance indicator is consistent with the out-of-sample measures given in cases iii and vi. To compute the Diebold and Mariano test, we use the Newey-West (1987) heteroskedasticity and autocorrelation consistent variance estimator. Panel B of table 4 reports the values of the statistic and associated significance levels. With very few exceptions, we reject the null hypothesis of equal forecast accuracy of model 1 compared to the benchmark models. We conclude that the out-of-sample superior performance of model 1 is statistically significant. Moreover, in the rare occasions in which model 1 underperforms the benchmarks, not

12. Note that it is not possible to calculate the mean percentage of correct prediction of the direction of change of implied volatility for the random walk model since this model implies zero predicted changes in implied volatility by construction.

13. In unreported results, we also studied out-of-sample performance for each of the four prediction windows. The overall picture remains favorable to our approach, although years of higher volatility and turbulent markets (such as 1994) deteriorate the performance of our approach. We also investigated the forecasting accuracy in multistep-ahead forecasting. We considered horizons of two, three, and five trading days. The ranking across models remains identical to the one from table 4: model 1 outperforms model 2 and the NGARCH(1, 1) benchmarks at all horizons. However, although model 1 is superior, its accuracy declines faster than that of model 2 and the NGARCH as the prediction horizon is increased.

TABLE 5 Out-of-Sample Average Prediction Errors by Moneyness and Maturity

	Short-Term		Medium-Term		Long-Term	
	%RMSE-V	%RMSE-P	%RMSE-V	%RMSE-P	%RMSE-V	%RMSE-P
DITM:						
Model 1	26.66	15.31	10.73	9.98	11.41	15.55
Model 2	28.37	30.85	14.70	20.78	12.76	24.46
AR(1)	26.74	49.35	19.61	32.63	22.08	18.03
ITM:						
Model 1	12.64	12.24	6.63	6.61	7.19	9.19
Model 2	16.85	27.40	10.76	14.53	12.79	14.96
AR(1)	15.62	37.09	11.52	14.23	10.70	7.62
ATM:						
Model 1	6.08	6.47	4.84	4.93	5.83	5.71
Model 2	13.19	14.30	10.28	10.42	9.37	9.23
AR(1)	6.63	7.08	6.05	6.09	6.32	6.27
OTM:						
Model 1	5.39	4.26	4.26	4.05	6.46	5.17
Model 2	11.71	5.98	9.39	6.62	10.75	9.20
AR(1)	16.53	5.47	9.55	7.09	5.44	7.62
DOTM:						
Model 1	4.23	2.97	3.98	2.95	7.24	4.40
Model 2	8.16	3.12	8.54	3.91	11.70	4.86
AR(1)	12.5	3.43	13.89	5.22	9.15	8.94

NOTE.—Model 1 corresponds to eq. (2) in the text, with p selected by the BIC criterion (starting with a maximum value of $p = 12$) and without any exogenous regressors. Model 2 is the Dumas et al. (1998) ad hoc straw man. The third model is an AR(1) model applied to each (log) implied volatility time series. Each model is estimated on four expanding windows of observations and then used to forecast implied volatilities on four successive windows of six months each. %RMSE-V and %RMSE-P are RMSEs for volatility and option prices, expressed as a percentage of the mean implied volatility and option price within the class, respectively. Each time series is formed by sampling contracts that in each available day come closer to class definitions based on moneyness and maturity.

only is the difference rather small in absolute terms, but we cannot reject the hypothesis of equal predictive accuracy.

The superior out-of-sample performance of model 1 relative to model 2, the static ad hoc model heavily used by practitioners, confirms that time variation in the IVS is statistically important. Economic models of the IVS such as those that allow for investors' learning to affect equilibrium option prices can explain these findings. If on a learning path beliefs are persistent because the updating occurs in a gradual fashion, the stochastic discount factor should inherit these properties and imply predictability of the IVS. This implies that model 2, which ignores such predictability—that is, a random walk for the first-stage coefficients—has a hard time capturing the dynamics of the IVS. Instead, model 1 represents a reduced-form framework able to capture the dynamic properties of the IVS. As often documented in forecasting applications, such a reduced-form approach works very well, outperforming the more complex structural model of Heston and Nandi (2000).

In order to further analyze the nature of the forecasting ability of model 1, table 5 reports out-of-sample average prediction errors by different option moneyness and maturity categories. Specifically, for each category we report the average out-of-sample RMSE for implied volatilities (and option prices), expressed as a percentage of the mean implied volatility (and mean option price) in that category. Scaling by mean volatility and price is important to gain comparative insight into the sources of model 1's outperformance. For comparison purposes, we also include model 2, the restricted (static) version of the more flexible dynamic model 1. In addition, we consider a simple AR(1) model for (log) implied volatilities, as in Christensen and Prabhala (1998). Contrary to model 1, this model does not exploit the panel structure of options data since it applies to a single time series of (log) implied volatilities. In particular, for a given options class, we create a time series of (log) implied volatilities by selecting each day the contract that is closest to the midpoint in this category.[14] Since this simple AR(1) model does not utilize any cross-sectional restrictions on implied volatilities, we expect it to perform worse than model 1.

Our findings are as follows. We start with model 1. For any given moneyness level, medium-term contracts are associated with the smallest prediction errors, both in implied volatilities and in option prices. The ranking between short-term and long-term contracts depends on moneyness. For ITM and ATM options, long-term contracts have smaller prediction errors than short-term contracts (in both the volatility and price metrics). For OTM options the opposite is true. For a given maturity level, RMSEs (in volatilities and option prices) are generally decreasing when moving from DITM to DOTM; that is, it is easier to predict OTM than ITM implied volatilities and option prices. The only exception to this pattern occurs when forecasting implied volatilities for long-term contracts, for which a U-shaped pattern of RMSE-V emerges.

In sum, the forecasting strength of model 1 seems to originate mainly from the short- and medium-term OTM and ATM segments of the market.

For the AR(1) model, RMSEs tend to decrease with maturity, given moneyness. One exception is the DOTM class, for which short-term options have the lowest RMSE-P. For any maturity level, the AR(1) model achieves, in general, lower RMSE-V for ATM implied volatilities than ITM or OTM contracts. For short-term and medium-term options, the RMSE-P decreases monotonically when moving from DITM to DOTM.

Table 5 shows that model 1 generally beats the AR(1) model across all moneyness and time to expiration classes.[15] Thus the gain in forecasting from

14. For a given options class, on each day for which there are options available in that class, we select the contract that solves the following problem:

$$\min_{m_i,\tau_i}\left[\frac{(m_i - m_c)^2}{\sigma_m^2} + \frac{(\tau_i - \tau_c)^2}{\sigma_\tau^2}\right],$$

where m_c and τ_c are the midpoints of the moneyness and maturity intervals defining the class, and σ_m^2 and σ_τ^2 are the variances of moneyness and time to expiration for all contracts in the class traded that day.

15. The AR(1) model outperforms model 1 only in two cases: for ITM, long-term options (when it achieves a smaller %RMSE-P) and for OTM, long-term options (with a smaller %RMSE-V).

our two-stage approach seems to come from the cross-sectional restrictions. The greatest improvements in RMSE-V occur for OTM short- and medium-term contracts; instead, the greatest gains in RMSE-P occur for ITM short- and medium-term contracts. The smallest gains are obtained for ATM contracts. This confirms that the additional information contained in the segments of the IVS far from at-the-money may be crucial in improving the forecasting performance of IVS models.

Model 1 also outperforms model 2 for all categories. The largest reductions in average prediction errors are obtained for ATM and OTM short- and medium-term options, when forecasting implied volatilities, whereas ATM and ITM short- and medium-term options show the largest reductions in RMSE-P. DITM options are in general associated with smaller reductions in implied volatilities prediction errors, suggesting that for this class of options the dynamics in the coefficients capturing the IVS shape is stable enough to allow accurate forecasting from model 2.

For OTM short- and medium-term options, model 2 yields lower average prediction errors than the AR(1) model, which suggests that for these classes it is more important to model the cross-section dimension of the options data than the time-series dimension. Instead, for ATM options, the simple AR(1) model outperforms model 2, suggesting that it is important to model the dynamics of implied volatilities for this class of options.

V. Economic Analysis

The results of Section IV suggest that implied volatilities (and corresponding option prices) are highly predictable in a statistical sense. The good out-of-sample statistical performance of our model and the fact that our approach can be viewed as a reduced-form approach that captures the dynamics in the IVS that could be generated by equilibrium-based economic models suggest some robustness of our results to data mining. However, we cannot exclude entirely the possibility that our results are subject to mining biases. Therefore, as an additional test, we now examine the economic consequences and significance of this predictability. In particular, we ask the following question: Would a hypothetical market trader be able to devise any profitable trading strategies based on the implied volatility forecasts produced by our two-stage dynamic IVS models? We follow Day and Lewis (1992), Harvey and Whaley (1992), and Noh et al. (1994) and evaluate the out-of-sample forecasting performance of a number of competing models by testing whether certain trading rules may generate abnormal profits, that is, profits that are not accounted for by the risk of the positions required by the strategies.[16]

16. These experiments might be also constructed as tests of the informational efficiency of the S&P 500 index options market. An efficient market ought to be able to produce option prices consistent with the implied volatility forecasts from our two-step estimation procedure. If abnormal profits can be made, the efficient market hypothesis is rejected. Alternatively, the most likely explanation is to be found in microstructural features that make the underlying index and option prices adjust to the flow of news at different speeds.

A. Trading Strategies and Rate of Return Calculations

The trading strategies we consider are based on out-of-sample forecasts of volatility. More specifically, if on a given day implied volatility is predicted to increase (decrease) the following day, the option is purchased (sold). Each day we invest \$1,000 net in a delta-hedged portfolio of S&P 500 index options, which is held for one trading day.[17] The trading exercise is repeated every day in the out-of-sample period, and a rate of return is calculated.

Implied volatility forecasts are obtained as in Section IV: on day t we use the time series of estimated coefficients $\hat{\boldsymbol{\beta}}$ describing the IVS, up to and including day t, to predict day $t + 1$'s coefficients $\hat{\boldsymbol{\beta}}_{t+1}$ by means of the VAR-type models estimated from the appropriate estimation window. The forecast of $\hat{\boldsymbol{\beta}}_{t+1}$ is then used to predict day $t + 1$'s implied volatility associated with a given option. Since the index price and interest rate at $t + 1$ are not known as of time t, we assume that today's prices of the primitive assets are tomorrow's best forecasts. To delta-hedge our options position, per each unit of call (put) options bought, we sell (buy) an amount of the underlying index equal to the Black-Scholes delta ratio (Δ), calculated using the implied volatility forecast. Similarly, if we sell one call (put) option, we buy (sell) an amount of the underlying index equal to the corresponding Black-Scholes hedge ratio.[18]

To compute the rate of return, we assume that funds may be freely invested at the riskless interest rate. Suppose that one particular trading rule has indicated that a certain subset of contracts Q should be traded at time t. Let C_{it} denote the price of a call contract i at time t and P_{it} the price of a put contract i at time t. The delta ratios corresponding to call and put options are denoted Δ^C_{it} and Δ^P_{it}, respectively. If no options are traded (i.e. Q is empty), we force the trader to invest her \$1,000 in the riskless asset for one trading period. We distinguish between two cases: a first case in which the overall time t net cost of the delta-hedged portfolio is positive and a second case in which the cost is negative.

Consider first the case in which the portfolio requires an injection of funds. Let V_t denote the price of a unit portfolio in which all contracts are sold or purchased in one unit:

$$V_t = \sum_{i \in Q_+^{\text{call}}} (C_{it} - S_t \Delta^C_{it}) + \sum_{i \in Q_+^{\text{put}}} (P_{it} + S_t \Delta^P_{it}) - \sum_{i \in Q_-^{\text{call}}} (C_{it} - S_t \Delta^C_{it}) - \sum_{i \in Q_-^{\text{put}}} (P_{it} + S_t \Delta^P_{it}), \tag{3}$$

where Q_+^{call} (Q_-^{call}) is the subset of Q for which a buying (selling) signal on

17. Delta-hedging is intended to render the portfolio's value insensitive to market movements so that our computed profits truly reflect profits in "trading in volatility."

18. In practice, hedging is accomplished by trading in S&P 500 futures with appropriate maturities. The resulting hedging is imperfect since the underlying consists of the spot index, and index and futures fail to be perfectly correlated (basis risk). For the sake of simplicity we ignore the complications arising from hedging with futures.

calls was obtained; similar definitions apply to puts. Then \$1,000 is invested in a portfolio in which all options in Q (and their associated delta-hedging positions in the S&P 500 index) are traded in the quantity $X_t = 1{,}000/V_t$, with a total cost of \$1,000. Hence the resulting portfolio is value-weighted. The net gain between t and $t+1$ can be determined as

$$\begin{aligned} G_{t+1}^{\text{out}} = {} & X_t\left[\sum_{i\in Q_+^{\text{call}}}(C_{i,t+1}-C_{it})+\sum_{i\in Q_+^{\text{put}}}(P_{i,t+1}-P_{it})\right] \\ & + X_t\left[\sum_{i\in Q_-^{\text{call}}}(C_{it}-C_{i,t+1})+\sum_{i\in Q_+^{\text{put}}}(P_{it}-P_{i,t+1})\right] \\ & - X_t(S_{t+1}-S_t)\Big(\sum_{i\in Q_+^{\text{call}}}\Delta_{it}^C+\sum_{i\in Q_-^{\text{put}}}\Delta_{it}^P\Big) \\ & + X_t(S_{t+1}-S_t)\Big(\sum_{i\in Q_-^{\text{call}}}\Delta_{it}^C+\sum_{i\in Q_+^{\text{put}}}\Delta_{it}^P\Big). \end{aligned} \tag{4}$$

Next, consider the case in which the portfolio generates cash inflows; for example, most or all of the trading signals are selling signals. Define V_t as in (3), except for the fact that now $V_t < 0$. In this case a portfolio worth \$1,000 is created by trading each contract for which there exists an active signal in the quantity $X_t = 1{,}000/|V_t|$. We assume that the \$1,000 option portfolio generated inflows plus the additional \$1,000 originally available is invested at the riskless interest rate r_t. The resulting net gain between t and $t+1$ can be calculated in a manner similar to (4):

$$G_{t+1}^{\text{in}} = G_{t+1}^{\text{out}} + 2{,}000\exp\left(\frac{r_t}{252}\right).$$

We consider several trading rules. In order to avoid noisy signals, all our trading strategies use a price deviation filter of 5 cents.[19] This implies that trading occurs only when the price difference between the predicted option price (i.e. the Black-Scholes predicted price based on our volatility forecast) and today's observed price is larger than the filter.[20] First, following Harvey and Whaley (1992), we consider a trading rule (henceforth trading rule A) in which trades occur only on the closest ATM, shortest-term contracts (thus $Q \leq 1$). Second, we consider a strategy (trading rule B) for which trading occurs only in two contracts, those for which the expected selling and the expected buying profits, respectively, are maximum. In this case $Q \leq 2$ obtains

19. Later we will increase the value of this filter.
20. Since the theta of a European option (the rate of change of its value as time to maturity decreases) is normally negative, comparing predicted and current implied volatilities contains a small bias, in the sense that, ceteris paribus, the option price implied by predicted volatility will be normally slightly smaller than the current price because of the mere passage of time. Applying some minimal filter to the differences in implied prices adjusts for this bias.

at all times. In a third set of simulations (trading rule C), we consider trading only in one contract, the one giving the highest expected trading profit, so that $Q \leq 1$ again.

B. *Trading Profits before Transaction Costs*

Table 6 presents summary statistics for profits deriving from trading rules A–C. We consider two measures of abnormal returns (profitability): the Sharpe ratio and a risk measure due to Leland (1999). The Sharpe ratio is an appropriate measure of profitability when investors have mean-variance preferences. This is hard to rationalize under nonnormal returns. Instead, Leland's risk measure allows for deviations from normality by taking into account skewness, kurtosis, and other higher-order moments of the returns distribution. It derives from a marginal utility–based version of the single-period capital asset pricing model (CAPM) as follows:

$$A = E\left[\frac{G_{t+1}}{1{,}000}\right] - r_t - B(E[r_{\mathrm{mkt}}] - r_t),$$

where r_{mkt} denotes the return on the market portfolio and B is conceptually similar to a preference-based CAPM beta (under power utility). Crucially, a positive A indicates performance that is abnormal even when the features of higher-order moments (like negative skewness or excess kurtosis) of the empirical distribution of trading profits are taken into account. Appendix A provides further details on the calculations underlying A and its inputs.

Three benchmarks are considered. One is the random walk model for implied volatilities. Since this model predicts tomorrow's implied volatility to be equal to today's value, it does not provide buy or sell signals, and therefore the resulting strategies trivially correspond to buying and holding T-bills every day in the prediction window. In this case, mean profits are negligible and the Sharpe ratio is zero by construction. One might wonder whether it is simply possible to make abnormal profits by randomly trading option contracts. We therefore include a random (delta-hedged) buy and sell option strategy as a benchmark: according to this rule, each option has a 0.5 probability of being traded; if selected, the option is sold with probability 0.5, otherwise it is purchased. The third benchmark we consider is the "S&P 500 buy and hold" rule, by which each day the $1,000 is simply invested in the underlying S&P 500 index, thus obtaining Sharpe ratios and A coefficients that are typical of the CAPM.

Table 6 shows that our two-step approach to modeling and forecasting the S&P 500 IVS is successful at generating profitable strategies. Indeed, model 1 yields statistically significant positive mean profits under all three trading rules. Trading rule A, based on trading the closest ATM, shortest-maturity contract, implies a daily mean profit of 0.083%, with a t-ratio of 4.2, followed by trading rule C (mean profit equal to 1.322%, with a t-ratio equal to 11.03) and by trading rule B (mean profit of 2.18%, with a t-ratio equal to 3.9). As

TABLE 6 Simulated Trading Profits before Transaction Costs

	Mean Moneyness	Mean Time to Maturity	Mean Profit (%)	Daily % Standard Deviation	*t*-Ratio	Sharpe Ratio (%)	*A* Coefficient (%)
Trading rule A:							
Model 1	.0009	38.83	.0830	.0198	4.198	14.800	.0418
Model 2	.0009	38.39	.0477	.0196	2.435	6.886	.0065
NGARCH(1, 1)	.0009	38.90	.0545	.0194	2.805	8.511	.0135
Trading rule B:							
Model 1	−.0170	91.04	2.1809	.5551	3.929	17.394	−2.8445
Model 2	−.0199	103.05	−.1166	.9415	−.1239	−.6357	−14.5058
NGARCH(1, 1)	−.0127	85.43	.7056	.3477	2.029	8.832	−1.2875
Trading rule C:							
Model 1	.0004	133.31	1.322	.1198	11.034	48.599	1.0548
Model 2	−.0119	113.25	1.3553	.1597	8.489	37.400	.9076
NGARCH(1, 1)	.0117	90.31	2.2146	.3751	5.905	26.146	−.0982
Benchmarks:							
S&P 500 buy and hold	NA	NA	.0166	.0287	.578	4.670	0
Random option portfolio	−.0119	51.23	−.1483	.1848	−.803	−4.008	−.7027
T-bill portfolio (random walk)	NA	NA	.0175	.0002	86.638	0	−.0174

NOTE.—Model 1 is a VAR model. Model 2 is the Dumas et al. (1998) ad hoc straw man. NGARCH(1, 1) is Heston and Nandi's (2000) model, estimated in the IVS. Each model is estimated on four expanding windows of observations and then used to forecast implied volatilities on four successive windows of six months each. Given implied volatility forecasts, Black-Scholes option prices are computed. If the observed option price of a contract is below (exceeds) the theoretical price, $1,000 of the options are purchased (sold) and the options position is hedged. According to trading rule A, trading only occurs on the closest ATM shortest-term contracts; in trading rule B, trading occurs only in two contracts, those for which the expected selling and the expected buying profits, respectively, are maximum; in trading rule C, trades concern only one contract, the one giving the highest expected profit.

expected, trading rule A is less successful than the remaining trading rules since it is constrained in terms of moneyness. All trading rules yield Sharpe ratios that easily outperform the 4.7 ensured by the S&P 500 buy and hold strategies; that is, they do reward risk in excess of the market portfolio. This conclusion is robust to the CAPM-based performance evaluation delivered by the coefficient A for trading rules A and C, for which A is positive. For trading rule B, a negative value of A is obtained, despite the large value of the Sharpe ratio (17.4). The empirical distribution of trading profits for this trading rule reveals that it is associated with very high values of excess kurtosis, which is negatively weighted under the A coefficient. Since the Sharpe ratio takes into account only the mean and variance of profits, it fails to include this feature, explaining the large value obtained. The negative value of A suggests that daily rewards in excess of 2% per day are insufficient to compensate for the risk absorbed under trading rule B.

A comparison between model 1 and the remaining models reveals that model 1 yields, in general, higher mean daily profits than model 2 and NGARCH(1, 1). One exception is trading rule C, for which the NGARCH(1, 1) model performs best, yielding a mean profit of 2.21% per day, against mean profits of 1.35% for model 2 and 1.32% for model 1. Nevertheless, the high profits obtained by the NGARCH(1, 1) under trading rule C are abnormally low as signaled by a negative value of A. Instead, models 1 and 2 are associated with large values of Sharpe ratios and positive values of A, suggesting that their performance is truly abnormal.

C. Trading Results after Transaction Costs

The results from subsection *B* suffer from two limitations. First, they ignore the effect of transaction costs. Second, trading rules A–C may be so narrowly defined as to imply that a very limited (typically, $Q = 1$) number of contracts are traded. Therefore, it is possible that a model that poorly predicts volatilities and prices out-of-sample does manage to provide correct buy and sell signals, either for ATM short-term contracts or for the most aberrant misspricings (maximizing expected profits).

Table 7 presents results that take transaction costs into account. We recompute rate of returns for trading rules A–C when the payment of a fixed transaction cost per contract traded (both options and the S&P 500 index) is imposed. We apply two different levels of unit cost, \$0.05 (panel A) and \$0.125 (panel B). Panel A shows that low transaction costs barely change the conclusions reached in table 6. As expected, after–transaction costs profits are lower on average, but the ranking of models is the same as in table 6. Model 1 outperforms model 2 and the NGARCH(1, 1) for trading rules A and B, achieving the highest daily mean percentage profits and Sharpe ratios. For trading rule C, model 1's performance is similar to that of model 2. Although both models yield lower daily mean profits than the NGARCH(1, 1) model, they both guarantee positive A coefficients, with model 1 achieving

TABLE 7 Simulated Trading Profits after Transaction Costs

	Mean Moneyness	Mean Time to Maturity	Mean Profit (%)	Daily % Standard Deviation	t-Ratio	Sharpe Ratio (%)	A Coefficient (%)
	A. Transaction Cost of 5 Cents Round-Trip						
Trading rule A:							
Model 1	.0009	38.83	.0554	.0198	2.799	8.557	.0141
Model 2	.0009	38.39	.0182	.0196	.927	.160	−.0230
NGARCH(1, 1)	.0009	38.90	.0257	.0195	1.319	1.880	−.0154
Trading rule B:							
Model 1	−.0170	91.04	.3940	.5530	.713	3.039	−4.5923
Model 2	−.0199	103.05	−1.7938	.5382	−3.333	−15.021	−6.5190
NGARCH(1, 1)	−.0127	85.43	−1.3928	.3702	−3.762	−17.002	−3.6472
Trading rule C:							
Model 1	.0004	133.31	1.2989	.1197	10.850	47.775	1.0319
Model 2	−.0119	113.25	1.3321	.1596	8.349	36.773	.8849
NGARCH(1, 1)	.0117	90.31	2.1868	.3758	5.820	25.768	−.1345
	B. Transaction Cost of 12.5 Cents Round-Trip						
Trading rule A:							
Model 1	.0009	38.83	.0140	.0198	.705	−.780	−.0273
Model 2	.0009	38.39	−.0261	.0196	−1.331	−9.914	−.0673
NGARCH(1, 1)	.0009	38.90	−.0177	.0195	−.908	−8.053	−.0588
Trading rule B:							
Model 1	−.0170	91.04	−2.3246	.7328	−3.172	−14.264	−11.0556
Model 2	−.0199	103.05	−3.3199	.3497	−9.494	−42.598	−5.3348
NGARCH(1, 1)	−.0127	85.43	−3.6611	.4567	−8.017	−35.953	−7.0729
Trading rule C:							
Model 1	.0004	133.31	1.2638	.1195	46.534	46.534	.9975
Model 2	−.0119	113.25	1.2974	.1594	35.831	35.831	.8509
NGARCH(1, 1)	.0117	90.31	2.1452	.3768	25.200	25.200	−.1894

NOTE.—The table reports trading profits when transaction costs of 5 cents (panel A) and 12.5 cents (panel B) per contract traded, on a round-trip basis, are imposed. Model 1 is a VAR model, and model 2 is the Dumas et al. (1998) ad hoc straw man. NGARCH(1, 1) is Heston and Nandi's (2000) model, estimated in the IVS. See also the note to table 6.

the largest percentage abnormal return. In contrast, the NGARCH(1, 1) implies a negative value of A.

To test the robustness of our results, panel B increases transaction costs to \$0.125 per traded contract (round-trip). In this case, positive and significant mean daily profits result for all models under trading rule C, with the best-performing model being the NGARCH(1, 1) model, followed by model 2 and model 1. As before, the performance of the NGARCH(1, 1) model cannot be considered abnormal as signaled by the (negative) value of the A coefficient, whereas the performance of models 1 and 2 can. Nevertheless, none of the models is able to produce significantly positive profits under the other two trading strategies (trading rules A and B).

One of the strengths of our two-step approach is that it allows us to model and forecast the entire S&P 500 IVS. The trading rules analyzed thus far are designed to pick a small number of option contracts (typically $Q = 1$ or 2) and therefore do not exploit entirely the flexibility provided by our approach. In order to allow for trade in a larger set of option contracts, we introduce a fourth type of trading strategy (trading rule D), which applies filter rules to the price deviation for selecting options to be traded. In particular, we consider filters equal to \$0.125, \$0.25, and \$0.50 and allow trades in all contracts for which the absolute value of the price deviation exceeds the filter. Under these filter arrangements, Q can contain a large number of contracts, not being constrained to be at most one or two contracts, as in trading rules A–C. In addition to the price filters, we apply transaction costs of the same magnitude on each contract traded on a round-trip basis, as in table 7.[21] High transaction costs such as \$0.50 are designed to represent the situation faced by retail customers, who often pay substantial commission fees in addition to the bid-ask spread.

Table 8 reports the results for trading rule D. It shows that the profitability of filtered-based trading rules depends heavily on the magnitude of the filter/transaction cost employed. For a filter/transaction cost of \$0.125, model 1 is the only model that is able to guarantee significant (statistically and economically) positive profits. This is in contrast with the static IVS model (model 2) and the NGARCH(1, 1) model, which predict negative (statistically significant) profits. Results not reported here show that most of model 1's profits come from trading short-term ATM and OTM contracts. Instead, DITM contracts yield losses on average, with profits being statistically significantly negative for medium-term contracts. This is consistent with our previous findings of smaller RMSE-P for OTM as compared to ITM options for model

21. Transaction cost–based filter strategies (i.e., strategies that discount the presence of a cost that is actually to be paid on each traded contract) have two opposing effects. On one hand, they may raise trading profits since they constrain Q to contain only signals that, at least in expectation, imply positive after–transaction cost profits. On the other hand, and because we apply transactions costs of the same magnitude as the filter, they obviously depress after–transaction cost realized profits. Which effect turns out to be stronger is an empirical issue. For instance, Harvey and Whaley (1992, table 5) find that *high* enough transaction costs used as filters induce positive and significant profits (however, their simulation does not apply transaction costs equal to filters).

TABLE 8 Simulated Trading Profits under Trading Rule D after Transaction Costs

	Mean Moneyness	Mean Time to Maturity	Mean Profit (%)	Daily % Standard Deviation	*t*-Ratio	Sharpe Ratio (%)	*A* Coefficient (%)
	A. Filter: Transaction Cost = 12.5 Cents Round-Trip						
Model 1	−.0034	48.33	.3118	.0896	3.479	14.657	.1468
Model 2	−.0056	52.08	−2.5551	.7851	−3.255	−14.626	−12.5705
NGARCH(1, 1)	−.0045	43.69	−.7388	.1392	−5.309	−24.257	−1.0874
	B. Filter: Transaction Cost = 25 Cents Round-Trip						
Model 1	−.0025	55.99	.0621	.1530	.406	1.303	−.3521
Model 2	−.0045	55.90	−2.9678	.7093	−4.184	−18.785	−11.1492
NGARCH(1, 1)	−.0045	45.07	−2.4101	.6435	−3.745	−16.837	−9.1506
	C. Filter: Transaction Cost = 50 Cents Round-Trip						
Model 1	−.0009	74.79	−.5466	.1187	−4.606	−21.213	−.8096
Model 2	−.0021	62.03	−4.807	.6988	−6.879	−30.814	−12.7497
NGARCH(1, 1)	−.0032	48.96	−5.1207	1.1002	−4.655	−22.655	−7.8994

NOTE.—The table reports trading profits from trading rule D. This strategy applies filter rules to price deviations for selecting options to be traded. In particular, we consider filters equal to \$0.125, \$0.25, and \$0.50 and allow trade in all contracts for which the absolute value of the price deviation exceeds the filter. Transaction costs are set at the same three round-trip levels. See also the note to table 6.

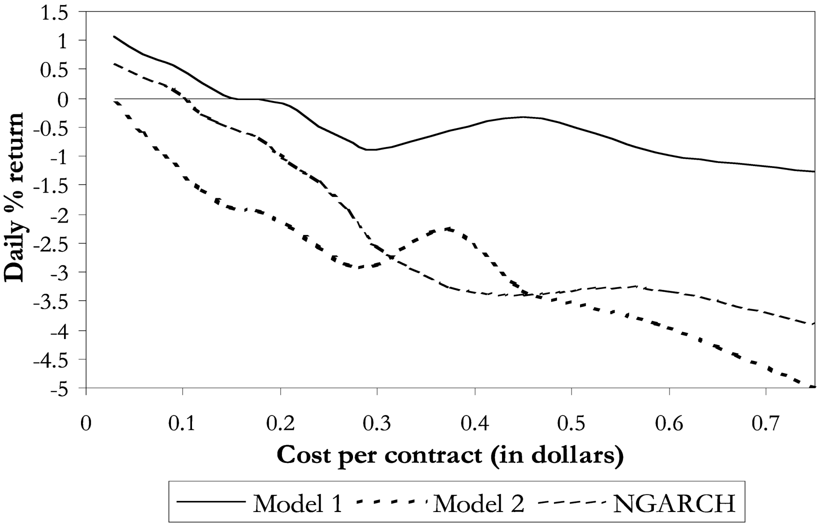
1 (cf. table 5). When we increase the filter/transaction cost to \$0.25, model 1 predicts positive profits, but they are not statistically significant; the implied Sharpe ratio is single-digit, below what would be guaranteed by a simple buy and hold daily strategy applied to the S&P 500 index; and the value of A becomes negative. All models predict negative profits when the filter/transaction cost of \$0.50 is applied. Therefore, it seems that as the level of transaction costs is progressively raised above \$0.25 (on a round-trip basis), mean daily profits for all models disappear; that is, for the levels of frictions that are most likely to be faced by small (retail) speculators, the strong statistical evidence of predictability in the IVS dynamics fails to be matched by equally strong evidence of a positive economic value.

To shed further light on the relationship between the profitability of trading rules that rely on our predictability findings and transaction costs, we perform a further experiment: we calculate the exact level/structure of transaction costs such that mean daily profits either are zero or stop being statistically significant at conventional levels. In particular, we apply a fixed \$10 commission to all transactions (i.e., an ex ante -1% return on a \$1,000 investment) and proceed to vary the per contract (round-trip) cost between \$0.02 and \$0.75. For comparison purposes with table 8, we apply this range of friction levels to trading rule D. We also apply the same structure of transaction costs to the underlying stock index. Results are reported in figure 5, where the upper panel reports mean daily percentage returns as a function of the per contract cost, and the lower panel shows related t-statistics. Clearly, the plots illustrate that both mean profits and their statistical significance disappear (and turn negative) as transaction costs are raised. In particular, it seems that for model 1, profits disappear when the cost per contract is around \$0.12–\$0.14, consistent with the findings in table 8. In practice, trading profits stop being significant already for \$0.10, whereas they eventually become significantly negative for per contract costs of approximately \$0.40. Interestingly, model 1 systematically outperforms both model 2 and the NGARCH model. In fact, model 2 never produces significantly positive profits, once the fixed commission is deducted.[22]

VI. Robustness

In this section, we present some additional results intended to check the robustness of our previous findings to two issues. One is the existence of measurement errors in the inputs entering the Black-Scholes formula (such as in the S&P 500 index level and/or in option prices). The second check we consider refers to the effects of bid-ask spreads on the rate of return calculations.

22. The plots display some nonlinear patterns that ought not be entirely surprising, since when transaction costs are raised, the implied filters are also increased in a way that makes trading (under rule D) more selective and possibly more profitable. This explains the flat (or even upward-sloping) segments generally obtained for intermediate costs, \$0.30–\$0.40.



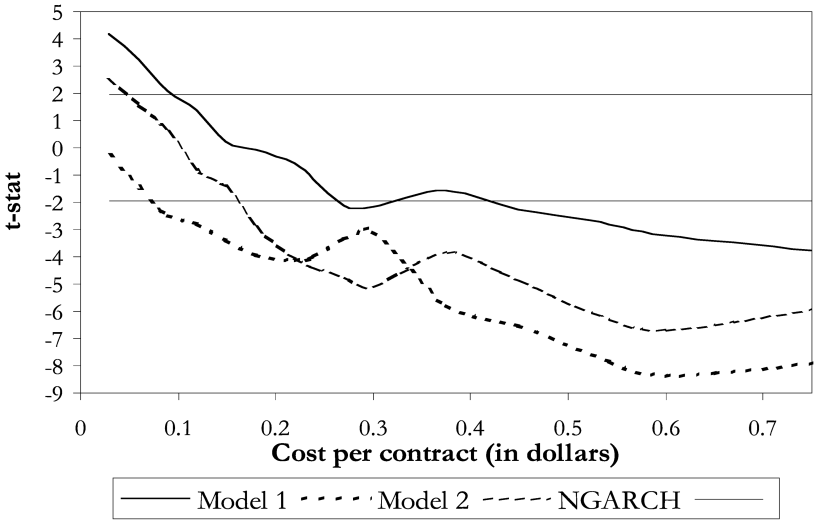


FIG. 5.—Mean percentage daily profits (with filters, rule D) as a function of the transaction cost per contract (plus a \$10 fixed cost): top: mean; bottom: *t*-statistics.

A. *Effects of Measurement Errors*

Hentschel (2003) has recently stressed that even small measurement errors in option prices or in the S&P 500 index level can produce large errors in implied volatilities for options away from the money. Thus it is important to investigate whether the presence of such measurement errors is driving our predictability results. As Hentschel shows, the existence of measurement errors in the underlying prices induces heteroskedasticity and autocorrelation in the errors of the cross-sectional IVS model (eq. [1] above). This implies that OLS estimates of $\boldsymbol{\beta}$ are inefficient. To obtain more efficient estimates of $\boldsymbol{\beta}$, and thus of implied volatilities, we follow Hentschel and reestimate equation (1), day by day, using a feasible GLS method. The details of the implementation of this method are in Appendix B.

Panel B of table 2 presents summary statistics for the feasible GLS estimates as well as for the adjusted R^2 and RMSE of implied volatilities. The estimates are, on average, similar to those obtained by OLS, with the exception of $\hat{\beta}_2$ and $\hat{\beta}_4$. The in-sample goodness of fit (as measured by $\bar{R}^2$ and RMSE) deteriorates only slightly under GLS estimation as compared with OLS. As before, the significant values of the LB statistics indicate that there is strong serial correlation (in levels and squares) in the estimates, suggesting a second-stage multivariate modeling approach.

Panel A of table 4 presents the out-of-sample forecasting measures i–vi defined in Section IV when the GLS estimates are used as the raw data in the second stage. On average, the RMSE and MAE of implied volatilities are slightly higher for all models, although interestingly the pricing RMSE and MAE are often lower than those obtained by OLS. Model 1 remains the best model out-of-sample, yielding a RMSE-V of 1.516 (vs. 1.429 under OLS) and a RMSE-P of 93 cents (vs. \$1 under OLS). It still clearly outperforms the benchmarks in terms of Black-Scholes pricing (MAE-P and RMSE-P) and percentage accuracy at predicting the direction of change.

In table 9 we present summary statistics for trading profits before transaction costs for trading rules A–C under GLS estimation. As obtained before under OLS (cf. table 6), model 1 performs best for trading rules A and B, yielding the highest average profit rates, with statistically significant *t*-ratios, large Sharpe ratios, and positive values of *A*. However, the use of GLS estimates implies a reduction of the mean profits for these trading rules, which is especially large in the case of trading rule B (the mean profit is now equal to 0.84% per day whereas before it was equal to 2.18%). Interestingly, for trading rule C, models 1 and 2 yield higher mean profits under GLS than under OLS.

Table 10 shows that these results are largely robust to the introduction of transaction costs, similarly to table 7. Even a commission fee of 12.5 cents per contract fails to completely remove the profitability of some of the trading rules, especially the selective rule C. Surprisingly enough, GLS estimation does even increase mean daily returns for trading rule C. This finding suggests that efficient estimation of the IVS may be important to improve the prediction accuracy in the segments of the IVS over which selective trading rules are most likely to produce buy and/or sell signals.

B. *Effects of Bid-Ask Spreads*

Although we have attempted to take into account the effects of transaction costs in computing trading profits, we have so far ignored the effects of bid-

TABLE 9 Trading Profits before Transaction Costs: Effects of Measurement Errors

	Mean Moneyness	Mean Time to Maturity	Mean Profit (%)	Daily % Standard Deviation	*t*-Ratio	Sharpe Ratio (%)	*A* Coefficient (%)
Trading rule A:							
Model 1	.0019	38.97	.0773	.0194	3.982	13.758	.0363
Model 2	.0019	38.51	.0480	.0197	2.436	6.913	.0068
NGARCH(1, 1)	.0009	38.90	.0545	.0194	2.805	8.511	.0135
Trading rule B:							
Model 1	−.0124	87.51	.8416	.1837	4.582	20.023	.2624
Model 2	−.0136	103.66	.0588	.1663	.354	1.110	−.4239
NGARCH(1, 1)	−.0127	85.43	.7056	.3477	2.029	8.832	−1.2875
Trading rule C:							
Model 1	.0052	115.92	1.6890	.1779	9.497	41.948	1.1419
Model 2	.0086	123.51	1.6530	.1868	8.847	39.070	1.0528
NGARCH(1, 1)	.0117	90.31	2.2146	.3751	5.905	26.146	−.0982
Benchmarks:							
S&P 500 buy and hold	NA	NA	.0166	.0287	.578	4.670	0
Random option portfolio	−.0119	51.23	−.1483	.1848	−.803	−4.008	−.7027
T-bill portfolio (random walk)	NA	NA	.0175	.0002	86.638	0	−.0174

NOTE.—This table reports trading profits deriving from various trading rules and models, as in table 6. The difference is that here we apply a feasible GLS procedure to estimate the cross-sectional parameters of the IVS each day. This method is more efficient than the OLS method applied before, under the presence of measurement error.

TABLE 10 Simulated Trading Profits under Trading Rule D after Transaction Costs: Effects of Measurement Errors

	Mean Moneyness	Mean Time to Maturity	Mean Profit (%)	Daily % Standard Deviation	t-Ratio	Sharpe Ratio (%)	*A* Coefficient (%)
			A. Filter: Transaction Cost = 12.5 Cents Round-Trip				
Model 1	−.0013	51.46	.0089	.2277	.039	−.168	−.8659
Model 2	−.0035	50.84	−1.8572	.4698	−3.953	−17.808	−5.4667
NGARCH(1, 1)	−.0045	43.69	−.7388	.1392	−5.309	−24.257	−1.0874
			B. Filter: Transaction Cost = 25 Cents Round-Trip				
Model 1	.0008	57.09	−.5682	.3554	−1.599	−7.354	−2.6487
Model 2	−.0029	53.50	−2.2630	.3665	−6.175	−27.773	−4.4726
NGARCH(1, 1)	−.0045	45.07	−2.4101	.6435	−3.745	−16.837	−9.1506
			C. Filter: Transaction Cost = 50 Cents Round-Trip				
Model 1	.0026	69.45	−.9700	.2321	−4.179	−18.987	−1.8774
Model 2	−.0020	57.18	−5.0065	.6158	−8.130	−36.411	−11.1828
NGARCH(1, 1)	−.0033	49.38	−3.5660	.5152	−6.921	−31.042	−7.8994

Note.—This table reports trading profits deriving trading rule D for various models, as in table 8. The difference is that here we apply a feasible GLS procedure to estimate the cross-sectional parameters of the IVS each day. This method is more efficient than the OLS method applied before, under the presence of measurement errors.

ask spreads since our simulated trading strategies have used observed closing prices (calculated as midpoints of the spread). Since actual transactions would have to take place inside the bid-ask spread but not necessarily at its midpoint, it is reasonable to assume that on average half of the bid-ask spread must be incurred as an additional transaction cost in the options market when a trade takes place, in addition to fixed commission costs. In this subsection, we try to take into account the effects of bid-ask spreads in our rate of return calculations.

Given that our data set does not include bid-ask spreads, we resort to Dumas et al.'s (1998) data set, which contains (transaction-based) information on bid-ask spreads at a weekly frequency (every Wednesday).[23] In order to complete our data set, we impute to all days within the same week of each Wednesday in Dumas et al.'s data set the bid-ask spreads sampled for that Wednesday.[24] Daily returns are computed as before, with the difference that we now simulate purchases at the ask (minus one-quarter of the spread) and sales at the bid (plus one-quarter of the spread), in addition to a fixed unit transaction cost. Obviously, these additional frictions represent an upper bound to the costs that would be actually incurred by a specialized trader, both because wholesale traders and market makers may essentially avoid the spread and because at times trades do take place well inside the spread.

Table 11 presents a summary of trading profits for trading rules A–C, under OLS and GLS estimation, when bid-ask spreads are taken into account. In addition to half of the bid-ask spread, we also apply a fixed commission of 5 cents per contract traded. Panel A of table 11 (OLS) is directly comparable to panel A of table 7. Clearly, incorporating bid-ask spreads lowers mean daily returns. Nevertheless, the strength of this reduction varies across strategies and models, as a function of the moneyness and time to maturity of the contracts traded. Out-of-sample results for trading rule C are particularly robust to the effects of bid-ask spreads. For this rule, large positive and abnormal returns remain after we introduce bid-ask spreads, with the more efficient GLS estimation yielding better out-of-sample results than OLS.

VII. Conclusion

Observed S&P 500 index option prices describe nonconstant surfaces of implied volatility versus both moneyness and time to maturity. The state-of-the-art practitioners' framework relies on simple linear regression models in which implied volatility is regressed on time to maturity and moneyness. The empirical evidence suggests that the coefficients of this model are strongly time-varying. In fact, structural models that have proposed economic justifications for the existence of an IVS also imply time variation in the IVS. When

23. The data were kindly provided by Bernard Dumas.
24. On average, the vector of spreads is (0.83 0.62 0.43 0.32 0.27)′ for DITM, ITM, ATM, OTM, and DOTM contracts, respectively.

TABLE 11 Simulated Trading Profits: Effects of Bid-Ask Spreads

	Mean Moneyness	Mean Time to Maturity	Mean Profit (%)	Daily % Standard Deviation	t-Ratio	Sharpe Ratio (%)
	A. Filter: Transaction Cost = 5 Cents					
OLS:						
Model 1	−.0026	78.43	.0005	.0747	.007	−1.012
Model 2	−.0028	64.42	−3.4873	.7377	−4.728	−21.206
NGARCH(1, 1)	−.0045	49.48	−2.8843	.5609	−5.143	−23.092
GLS:						
Model 1	.0022	72.99	−.0093	.1504	−.614	−3.258
Model 2	−.0022	59.12	−3.6985	.5278	−7.007	−31.423
NGARCH(1, 1)	−.0045	49.48	−2.8843	.5609	−5.143	−23.092
	B. Filter: Transaction Cost = 25 Cents					
OLS:						
Model 1	−.0005	78.84	−.2765	.0915	−3.022	−14.339
Model 2	−.0028	64.46	−4.1461	.8433	−4.916	−22.035
NGARCH(1, 1)	−.0045	49.48	−4.3600	.9486	−4.596	−20.596
GLS:						
Model 1	.0022	72.99	−.5833	.1803	−3.236	−14.867
Model 2	−.0022	59.16	−4.0547	.6412	−6.323	−28.344
NGARCH(1, 1)	−.0045	49.48	−4.3600	.9486	−4.596	−20.596

NOTE.—Transaction costs are set at 25 cents per contract, and bid-ask spreads are a function of the contract moneyness. Bid-ask spreads and transaction costs are applied on a round-trip bases as filters to obtain buy and sell signals. The first-stage cross-sectional IVS coefficients are estimated either by OLS or by GLS (adjusting for the likely effects of measurement errors involving option prices and the underlying index).

persistent latent variables drive the fundamental pricing equation, not only smiles, skews, and term structure effects in implied volatility are derived in equilibrium, but the resulting IVS is time-varying and therefore forecastable on the basis of information related to the latent factors. In this paper, we try to exploit this predictability by proposing a simple extension of the ad hoc practitioners model. We propose a two-step procedure for jointly modeling the cross-sectional and time-series dimensions of the S&P 500 index options IVS. In the first step, we model the cross-sectional variation of implied volatilities as a function of polynomials in moneyness and time to expiration (or functions thereof). Although the cross-sectional fit achieved by this step is largely satisfactory, we document the presence of considerable time variation and instability in the estimated coefficients. In the second step, we model the dynamics of the IVS by estimating parametric VAR-type models. We find that the two-step estimators produce a high-quality fit of the surface and of its changes over time.

We evaluate the forecasting accuracy of our modeling approach using both standard statistical measures and profitability-based criteria. In particular, the economic criteria assess the ability of generating abnormal profits by performing volatility-based trading that reflects the one-step-ahead predictions produced by the models.

Under a statistical perspective, we find that two-stage models provide accurate forecasts of future implied volatility and also satisfactory option price predictions (using the Black-Scholes formula, similarly to Noh et al. [1994]). These performances are competitive (often superior) to hard-to-beat benchmarks, such as a contract-by-contract random walk model.

Under an economic perspective, our evidence is mixed and depends heavily on the magnitude of transactions costs employed in the rate of return calculations and on how selective trading rules are. For less selective trading rules that imply a potentially large number of trades along the entire IVS (such as trading rule D), our volatility forecasts can support profitable trading strategies under low to moderate transaction costs only. However, when more selective rules are employed (such as trading rule C), we find that even under realistic assumptions on commission fees and bid-ask spreads, mean daily returns remain positive, statistically significant, and often truly in excess of what could be justified by their covariation with the returns on the market portfolio. Thus our finding that abnormal profitability depends on fine details of the trading rules and on assumptions on the strength of market frictions confirms that the existence of predictability patterns is not necessarily in contradiction with the notion of market efficiency.

There are a number of directions for future research that this paper leaves open. First, in this paper we have followed a two-step approach by first estimating the cross-sectional IVS coefficients each day and then modeling and forecasting the time series of these coefficients. An alternative method of estimation would consist of the simultaneous estimation of the cross-sectional coefficients and their dynamics by writing the IVS model in a state-space form and applying the Kalman filter to obtain maximum likelihood estimates. The one-step Kalman filter approach is theoretically more efficient than our two-step approach. Our main motivation for pursuing a two-step approach instead of an optimal one-step approach is simplicity: we view our method as a simple extension of what practitioners do already in practice and we show that it works well. We nevertheless realize that further gains in forecasting the IVS could potentially be obtained with a Kalman filter approach. We leave this interesting extension for future research. Second, additional experiments could be useful in terms of specifying the most useful prediction models. For instance, both Harvey and Whaley (1992) and Noh et al. (1994) find that there are substantial days-of-the-week effects in ATM implied volatility. It might be important to account for these kinds of effects also when modeling the entire surface. Additionally, Noh et al. show that there are considerable advantages in separately modeling the implied surface for call and put options. In this paper we have used data from both calls and puts, but we do not claim that this is an optimal choice. Finally, in our approach we estimate an unrestricted VAR model that does not exploit any nonarbitrage restrictions. Imposing such restrictions in our framework would entail writing a structural model for the IVS, which is beyond the scope of the present paper. We note, however, that imposing no-arbitrage conditions does not necessarily entail better forecasts. Indeed, our results suggest that our model (which does not exploit nonarbitrage conditions) outperforms Heston and Nandi's (2000) model, which is arbitrage-free.

Appendix A

Details on the Calculation of Leland's A Coefficient

This appendix provides additional details on the computation of Leland's (1999) risk measure. Following Rubinstein (1976) and Leland (1999), we make two fairly general assumptions: (i) the agent has power utility characterized by constant relative risk aversion coefficient γ, and (ii) the returns on the market portfolio are i.i.d. over time. Notice that assumption ii requires i.i.d.-ness of market portfolio returns only, not of the returns on all the existing assets, so that arbitrary patterns of dependence may be accommodated. Under these assumptions, it can be shown that for a generic portfolio characterized by gain process G,

$$E\left[\frac{G_{t+1}}{1{,}000}\right] = r + B(E[r_{\text{mkt}}] - r),$$

where

$$B \equiv \frac{\text{Cov}\,[E[G_{t+1}/1{,}000],\ (1 + r_{\text{mkt}})^{-\gamma}]}{\text{Cov}\,[r_{\text{mkt}},\ (1 + r_{\text{mkt}})^{-\gamma}]}.$$

This is a marginal utility–based version of the single-period CAPM, whose closed-form solution depends on the assumption of power utility and the identification of final wealth with the market portfolio. Interestingly, no assumptions are required for the preference parameter γ, since it turns out that

$$\gamma = \frac{\ln(E[1 + r_{\text{mkt}}]) - \ln(1 + r)}{\text{Var}\,[\ln(1 + r_{\text{mkt}})]}.$$

Once γ and B are estimated from the data, it is straightforward to calculate a marginal utility–adjusted abnormal return measure A as

$$A = E\left[\frac{G_{t+1}}{1{,}000}\right] - r - B(E[r_{\text{mkt}}] - r).$$

Measure $A > 0$ implies a return that exceeds what is accounted for by the quantity of risk absorbed by the agent, taking into account the shape of her utility function and therefore all higher-order moments of her wealth process.

For the purposes of our application, we proceed first to estimate γ from sample moments implied by 1992–96 S&P 500 index returns, obtaining a plausible $\hat{\gamma} = 6.81$. Next, we calculate B using data on daily trading strategy returns and the S&P 500. At that point calculation of (percentage) A is straightforward.

Appendix B

Details on the GLS Method Used to Filter Measurement Errors

This appendix gives some details on how to apply the GLS method proposed by Hentschel (2003) to obtain more efficient estimates of the parameters describing the cross-sectional IVS model used in the first stage of our approach. Consider the following equation:

$$\ln \sigma_i = X_i' \boldsymbol{\beta}, \tag{B1}$$

where $X_i = (1, M_i, M_i^2, \tau_i, M_i \times \tau_i)'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_4)'$. Here, σ_i denotes the true Black-Scholes implied volatility, and it is a function of the option pricing inputs (S, r, τ_i, and K_i) and of the option price P_i. The presence of measurement errors in the observed prices (such as S and P_i) implies that in practice we do not observe σ_i. Instead, we observe σ_i with an error. In our context, one way to formalize this idea is to suppose that the *observed* log-implied volatility, $\ln \tilde{\sigma}_i$, is equal to the true log volatility, $\ln \sigma_i$, plus an error $d \ln \sigma_i$:

$$\ln \tilde{\sigma}_i = \ln \sigma_i + d \ln \sigma_i. \tag{B2}$$

Substituting (B2) into (B1), we obtain

$$\ln \tilde{\sigma}_i = X_i' \boldsymbol{\beta} + d \ln \sigma_i. \tag{B3}$$

Equation (B3) is the cross-sectional IVS model that we will estimate in practice. It corresponds to our previous equation (1), with σ_i replaced by $\tilde{\sigma}_i$ and where $\varepsilon_i = d \ln \sigma_i$; that is, we identify the error term with the measurement error in implied volatility. Assumptions on the source and nature of this measurement error will enable us to further characterize the structure of the regression error. In particular, suppose that only measurement errors in S and P_i are present.[25] Then it follows that

$$d \ln \sigma_i = \frac{1}{\sigma_i} d\sigma_i = \frac{1}{\sigma_i} \frac{\partial \sigma_i}{\partial P_i} \left(dP_i + \frac{\partial P_i}{\partial S} dS \right) = \frac{1}{\sigma_i} \frac{\partial \sigma_i}{\partial \mathbf{x}_i'} d\mathbf{x}_i \equiv \frac{\partial \ln \sigma_i}{\partial \mathbf{x}_i'} d\mathbf{x}_i,$$

where $\mathbf{x}_i = (P_i, S)'$ collects the underlying prices and $d\mathbf{x}_i$ denotes the vector of corresponding measurement errors. Notice that

$$\frac{\partial \sigma_i}{\partial \mathbf{x}_i'} = \left(\frac{\partial \sigma_i}{\partial P_i}, \frac{\partial \sigma_i}{\partial S} \right) = \left(\frac{\partial \sigma_i}{\partial P_i}, \frac{\partial \sigma_i}{\partial P_i} \frac{\partial P_i}{\partial S} \right) = (\mathcal{V}_i^{-1}, \mathcal{V}_i^{-1} \Delta_i),$$

where $\mathcal{V}_i \equiv \partial P_i / \partial \sigma_i$ is the option's Black-Scholes vega, and $\partial P_i / \partial S = \Delta_i$ is the option's Black-Scholes delta. As in Hentschel (2003), we assume that measurement errors are mean zero and independent of each other, implying that

$$\operatorname{Var}(d \ln \sigma_i) = \frac{1}{\sigma_i^2} \frac{\partial \sigma_i}{\partial \mathbf{x}_i'} E(d\mathbf{x}_i d\mathbf{x}_i') \frac{\partial \sigma_i}{\partial \mathbf{x}_i} = \frac{1}{\sigma_i^2} \frac{\partial \sigma_i}{\partial \mathbf{x}_i'} \boldsymbol{\Lambda}_i \frac{\partial \sigma_i}{\partial \mathbf{x}_i} \equiv \frac{\partial \ln \sigma_i}{\partial \mathbf{x}_i'} \boldsymbol{\Lambda}_i \frac{\partial \ln \sigma_i}{\partial \mathbf{x}_i},$$

where $\boldsymbol{\Lambda}_i$ is a diagonal matrix with $\operatorname{Var}(dP_i)$ and $\operatorname{Var}(dS)$ on the diagonal, that is, $\operatorname{diag}(\boldsymbol{\Lambda}_i) = (\operatorname{Var}(dP_i), \operatorname{Var}(dS))'$. Because σ_i and the elements entering $\partial \sigma_i / \partial \mathbf{x}_i'$ (such as $\mathcal{V}_i$, Δ_i, and P_i) are option-specific, the above formula shows that the existence of measurement errors in option prices and index prices introduces heteroskedasticity.

25. Hentschel (2003) argues that this is the case for plausible values of the parameters.

Moreover, measurement errors in observed underlying prices (such as S) induce correlation among errors in implied volatilities. Thus OLS is inefficient, and we should instead use GLS to obtain more efficient estimates of $\boldsymbol{\beta}$ (and hence of fitted implied volatilities).

For a cross section of N options, we can rewrite (B3) in a compact form as follows:

$$\ln\tilde{\boldsymbol{\sigma}} = \mathbf{X}\boldsymbol{\beta} + d\ln\boldsymbol{\sigma},$$

with obvious definitions (e.g., $\ln\tilde{\boldsymbol{\sigma}}$ is the vector of the N observed implied volatilities). In particular, we can write

$$d\ln\boldsymbol{\sigma} = \frac{\partial\ln\boldsymbol{\sigma}}{\partial\mathbf{x}'}d\mathbf{x},$$

where $\mathbf{x} = (P_1, \ldots, P_N, S)'$ and $\partial\ln\boldsymbol{\sigma}/\partial\mathbf{x}'$ is the Jacobian matrix of log-implied volatility derivatives, $\partial\ln\sigma_i/\partial x_j$, with x_j denoting the jth element of $\mathbf{x}$. The variance-covariance matrix of the error vector $d\ln\boldsymbol{\sigma}$ is given by

$$\boldsymbol{\Sigma} = E(\ln\boldsymbol{\sigma}\ln\boldsymbol{\sigma}') = \frac{\partial\ln\boldsymbol{\sigma}}{\partial\mathbf{x}'}\boldsymbol{\Lambda}\frac{\partial\ln\boldsymbol{\sigma}}{\partial\mathbf{x}},$$

where $\boldsymbol{\Lambda} = E(d\mathbf{x}d\mathbf{x}')$ is a diagonal matrix with $\text{diag}(\boldsymbol{\Lambda}) = (\text{Var}(dP_1), \ldots, \text{Var}(dP_N), \text{Var}(S))'$. The GLS estimator of $\boldsymbol{\beta}$ is given by the well-known GLS formula

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\ln\tilde{\boldsymbol{\sigma}}.$$

In practice, this GLS estimator is not feasible since $\boldsymbol{\Sigma}$ is unknown. In particular, it depends on the measurement error variances (i.e. on $\boldsymbol{\Lambda}$) and on the unobserved values of S and σ_i.

In our application, we implement a data-driven choice of the elements of $\boldsymbol{\Lambda}$. For the choice of $\text{Var}(dS)$, we follow Hentschel (2003, 8) in computing an implicit "bid-ask spread" for the index level and then set $\sqrt{\text{Var}(dS)}$ equal to one-quarter of this bid-ask spread. More specifically, if returns are an i.i.d. random walk in calendar time with annual volatility σ^2, then the standard deviation of half-hour returns is approximately $\sigma_h^2 = \sigma^2/(365 \times 48)$. An implicit bid-ask spread can then be calculated as $\sigma_h^2 \times S$ so that $\sqrt{\text{Var}(dS)} = \frac{1}{4}(\sigma_h^2 \times S)$. In practice, we make $\sqrt{\text{Var}(dS)}$ explicitly time-varying by using each day the actual, closing S&P 500 index level and by calculating a time-varying σ^2 as the one-step-ahead predicted, annualized GARCH(1, 1) forecast obtained by using a rolling window of 10 years of daily S&P 500 returns data.[26] This feature accommodates the fact that time misalignments are bound to create larger measurement errors in days in which stock prices are more volatile, whereas GARCH models seem to offer, on average, good forecasting performance for volatility. As for our choice of $\text{Var}(dP_i)$, our main difficulty lies in the fact that our data set does not have options bid-ask spreads. We proceed as in Section VI.*B* by resorting to Dumas et al.'s (1998) data set to impute bid-ask spreads to our data. We follow Hentschel (2003) and set $\sqrt{\text{Var}(dP_i)}$ to one-quarter of the bid-ask spread. The time variation observed in the options spreads carries over to $\text{Var}(dP_i)$, which thus becomes time-varying.

We choose to replace the unobserved value of S by its observed level $\tilde{S}$. This is

26. Clearly, this contradicts the assumption of i.i.d.-ness of stock returns. However, this method seems to match common practice in applied finance.

consistent with the idea that measurement errors are zero mean so that the true, unobservable index is likely to be distributed around $\tilde{S}$ itself. As for $\boldsymbol{\sigma}_i$, we resort to Hentschel's iterative approach. We calculate first-step GLS estimates $\hat{\boldsymbol{\beta}}^{(1)}(\tilde{\boldsymbol{\sigma}})$ that use the "observed" implied volatilities $\tilde{\boldsymbol{\sigma}}$; on the basis of these first-step estimates, we obtain fitted implied volatilities $\hat{\boldsymbol{\sigma}}^{(1)} = \exp[X\hat{\boldsymbol{\beta}}^{(1)}(\tilde{\boldsymbol{\sigma}})]$, which are then used to calculate a second-step GLS estimate $\hat{\boldsymbol{\beta}}^{(2)}(\hat{\boldsymbol{\sigma}}^{(1)})$. The iterative process is applied until convergence of the resulting (feasible) GLS estimates is obtained, that is, when $\hat{\boldsymbol{\beta}}^{(k+1)} \simeq \hat{\boldsymbol{\beta}}^{(k)}$.

References

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen. 1997. Empirical performance of alternative option pricing models. *Journal of Finance* 52:2003–49.

Bakshi, Gurdip, and Zhiwu Chen. 1997. An alternative valuation model for contingent claims. *Journal of Financial Economics* 44:123–65.

Black, Fischer, and Myron Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81:637–54.

Campa, José Manuel, and Kevin Chang. 1995. Testing the expectations hypothesis on the term structure of volatilities. *Journal of Finance* 50:529–47.

Canina, Linda, and Stephen Figlewski. 1993. The informational content of implied volatility. *Review of Financial Studies* 6:659–81.

Christensen, B. J., and Nagpurnanand Prabhala. 1998. The relation between implied and realized volatility. *Journal of Financial Economics* 50:125–50.

Christoffersen, Peter, and Kris Jacobs. 2004. The importance of the loss function in option valuation. *Journal of Financial Economics* 72:291–318.

David, Alexander, and Pietro Veronesi. 2002. Option prices with uncertain fundamentals: Theory and evidence on the dynamics of implied volatilities. Working paper, University of Chicago.

Day, Theodore, and Craig Lewis. 1988. The behavior of the volatility implicit in the prices of stock index options. *Journal of Financial Economics* 22:103–22.

———. 1992. Stock market volatility and the information content of stock index options. *Journal of Econometrics* 52:267–87.

Diebold, Frank, and Canlin Li. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* 130:337–64.

Diebold, Frank, and Robert Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13:253–63.

Dumas, Bernard, Jeff Fleming, and Robert Whaley. 1998. Implied volatility functions: Empirical tests. *Journal of Finance* 53:2059–2106.

Fleming, Jeff. 1998. The quality of market volatility forecasts implied by S&P 100 index option prices. *Journal of Empirical Finance* 5:317–45.

Garcia, René, Eric Ghysels, and Eric Renault. 2005. The econometrics of option pricing. In *Handbook of financial econometrics*, ed. Y. Aït-Sahalia and L. P. Hansen. Amsterdam: Elsevier, North-Holland.

Garcia, René, Richard Luger, and Eric Renault. 2003. Empirical assessment of an intertemporal option pricing model with latent variables. *Journal of Econometrics* 116:49–83.

Gonçalves, Silvia, and Massimo Guidolin. 2003. Predictable dynamics in the S&P 500 index options implied volatility surface. Working paper, Université de Montreal and University of Virginia.

Gross, Larry, and Nicholas Waltner. 1995. S&P 500 options: Put volatility smile and risk aversion. Mimeo, Salomon Brothers, New York.

Guidolin, Massimo, and Allan Timmermann. 2003. Option prices under Bayesian learning: Implied volatility dynamics and predictive densities. *Journal of Economic Dynamics and Control* 27:717–69.

Harvey, Campbell, and Robert Whaley. 1992. Market volatility prediction and the efficiency of the S&P 100 index options market. *Journal of Financial Economics* 31:43–73.

Hentschel, Ludger. 2003. Errors in implied volatility estimation. *Journal of Financial and Quantitative Analysis* 38:779–810.

Heston, Steven, and Saikat Nandi. 2000. A closed-form GARCH option valuation model. *Review of Financial Studies* 13:585–625.

Jackwerth, Jens. 2000. Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* 13:433–51.
Jorion, Philippe. 1995. Predicting volatility in the foreign exchange market. *Journal of Finance* 50:507–28.
Leland, Hayne. 1999. Beyond mean-variance: Performance measurement in a nonsymmetrical world. *Financial Analysts Journal* 55 (January/February): 27–35.
Newey, Whitney, and Kenneth West. 1987. A simple positive semi-definite, heteroskedastic and autocorrelation consistent covariance matrix. *Econometrica* 55:703–8.
Noh, Jaesun, Robert Engle, and Alex Kane. 1994. Forecasting volatility and option prices of the S&P 500 index. *Journal of Derivatives* 1:17–30.
Peña, Ignacio, Gonzalo Rubio, and Gregorio Serna. 1999. Why do we smile? On the determinants of the implied volatility function. *Journal of Banking and Finance* 23:1151–79.
Poterba, James, and Lawrence Summers. 1986. The persistency of volatility and stock market fluctuations. *American Economic Review* 76:1142–51.
Rosenberg, Joshua, and Robert Engle. 2002. Empirical pricing kernels. *Journal of Financial Economics* 64:341–72.
Rubinstein, Mark. 1976. The valuation of uncertain income streams and the pricing of options. *Bell Journal of Economics* 7:407–25.
———. 1994. Implied binomial trees. *Journal of Finance* 49:781–818.
Tompkins, Robert. 2001. Implied volatility surfaces: Uncovering regularities for options on financial futures. *European Journal of Finance* 7:198–230.