Forecasting the Distribution of Option Returns*

Roni Israelov
AQR Capital Management

Bryan Kelly
University of Chicago

September 7, 2017

Abstract

We propose a method for constructing conditional option return distributions. In our model, uncertainty about the future option return has two sources: Changes in the position and shape of the implied volatility surface that shift option values (holding moneyness and maturity fixed), and changes in the underlying price which alter an option's location on the surface and thus its value (holding the surface fixed). We estimate a joint time series model of the spot price and volatility surface and use this to construct an ex ante characterization of the option return distribution via bootstrap. Our "ORB" (option return bootstrap) model accurately forecasts means, variances, and extreme quantiles of S&P 500 index conditional option return distributions across a wide range of strikes and maturities. We illustrate the value of our approach for practical economic problems such as risk management and portfolio choice. We also use the model to illustrate the risk and return tradeoff throughout the options surface conditional on being in a high or low risk state of the world. Comparing against our less structured but more accurate model predictions helps identify misspecification of risks and risk pricing in traditional no-arbitrage option models with stochastic volatility and jumps.

*We are grateful to Nicola Fusari, Don Chambers, and Qin Lu for sharing replication code for their papers, to Torben Andersen, Peter Carr, George Constantinides, Ian Dew-Becker, Christian Dorion, Darrell Duffie, Rob Engle, Steve Figlewski, Xavier Gabaix, Stefano Giglio, Lars Hansen, Zhiguo He, John Heaton, Steve Heston, Cliff Hurvich, Ronen Israel, Anil Kashyap, Lubos Pastor, Lasse Pedersen, Ernst Schaumburg, Marti Subramanyam, Viktor Todorov, Stijn Van Nieuwerburgh, Pietro Veronesi, Rob Vishny, and Dacheng Xiu, seminar participants at AQR, Maryland, Northwestern, NYU, and Chicago, and conference participants at the NYU Derivatives and Volatility Conference for helpful comments. We thank Mihir Gandhi and Carter Davis for excellent research assistance. The views and opinions expressed are those of the authors and do not necessarily reflect the views of AQR Capital Management, its affiliates, or its employees; do not constitute an offer, solicitation of an offer, or any advice or recommendation, to purchase any securities or other financial instruments, and may not be construed as such.

1 Introduction

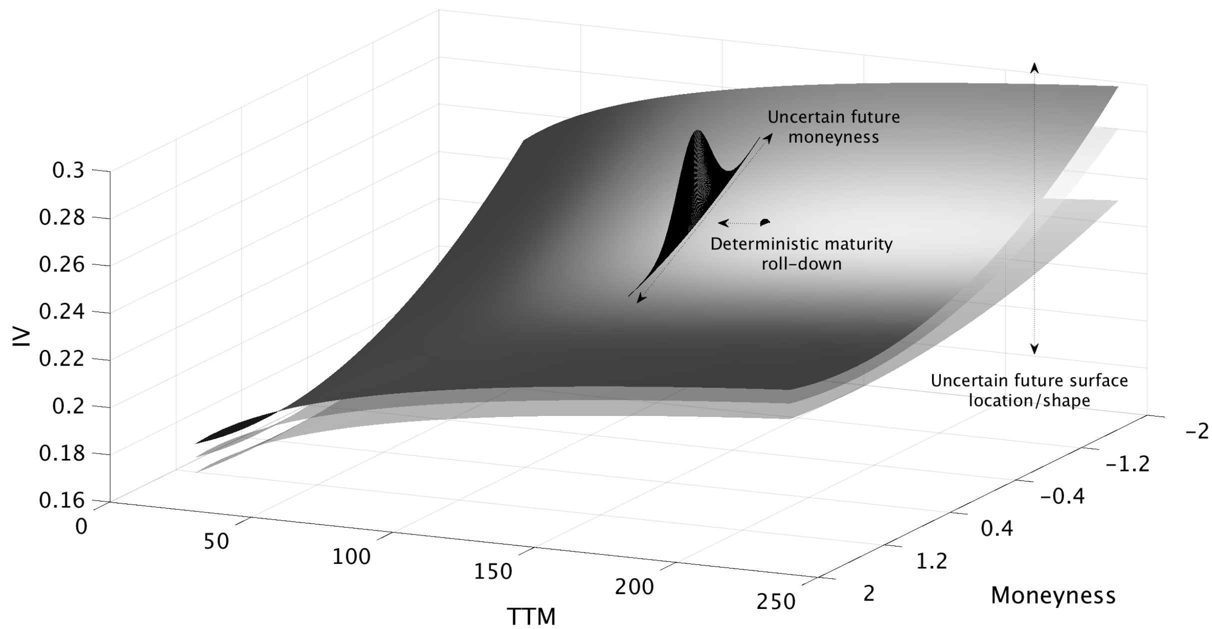
Equity index options allow investors to take a position in the aggregate equity claim that is contingent on the particular state of the world being realized. For example, selling a deep out-of-the-money put establishes a positive market exposure whose payoff only becomes activated in the event of a severe market downturn. The distribution of returns to this put encodes rich economic information. Dispersion in the put's returns, measured for example as standard deviation or extreme quantiles, characterizes the market risks an investor faces when holding this crash-contingent equity exposure. Measures of central tendency, such as the option's expected return and Sharpe ratio, describe the reward that investors demand in order to bear market exposure in crash states. More broadly, options markets present an opportunity to understand the risk and associated reward for bearing state-contingent exposures to the aggregate equity claim. Because options provide such a clear partitioning of the aggregate market state space, measuring their return distributions helps us understand investors' preferences regarding specific market outcomes.

The literature traditionally studies index option risk and reward through parametric models of the underlying price process. These models fully specify the distributional properties of the index spot price and derive options prices from no-arbitrage conditions. Among the many benefits of this approach are strict adherence to arbitrage-free pricing and the mathematical elegance of closed-form pricing formulae. At the same time, the parametric structure that permits these benefits also limits the specifications that can be tractably solved and reliably estimated. As we will show, a leading model calibration from this literature produces starkly counterfactual predictions for option return distributions. This reflects misspecifications that distort the models' description of risk premia, thus limiting the suitability of traditional models for inferring investor risk attitudes from options data.

1.1 Model Overview

Our approach to estimating state-contingent risks and risk compensation from the options data differs from the traditional approach. We use semi-parametric time series econometrics to model option return distributions. We begin with the proposal that uncertainty at time t about the future

Figure 1: Illustration of Option Forecast Uncertainty



Note. Hypothetical implied volatility surface with maturity axis described in days and moneyness axis described as the log distance between strike price and current spot price in annualized volatility units.

option price h periods ahead is describable in two layers, and illustrated in Figure 1. The first layer of uncertainty is summarized by the question "Where will the Black-Scholes implied volatility (IV) surface be located at time $t + h$?" Over time the empirical IV surface experiences frequent vertical shifts associated with the changing level of market volatility. The shape of the surface also evolves as changes in the relative pricing of claims alter surface slope and curvature in the moneyness and maturity dimensions. If one were to hypothetically hold contract moneyness and maturity fixed over time, uncertainty about future prices of all option contracts would be jointly summarized by uncertainty about the surface's future location.

The second layer of uncertainty is captured by the question "Where on the surface will the contract migrate to at $t + h$?" As time passes, the contract's time-to-maturity coordinate rolls toward zero deterministically. Its future moneyness coordinate, on the other hand, is unknown, because the distance between the underlying spot price and the contract's strike price depends on the realized index return. We thus describe the distribution of future options returns by jointly modeling these two sources of randomness: the future location of the surface and the future spot price. Then we translate the model's description of spot and surface uncertainty into uncertainty in option return space.

We implement our model in the following steps. First, on each day, we convert every contract price to an equivalent Black-Scholes implied volatility. Option prices can differ by orders of magnitude depending on how much time they have before expiration and whether they are at-the-money (ATM) or out-of-the-money (OTM). This transformation homogenizes prices to a comparable annualized volatility scale across strike price and maturity, which helps achieve stable estimates of option price dynamics.

Following a common practice in the literature, we postulate an implied volatility surface where individual traded contracts are the points in this surface observable by investors. On the x-axis is the option's "moneyness," which describes the distance between the current underlying price and the contract's strike price. On the y-axis is the contract's maturity, which represents number of days before the (European) option may be exercised. On the z-axis is the contract's implied volatility. The second step in our implementation is to build this IV surface for each day in our sample by interpolating the IV of traded contracts each day to a set of fixed moneyness and maturity coordinates. In doing so, we synthesize contracts whose identity is constant through time. This simplifies time series analysis of the surface by allowing us to estimate our model at static grid points. This avoids the complicated problem of modeling dynamics of traded contracts, whose identity is constantly changing as their moneyness fluctuates and maturity winds down.

Third, we estimate the model using data on the underlying index return and the panel of implied volatilities at constant moneyness/maturity grid points. We specify the system as a dynamic factor model whose backbone is a low-dimensional vector containing the key statistical factors that drive the IV surface. Surface factors follow a vector autoregression (VAR), and factor innovations are described by a multivariate GARCH model. To describe the full IV surface, common factors are mapped to all individual points on the grid using static loadings on contemporaneous factors. Thus, the entire forecasting machinery of our model is described by the factor VAR, and the static loadings are conduits to distribute VAR forecasts to all points in the surface.

The last step bootstraps factor model residuals to forecast the entire joint distribution of option prices. The economic questions that we pursue—regarding risk and risk compensation of option-based exposures—require not only an accurate mean forecast but a forecast of the entire return distribution. Rather than relying on a Gaussian or other parametric distribution for model innovations, bootstrapping builds up forecast distributions from the empirical distribution of innovations to non-parametrically match historical data. Each bootstrap draw of factor innovations is fed through the VAR, and forecasted factor values are mapped into forecasts for individual points on the surface via the estimated static factor loadings. This results in a joint forecast distribution for the future spot price and option IV's, thus summarizing both layers of model uncertainty—that regarding contracts' future moneyness and that of the surface's imminent location. Lastly, the resampled outcomes for the IV surface and contract moneynesses are converted to option prices by evaluating the Black-Scholes formula bootstrap draw by bootstrap draw. The final product is the joint forecast distribution for prices of all outstanding option contracts as well as the spot index itself.

1.2 Findings

Our model, which we refer to as the "ORB" (option return bootstrap) model for brevity, delivers a highly accurate description of the distribution of option returns. We forecast return horizons of one day up to two weeks. This choice is driven by the inherently short life span of options—the vast majority of contracts expire within one year, and a large mass of these have maturity less than one month. We construct multi-period forecasts by iterating one day forecasts, as opposed to re-estimating the model at a lower frequency. We also focus on daily delta-hedged option returns. The delta hedge purges option returns of mechanical variation associated with exposure to the underlying spot return. By dispensing with the comparatively well understood index return component, we are able to focus our analysis more narrowly on the component of return variation unique to the options market and arising, for example, from investor perception and pricing of variance and tail risk.

When evaluating forecast accuracy, we construct our predicted values on a purely out-of-sample basis so that an observation being forecasted never enters into any aspect of forecast construction.[1] We compare our model against two benchmarks. The first uses unstructured regressions of future option returns on contract attributes such as moneyness, maturity, and Black-Scholes "Greeks." These regressions are motivated by a growing empirical literature that relates expected option returns to contract characteristics. The second benchmark forecasts the option return distribution by simulating a traditional no-arbitrage model with stochastic volatility and price jumps (which we refer to throughout as the "SVJ" model).

Our first finding is that ORB possesses strong predictive power for mean option returns. For each contract day, the ORB model generates a complete forecast distribution for the future return. When we regress realized returns on the model's out-of-sample mean forecast in Mincer-Zarnowitz regressions, we find a predictive R^2 of 5.8% at the two week horizon (with comparable performance at the one day and one week horizon). Forecasts are equally powerful for short-dated versus long-dated options, and somewhat more powerful for OTM calls versus OTM puts. Overall the evidence indicates that ORB provides a vastly improved description of option risk premia compared to the characteristic and SVJ benchmarks, which deliver R^2's of 0.1% and 0.2%, respectively.

Next, we analyze our model's ability to accurately describe the ex ante volatility of an option position. We forecast realized absolute option returns using the average absolute return in the out-of-sample ORB forecast distribution. At the one day horizon our model predicts realized absolute returns with an R^2 of 30.6%. The benchmarks also have success in terms of volatility forecasts, with an R^2 of 24.4% based on a slew of characteristics and 4.5% from the SVJ model. When all predictors are included together, the bootstrap method stands out as the single most informative predictor of option return volatility. This remains qualitatively true at the two week horizon, though characteristic and SVJ forecasts contribute a relatively larger amount of predictive information at this longer horizon.

More than simply describing volatility, our model successfully forecasts the entire shape of

[1] In-sample, bootstrap residual distributions closely match the moments of residuals by construction (up to bootstrap uncertainty). This is not so on an out-of-sample basis, which is why our we focus on out-of-sample forecast performance in our analysis.

the future option return distribution. We show that ORB is remarkably accurate in predicting all return quantiles. For example, we find that model-based forecasts for the 1^{st} percentile of one day ahead option prices exceed the realized values for 0.9% of observations. Likewise, the model's conditional 99^{th} percentile exceeds the future realization in 98.9% if the data. Again, this is on a purely out-of-sample basis. The corresponding frequencies for quantile forecast exceedences in the SVJ model are 22.7% and 65.8%, indicating that the SVJ model fails to produce sufficient dispersion option price outcomes compared to the data. At the two week horizon, ORB tail risk forecasts deteriorate only slightly, with 1^{st} and 99^{th} percentile forecasts exceeding the realized value for 0.7% and 99.1% of observations, respectively. At longer horizons, SVJ quantile forecasts become more competitive, exceeding realized values 6.2% and 87.8% of the time, suggesting that traditional affine no-arbitrage models are better suited for describing lower frequency behavior of option prices.

Most importantly, our method establishes a new set of stylized facts about risk premia in options markets. Our model describes expected return, volatility, and Sharpe ratio "surfaces," which trace out the state-contingent risk and return relationship throughout the moneyness and maturity plane. We draw both unconditional surfaces and surfaces conditional on varying degrees of market risk. Based on the predictive accuracy of our approach, we argue that our model provides a far more accurate representation of ex ante option return distributions than available from alternative models. We show that especially large risk premia accrue to sellers of short-dated options, and particularly to sellers of OTM puts. These unconditional patterns may not be particularly surprising to those familiar with the empirical options literature. The great power of our approach is its ability to describe the *conditonal* risk-return tradeoff along the options surface, day by day. We show that the conditional expected return surface steepens dramatically in turbulent markets, as it becomes especially rewarding to sell shorter-dated options and to sell puts options with more negative moneyness. Because these are conditional moments, they generally cannot be calculated non-parametrically—i.e., from averages of historical data alone. As the conditioning set becomes finer and finer (for example, conditioning on days with higher and higher levels of market volatility), historical conditional averages begin to represent realizations instead of converging to the conditional expectation. Our model incorporates enough parametric structure to provide a true ex ante description of conditional option distributions, and the high degree of accuracy from our out-of-sample forecasts validates that ORB accurately reflects the true conditional distribution.

While other parametric models can also provide an ex ante description of the conditional return distribution, the key question is whether it is an accurate description. As an example, we show that some of the basic patterns regarding the risk-return tradeoff are reversed for the SVJ model, even unconditionally. The term structure of unconditional SVJ Sharpe ratios is increasing for puts, while in the data it is decreasing. And Sharpe ratios throughout the SVJ surface are orders of magnitude larger than those in the data. Understanding these types of discrepancies offers a valuable insight about how no-arbitrage model specifications can be improved to better fit the options market.

Finally, we illustrate the usefulness of ORB forecasts for solving practical economic problems such as risk management and portfolio choice. Because the model generates the joint conditional distribution between an option contract return and the underlying index return, it is simple to construct delta hedges and conditional value-at-risk estimates from the bootstrap forecast distribution. We show, for example, that out-of-sample bootstrap delta more effectively hedges exposure to the underlying market return than the standard Black-Scholes delta or the SVJ delta. More generally, the ORB model generates the *entire* joint conditional distribution among *all* outstanding option contracts at any given point in time. This means that the model can be used to construct ex ante optimized portfolios for arbitrary objective functions. As an example, we construct out-of-sample, conditionally mean-variance optimized portfolios that handily outperform simple strategies like naked selling of OTM puts and static risk reversals (selling OTM puts and buying OTM calls in a fixed proportion).

1.3 Related Literature

The literature on option returns can be divided into two categories. The first is purely empirical and typically studies option returns by sorting options into portfolios on the basis of some characteristic of the option contract or the underlying asset. It then tracks the subsequent returns of sorted portfolios and studies unconditional average returns, volatility, and Sharpe ratios of the characteristic-sorted option portfolios without a formal statistical model. Examples of this strand of literature include Coval and Shumway (2001), Bakshi and Kapadia (2003), Goyal and Saretto (2009), Frazzini and Pedersen (2012), Cao and Han (2013), Boyer and Vorkink (2014), Karakaya (2014), Vasquez (2016), and Israelov and Tummala (2017). Our approach differs from this literature in several ways. Most importantly, we develop a formal statistical model, and in doing so provide a comparatively rich description of risk and return in options markets. It characterizes the complete joint return distribution among all outstanding contracts and the spot, as opposed to specific moments of portfolios' marginal distributions. From this, one can immediately calculate a likelihood for any event in the bootstrap sample to, for example, measure ex ante crash probabilities, tail dependence, and associated hedging opportunities among contracts. And our model is specifically designed to move beyond unconditional characterizations and instead describe dynamic conditional distributions. The main thrust of our model evaluation shows that ORB forecasts are highly accurate representations of conditional option return distributions.

The second approach does not focus on option returns explicitly, but formally models option prices. This literature posits a distribution for the underlying spot price and for the stochastic discount factor then derives option pricing formulae by imposing no-arbitrage. The leading models in this tradition are those with affine stochastic volatility and jump specifications and are exemplified by Heston (1993) and Duffie et al. (2000). This literature does not study the distribution of option returns per se, perhaps in large part due to its analytical intractability.[2] The full distribution of option returns is nonetheless implicit in an affine model's physical ($\mathbb{P}$) and risk-neutral ($\mathbb{Q}$) densities for the underlying asset. Thus these models can be used to study option returns explicitly via simulation, as we do in this paper for the SVJ model, though this is not common in the literature.

While this literature has the great benefit of imposing economically meaningful no-arbitrage restrictions among contracts, there are also important disadvantages of the affine no-arbitrage approach. The range of specifications that retain closed form option price formulae is limited,[3] and their estimation is computationally intensive even in simple specifications and it is particularly

[2] Broadie et al. (2009) and Hu and Jacobs (2016) provide analytical characterizations of one aspect of the distribution, expected option returns, for certain affine models.

[3] For example, affine log volatility processes appear more consistent with the data than those affine in volatility levels (Chernov et al. (2003) and Amengual and Xiu (2016)) but do not admit closed form prices, thus suggesting that one must accept some amount of model misspecification in order to work with affine stochastic volatility models.

difficult to identify parameters of the $\mathbb{P}$ model. Perhaps the most challenging problem facing existing affine no-arbitrage frameworks is a tension between model complexity and feasibility of estimation. Option price behavior is complex—$\mathbb{Q}$ distributions appear to require several factors in order to describe the rich empirical patterns in option prices at various strikes and maturities.[4] On the other hand, the $\mathbb{P}$ distribution of the underlying index return can be very difficult to estimate. There is only so much resolution regarding stochastic jump intensities or volatility at multiple frequencies that one can glean from the time series of index returns. So, as $\mathbb{P}$ specifications become complex, the corresponding $\mathbb{P}$ model parameters become poorly identified. As a result, parsimony constraints on the $\mathbb{P}$ specification tend to bind and models have limited success in fitting the data.[5]

Our paper is related to the affine no-arbitrage literature in that we provide a structured description of joint pricing for all contracts based on a dynamic model with a small number of underlying factors. We diverge from this literature by advancing a statistical model without strictly imposing no-arbitrage. The key advantage of this approach is flexibility. We can estimate a variety of richly parameterized specifications, and do so with traditional time series models that can be estimated at trivial computational cost. Furthermore, to the extent that violations of no-arbitrage exist in the data (they do),[6] our model will naturally be able to better capture this behavior than a model that rules out arbitrage a priori. But our approach should be viewed as a complement to and not a substitute for models with no-arbitrage restrictions. Ours is a reduced form statistical approach for understanding empirical patterns in options markets. Its semi-parametric nature helps ensure that the patterns we identify are truly data driven rather than a distorted view of the data shaped by model misspecification. From the empirical insights of our statistical model, the literature can work toward developing improved no-arbitrage frameworks to match these insights. The goal in this

[4]For example, option prices are affected by time-varying demand pressures that may be easier to capture in a non-parametric model than in traditional no-arbitrage models (Garleanu et al., 2009).

[5]The need for rich $\mathbb{Q}$ models clashes with the need for a parsimonious $\mathbb{P}$ model. To avoid this tension, some literature uses rich $\mathbb{Q}$ specifications and abstracts from modeling $\mathbb{P}$ altogether. Leading examples in this vein are Bakshi et al. (1997) and Andersen et al. (2015a,b). This approach produces models that are very successful in terms of fitting options prices. But because $\mathbb{P}$ is unspecified, such models have no predictions regarding risk premia, and cannot be used to construct objects that are inherently tied to the $\mathbb{P}$ measure, such as the forecasts of realized option return distributions that we focus on in this paper. In related work, Broadie et al. (2007, 2009) emphasize the role of the $\mathbb{P}$ model for discriminating between affine model specifications and for drawing inferences about the sources of risk premia in the economy.

[6]See, for example, Ofek and Richardson (2003), Ofek et al. (2004), and Constantinides et al. (2009).

paper is to more directly infer the risk and compensation of state-contingent exposures with a flexible statistical approach directly modeling the behavior of options prices, rather than inferring this behavior from an overly restrictive and potentially severely misspecified model of the underlying.

Lastly, our paper is related to literature that models the volatility surface and its dynamics directly. This includes, among many others, Dumas et al. (1998), Aït-Sahalia and Lo (1998), Cont and Da Fonseca (2002), Fengler et al. (2003), Gatheral (2004), Daglish et al. (2007), Fengler et al. (2007), Gatheral and Jacquier (2014), Carr and Wu (2016), and Fengler and Hin (2015). We differ from this work by focusing on option return forecasts as opposed to price or implied volatility forecasts, and by focusing on forecasts of the entire joint distribution of returns via bootstrap.

The remainder of the paper proceeds as follows. Section 2 describes the data and our implied volatility surface construction. Section 3 specifies the statistical model. Section 4 evaluates the model's forecast distribution accuracy. Section 5 examines conditional risk premium surfaces and analyzes risk management and portfolio choice applications. Section 6 discusses outstanding issues, strengths and weaknesses of our approach relative to alternatives, and potentially fruitful follow-on work.

2 Data

Our empirical analysis focuses on the market for S&P 500 index options. Data are from OptionMetrics and cover the period from January 1996 through August 2015. This includes contract prices, underlying index values, and historical dividend yields and interest rates. We supplement this with data from the CBOE for the VIX index.

2.1 Implied Volatility Surface Construction

We construct an interpolated implied volatility surface that is the primary input to our statistical model. We track the surface on a fixed two dimensional grid of option moneyness and maturity coordinates. A contract i is defined by it's maturity date T_i and strike price K_i. The moneyness of i at time t is defined as:

$$m_{i,t} = \frac{\log(K_i/S_t)}{\mathrm{VIX}_t\sqrt{T_i - t}}, \tag{1}$$

where S_t denotes the underlying spot price and VIX_t level of the VIX index. This definition of moneyness describes the log distance between the prevailing spot price and the contract strike price in annualized volatility units. The literature often defines moneyness using contract-specific Black-Scholes implied volatility as the unit of volatility. We instead use VIX volatility units because it is convenient in our forecasting procedure to have moneyness of all contracts on a common volatility scale.

The grid points are set at 30, 60, 91, 122, 152, 182, 273, and 365 days to maturity, and moneyness values from -2 to 1 at increments of 0.25. When we interpolate to these points, we consider options with moneyness as low as -2.5 and as high as 1.5, and maturities of up to 450 days. This range contains the great majority of liquid contracts, spanning 90% of contract volume in the full sample, and 95% of the volume in the early (pre-2000) sample.

On each day, we construct synthetic IV for each grid point by interpolating traded contracts with a thin plate spline.[7] We focus on ATM and OTM contracts which are typically more liquid than ITM options, and perform separate interpolations for puts and calls due to frequent disagreement in their IV's at the same strike.[8] The data set that we ultimately bring to the statistical model consists of 4,949 repeated daily interpolated IV observations at 144 moneyness/maturity grid points.

3 Model

Our model is built on the idea that uncertainty about future option returns is describable in two layers: Uncertainty about the position and shape of the future IV surface, and uncertainty about the future moneyness of contracts. We build a low-dimensional vector autoregression for the common factors in our system, which is the main forecasting component of our model. Then, we use static factor loadings to map variation in the common factors to all points in the IV surface. Lastly, we convert variation at each point in the surface into a forecast for the distribution of future option

[7] We implement the spline interpolation in Matlab via the `fit` function with `thinplateinterp` fit type.

[8] Moneyness of the put surface ranges from -2 to 0.5, and from -0.5 to 1 for calls.

prices.

3.1 Specification

Let X_t denote the vector of L common factors. The first two elements of X_t are the underlying S&P 500 index return and the log of the VIX index. Uncertainty about future moneyness of all contracts is captured by the distribution of these two factors. When the S&P 500 index or the VIX moves, all contracts experience correlated shifts in their scaled distance between strike and spot, $m_{i,t}$.

The remaining elements of X_t describe common variation in the IV surface. IV for contracts at all moneynesses and maturities are strongly correlated with VIX. Thus, having VIX as the second element of X_t allows it to serve a dual role as the level factor for the IV surface. We supplement this with additional principal components of the IV surface. We build these additional factors by first orthogonalizing the time series log IV at each grid point against log VIX, and then extracting PCs from the orthogonalized log IV surface. These PCs tend to describe the changing slope and curvature of the surface in the moneyness and maturity dimensions.[9] We vary the number of PCs used across model specifications. We denote the general vector of estimated components as PC_t, and thus the full vector of common factors is $X_t = [r_t, \log \mathrm{VIX}_t, \mathrm{PC}_t']'$.

The time series model for factors is a VAR(1):

$$X_t = \mu + \rho X_{t-1} + \Sigma_{t-1}\epsilon_t. \tag{2}$$

Factor intercepts and autoregressive coefficients are static. Factor innovations are determined by an i.i.d. shock ϵ_t whose distribution is unspecified but that we assume has mean zero, variance one, and constant (potentially non-zero) correlations. Innovations have dynamic conditional covariances via Σ_{t-1}, which is a diagonal matrix of GARCH(1,1) volatilities.

The system's dynamics are extended from the factors to all points in the surface via static factor loadings. Denoting a given set of moneyness and maturity coordinates as (m, τ), the surface's factor

[9] The first PC of the surface is over 99% correlated with VIX.

structure specification is

$$\log IV(m,\tau)_t = \beta(m,\tau)[1, X_t']' + u(m,\tau)_t. \tag{3}$$

It is important that we specify the IV surface model and its factors in logs to avoid the possibility of negative variances at all points in the surface. The first element of β contains the grid-point specific intercept and the remaining elements describe contemporaneous regression sensitivities of gridpoint log IV to the common factors. Finally, we allow for residual serial correlation at each grid point:

$$u(m,\tau)_t = \psi(m,\tau)'[1, u(m,\tau)_{t-1}] + \phi(m,\tau)_{t-1}\eta(m,\tau)_t. \tag{4}$$

The residual innovation at each gridpoint, $\eta(m,\tau)_t$, is an i.i.d. shock with mean zero, variance one, and may have constant correlation across grid points. Residuals are subject to GARCH(1,1) volatility through $\phi(m,\tau)_{t-1}$.

3.2 Estimation and Out-of-Sample Forecast Construction

Our analysis focuses on out-of-sample forecasts. Let $\mathcal{T}$ = 4,949 denote the number of daily observations in the full data set from 1/1996–8/2015, and use an initial estimation window of 1,000 days and allow the estimation sample to expand over time. We construct one-step-ahead forecasts for each observation $t+1$ >1,000 as follows:

Step 1: Estimate using historical data. Define the estimation sample as the set of daily observations ending at t. In particular, the estimated model does not use in any way the $t+1$ contract observations to be forecasted.

Next, estimate static model parameters μ, ρ, and β, conditional volatilities Σ_t and ϕ_t, and the $t \times (K+N)$ matrix of historical model residuals $\mathcal{E} \equiv \{[\epsilon_\tau', \eta_\tau']\}_{\tau=1}^t$. Estimate the factor VAR coefficients μ and ρ via OLS regression and then estimate Σ_t from the OLS residuals. Similarly, estimate the log IV factor model from a time series OLS regression of log IV at each gridpoint onto the factors, then estimate residual serial correlation and GARCH from the regression residuals.[10]

[10]This multi-step approach is inefficient but practically useful. Estimating coefficients with regression, as opposed to using one-step joint MLE, greatly simplifies the computation of our procedure by avoiding repeated large scale

Finally, recover estimates of factor and surface innovation shocks as $\hat{\epsilon}_t$ and $\hat{\eta}_t$ by scaling regression residuals with their conditional volatility estimates. In what follows, hat superscripts indicate that a parameter, conditional variance, or residual is estimated with data ending at time t.

Step 3: Construct the option price forecast distribution via bootstrap. For each bootstrap draw $b = 1, ..., 5000$, randomly sample one row from $\hat{\mathcal{E}}$ maintaining its column ordering, and denote the sampled residuals by $\hat{\epsilon}^b_{t+1}$ and $\hat{\eta}^b_{t+1}$. Feed the bootstrap draws of $\hat{\epsilon}^b_{t+1}$ through the estimated VAR in equation (2) to construct the forecast distribution of one-period-ahead factors:

$$\hat{X}^b_{t+1} = \hat{\mu} + \hat{\rho} X_t + \hat{\Sigma}_t \hat{\epsilon}^b_{t+1}, \quad b = 1, ..., 5000.$$

This bootstrap sample includes the forecast distribution for the underlying index value at $t+1$ (via the index return, which is the first element of X) and for the VIX index at $t+1$ (the exponentiated second element of X). Together, these imply a forecast for the $t+1$ surface coordinates of each contract. In particular, for a contract i that matures at T_i with strike price K_i, its distribution of surface coordinate forecasts is given by

$$\tau_{i,t+1} = T_i - (t+1) \quad \text{and} \quad \hat{m}^b_{i,t+1} = \frac{\log\left(K_i/S_t \exp(\hat{r}^b_{t+1})\right)}{\hat{\text{VIX}}^b_{t+1}\sqrt{\tau_{i,t+1}}}, \quad b = 1, ..., 5000.$$

Next, construct the distribution of forecasted implied volatilities for each contract i. To do so, feed the bootstrapped factors ($\hat{X}^b_{t+1}$), surface residuals ($\hat{\eta}^b_{t+1}$), and moneyness coordinates ($\hat{m}^b_{i,t+1}$) through the estimated surface factor model of equations (3) and (4). Note that the model in (3) is only defined on a fixed set of grid points, while the contract's forecasted coordinates ($\tau_{i,t+1}, \hat{m}^b_{i,t+1}$) will generally lie between grid points. Therefore, interpolate the model estimates $\hat{\beta}$, $\hat{\psi}$, and $\hat{\phi}_t$ to each bootstrapped coordinate, and denote the interpolated values as $\hat{\beta}(\hat{m}^b_{i,t+1}, \tau_{i,t+1})$, $\hat{\psi}(\hat{m}^b_{i,t+1}, \tau_{i,t+1})$, and $\hat{\phi}(\hat{m}^b_{i,t+1}, \tau_{i,t+1})^b_t$. Use these in conjunction with the bootstrap draws to pro-

numerical optimization in our recursive out-of-sample approach. We do, however, use numerical optimization to estimate variance models. Because residual correlations are assumed constant, we need only estimate a sequence of univariate GARCH models, each having low computing cost.

duce implied volatility forecasts for individual contracts:

$$\hat{IV}^{b}_{i,t+1} = \exp\left\{\hat{\beta}(\hat{m}^{b}_{i,t+1}, \tau_{i,t+1})[1, \hat{X}^{b\prime}_{t+1}]' + \hat{u}(\hat{m}^{b}_{i,t+1}, \tau_{i,t+1})^{b}_{t+1}\right\}. \tag{5}$$

Finally, convert the implied volatility bootstrap forecast distribution to the distribution of forecasted option prices by evaluating the Black-Scholes formula at each individual bootstrap draw of $\{\hat{IV}^{b}_{i,t+1}\}^{5000}_{b=1}$.

The result of this procedure is a conditional one-step-ahead forecast for the price distribution of an individual contract, $\{\hat{P}^{b}_{i,t+1}\}^{5000}_{b=1}$. From the price distribution, the forecasted distribution of option returns is immediate. Importantly, because the bootstrap procedure draws factor innovations and surface errors row-wise (i.e., preserving their cross-sectional dependence), this procedure in fact forecasts the entire *joint* distribution of prices for option contracts across all outstanding strikes and maturities. Furthermore, it provides the joint distribution of these contract prices with any function of the future factor vector, such as the future index spot price, the VIX, and principal components of the IV surface.

3.2.1 Multi-period Forecasts

Our bootstrap procedure can be extended to form a J-step-ahead option price forecast with the following modifications. In Step 2, instead of drawing a single row of $\hat{\mathcal{E}}$, each bootstrap draw b will randomly sample J rows of $\hat{\mathcal{E}}$ to construct the sequence $[\hat{\epsilon}^{b\prime}_{t+1}, \hat{\eta}^{b\prime}_{t+1}], ..., [\hat{\epsilon}^{b\prime}_{t+J}, \hat{\eta}^{b\prime}_{t+J}]$. From this draw b, the J-step factor vector forecast is

$$\hat{X}^{b}_{t+J} = \left(\sum_{j=1}^{J} \hat{\rho}^{j-1}\right)\hat{\mu} + \hat{\rho}^{J} X_t + \sum_{j=1}^{J} \hat{\rho}^{J-j}\hat{\Sigma}_{t+j-1|t}\hat{\epsilon}^{b}_{t+j},$$

where the notation $\hat{\Sigma}_{t+j-1|t}$ emphasizes that the conditional information set for the volatility forecast is fixed at t. In addition to producing factor forecasts, this also delivers contracts' moneyness forecasts $(\hat{m}^{b}_{i,t+J})$. Similarly, $\hat{u}(\hat{m}^{b}_{i,t+J}, \tau_{i,t+J})^{b}_{t+J}$ is constructed from the sequence of $\hat{\eta}$ draws by iterating forecasts from equation (4) using parameters interpolated to gridpoint $(\hat{m}^{b}_{i,t+J}, \tau_{i,t+J})$.

Finally, the multi-step IV forecast is

$$\hat{IV}^b_{i,t+J} = \exp\left\{\hat{\beta}(\hat{m}^b_{i,t+J}, \tau_{i,t+J})[1, \hat{X}^{b\prime}_{t+J}]' + \hat{u}(\hat{m}^b_{i,t+J}, \tau_{i,t+J})^b_{t+J}\right\},$$

from which the J-step-ahead price forecast is immediately obtained by evaluating the Black-Scholes formula at this value.

3.3 Benchmark: SVJ

Throughout our analysis we provide forecasts from a traditional no-arbitrage model as a benchmark for comparison. We focus on the specification studied by Broadie et al. (2009) that incorporates stochastic volatility and jumps in the underlying index spot price, and which we refer to throughout as an "SVJ" model. We focus on the model and calibration in Broadie et al. (2009) for a number of reasons.

First, Broadie et al. (2009), along with the related paper of Broadie et al. (2007), emphasize estimating the physical distribution of the underlying index, and successful option return forecasts fundamentally rely on accurate estimates of physical distributions. Yet an important strand of the no-arbitrage options pricing literature does not specify the physical distribution of the underlying at all and instead focuses solely on estimating the risk-neutral model representation. In doing so, this literature excels at delivering accurate descriptive models with tiny option pricing errors. However, any model that abstracts from physical dynamics is unusable for the kinds of forecast applications that are the focus of our paper.[11] The emphasis that Broadie et al. (2009) place on estimating

[11] There is a clear tension between jointly estimating the risk-neutral *and* physical distributions of the index, versus only estimating the risk-neutral distribution. The availability of option prices for a wide cross section of strikes and maturities provides an abundance of detail about, and thus accurate estimates of, the risk-neutral distribution of the underlying index. Because of this, a researcher interested in a descriptive model that delivers small option pricing errors may focus on estimating the risk-neutral model alone. But enriching the risk-neutral specification and achieving more accurate option price fits faces the challenge that it is exceedingly difficult to estimate the corresponding physical distribution of the underlying. Investor risk preferences act as a wedge between option prices and the physical distribution of the underlying, and researchers typically rely on either the time series of underlying index returns or strong assumptions about the functional form of risk preferences in order to estimate the physical return distribution. Successful no-arbitrage models involve complex specifications for underlying index dynamics with multiple volatility processes and components of jump risk variation that are independent from volatility, as in Andersen et al. (2015a,b). There is no realistic hope of estimating the physical counterparts of these models from the time series of index returns, which is why researchers in search of a descriptive option pricing model sometimes abstract from specifying the physical distribution entirely.

the physical distribution make it an ideal benchmark for the behavior of forecast distributions in no-arbitrage models.

The second reason why Broadie et al. (2009) is an ideal benchmark SVJ calibration is because it uniquely focuses on option *return* implications of SVJ models, as opposed to targeting option price levels like most of the literature. This puts it especially closely in line with our focus on the behavior of option returns in our statistical model. Third, their calibration is based on data that most closely overlaps with the data we analyze.[12]

Details of the Broadie et al. (2009) model and calibration are described in Appendix A. The key model features are that the index follows a Brownian motion with diffusive stochastic volatility (as in Heston (1993)) and discontinuous price jumps that follow a Poisson mixture of normals (as in Merton (1976)) with static arrival intensity. Many other complicated no-arbitrage specifications have been proposed, including those with stochastic jump intensities, correlated price and volatility jumps, and non-Brownian increments. Our analysis of the Broadie et al. (2009) model is meant to serve as an indicative benchmark with many advantages discussed above, though it is admittedly not an exhaustive representation of the literature.

Because the SVJ model is fully parametric, we construct its conditional forecast distributions for option returns by simulating 5,000 future realizations for the index price and index volatility based on the calibrated physical model parameters reported in Broadie et al. (2009), then convert these simulated values to option prices by evaluating the model's pricing formula based on the calibrated risk-neutral parameters. Conditional forecasts on a given day require the value of the latent state variable (the stochastic volatility) to initialize forecasts. We retrieve the latent state on each day by inverting the model's pricing formula using the call option with maturity closest to 90 days and nearest-to-the-money strike.

[12] Broadie et al. (2009) study S&P 500 futures options from 1987 to 2003. Our data are on S&P 500 index options from 1996 to 2015.

4 Empirical Results

In this section we describe the empirical performance of our model. We defer our discussion of estimated model parameters and surface factors to Appendix B, and instead begin immediately describing behavior of the model's option price and return forecasts.

4.1 Case Study

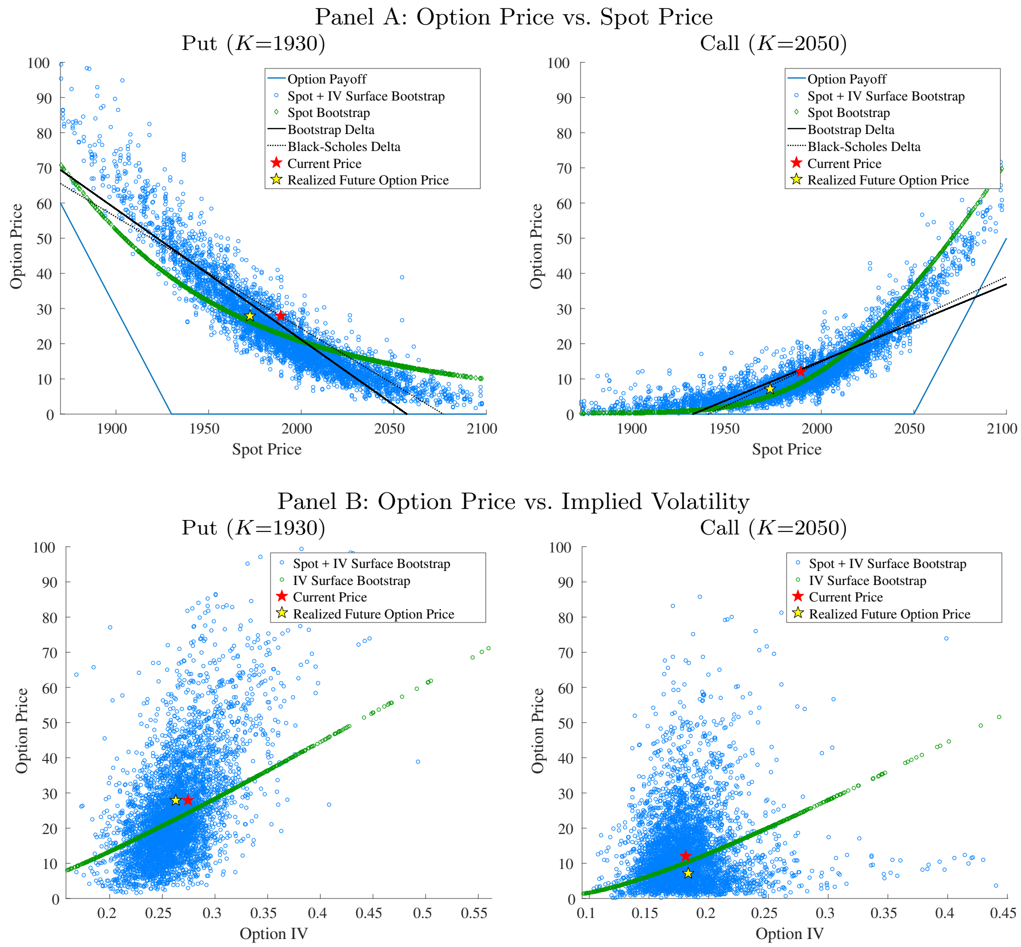
To provide a tangible introduction to the mechanics of the model, we describe how forecasts are generated for a handful of contracts on a single day. This case study takes the perspective of the last day in our sample, August 28, 2015, and considers forecasting one day ahead prices and returns using the ORB model with five surface factors (factor construction is Section 4.3).

On August 28, 2015, the S&P 500 index closed at 1989. The VIX index stood at 26, above its 1996-2015 sample mean of 21, amid a turbulent summer market whose volatility was driven in large part by a correction in the Chinese market of 40% between June and August. The top panel of Figure 2 describes one day ahead bootstrap forecasts for the price of an OTM put ($K = 1930$, $m = -0.5$) and OTM call ($K = 2050$, $m = 0.5$) with 21 days to maturity. The solid blue line plots the option payoff at maturity as a function of the underlying price on the horizontal axis. Red stars show the prevailing price of each contract.

First, in Panel A, we examine how price forecasts would behave if the only thing to change were the underlying spot price. The green diamonds show the distribution of future option prices when we bootstrap the underlying index return but hold the implied volatility surface fixed at its closing position on August 28. This distribution embodies the portion of the forecast distribution arising only from uncertainty about the contracts' future moneyness. This partial forecast traces out a curve relating future spot price to option price, ceteris paribus.

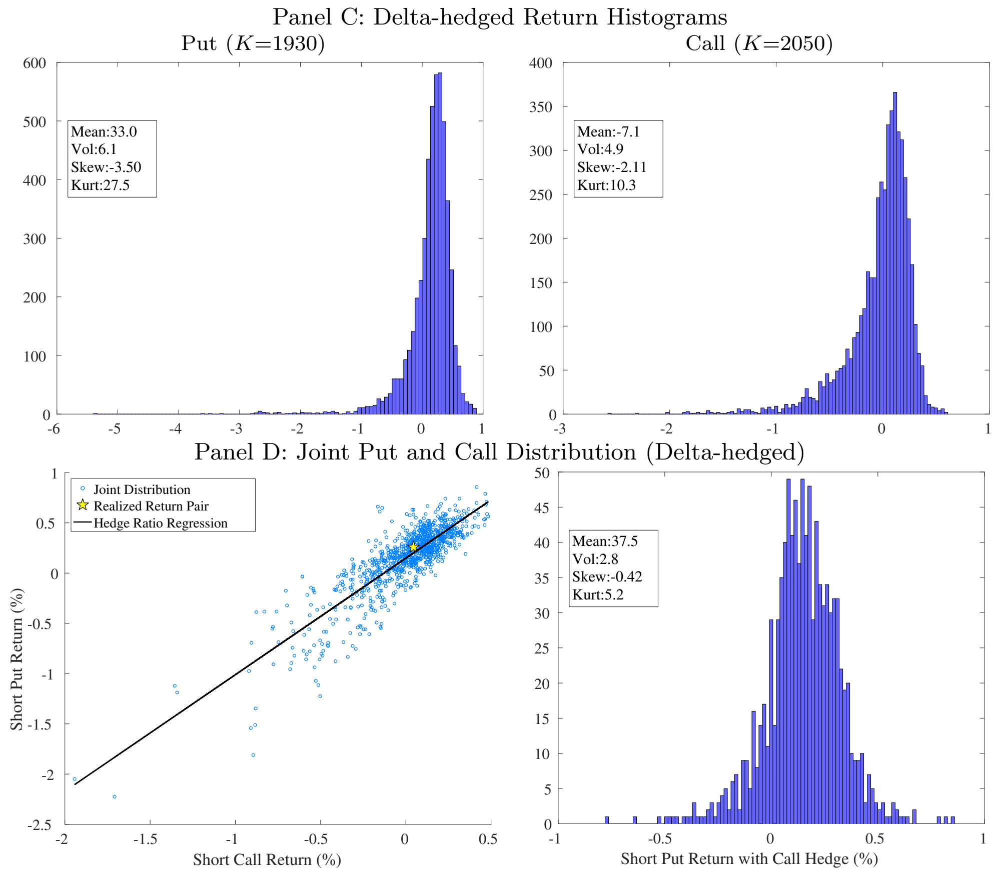
Similarly, in Panel B, green diamonds plot the forecasted price distribution holding the spot price fixed and only bootstrapping the future IV surface. This distribution embodies the portion of the forecast distribution arising only from uncertainty about the future position of the IV surface. This shows a smooth curve for the partial relationship between future IV to option prices, ceteris paribus.

Figure 2: Case Study: 21-day Options on August 28, 2015



Next, by bootstrapping both the surface factors and the spot price, the scatters of blue points in Panels A and B show the forecasted price distribution when uncertainty about both moneyness and the IV surface are accounted for. Surface uncertainty contributes a large amount of uncertainty. For example in Panel A, even if the spot price were to remain constant the distribution of price forecasts for the OTM put range from \$10 to \$35. More specifically, the cloud of blue points describes the conditional joint distribution of future spot prices and prices of the option contract. As such, it immediately delivers a bootstrap counterpart to the Black-Scholes delta. The bootstrap delta is simply calculated as the slope coefficient from a regression of bootstrapped put price

Figure 2 (cont.): Case Study: 21-day Options on August 28, 2015



Note. Out-of-sample forecast distributions for one day ahead prices and returns from the ORB model with five surface factors. Panel A shows scatter plots of bootstrap option prices versus bootstrap spot prices. Panel B scatters bootstrap option prices versus bootstrap IV. Panel C shows histograms and summary statistics of annualized delta-hedged percentage returns to selling options. Panel D shows a scatter of delta-hedged put returns versus call returns and the histogram and annualized summary statistics of returns to a delta-hedged risk reversal.

outcomes on the bootstrapped spot prices (the fitted regression line is shown in black, and the Black-Scholes delta is represented by the slope of the gray dashed line). The shape of the blue cloud also highlights important joint dynamics in the index prices and the surface. In states where the index falls, forecasted put and call prices are especially high relative to forecasts with a static surface (in green), reflecting the negative correlation between index returns and volatility. Similarly surface IVs and hence option prices are comparatively low when the index rises. As an implication, the slope of the black line in the put figure is steeper than a regression line fitted to the green curve would be, recommending a more aggressive delta hedge than that from Black-Scholes. The opposite is true for calls: The negative spot/volatility correlation dampens the bootstrap delta relative to Black-Scholes.

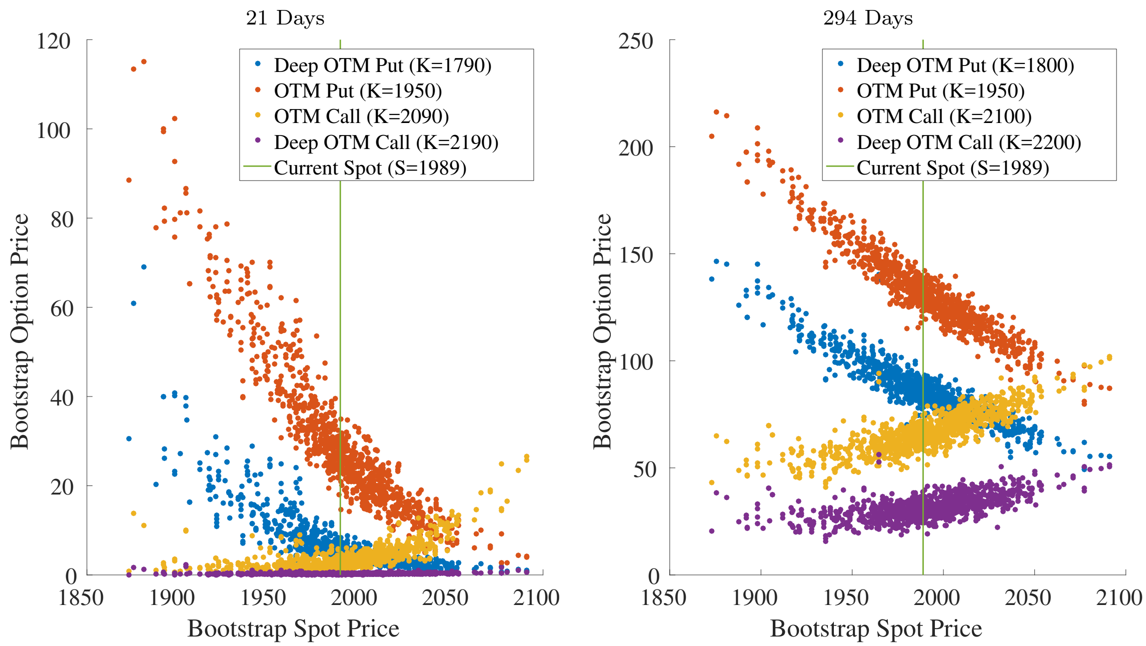
In turn, the bootstrap delta together with the simulated outcomes allows us to recover the conditional distribution of delta-hedged option returns. The ability to easily generate distributions of delta-hedged option returns is a boon for understanding subtler pricing behaviors of the option market by stripping out the large component of option returns that is mechanically correlated with the underlying index return. Panel C of Figure 2 shows histograms and annualized summary statistics for the conditional delta-hedged return distribution facing a seller of each contract. The bootstrap distribution highlights large deviations from normality in option return distributions. Sellers of these OTM contracts face extreme downside risks in the form of negative skewness (-3.5 for the put, -2.1 for the call) and excess kurtosis (27.5 and 10.3 for the put and call, respectively).

The right and left plots in Panel C are in fact closely related. They are both produced from the same set of bootstrap draws—that is, together they comprise a joint distribution of returns on two contracts. In the left plot of Panel D, we show this joint distribution explicitly by scattering outcomes for the delta-hedged put return against the corresponding draws for the call return. This scatter begs the question of whether the two contracts can be advantageously combined in a portfolio. A common option trading strategy designed to harvest the relative richness in put prices relative to calls is a risk reversal.[13] This strategy sells an OTM put option and hedges with a long OTM call position. But what is the risk-minimizing hedge ratio? In the model, a conditionally minimum variance hedge is constructed by regressing bootstrapped put returns on the corresponding bootstrapped call returns, represented by the black line. The slope of this line is the ex ante optimal hedge ratio, which in this example is equal to 1.15.[14] The bootstrapped return distribution for the risk reversal with a bootstrap hedge ratio is shown in the right figure. Compared to the individual put and call returns, the risk reversal has attractive ex ante expected performance. It earns nearly the same expected returns as selling the put outright, but with half

[13] See, for example, the CBOE's risk reversal index: http://www.cboe.com/products/strategy-benchmark-indexes/risk-reversal-index.

[14] For comparison, the hedge ratio to make the risk-reversal zero net investment is 2.30, to make it delta-neutral is 1.31, to make it vega-neutral is 1.14, and to make it gamma-neutral is 0.75.

Figure 3: Case Study: Options on August 28, 2015



Note. Out-of-sample forecast distributions for one day ahead prices from the ORB model with five surface factors. The figure shows scatter plots of bootstrap option prices versus bootstrap spot prices for a variety of strikes and maturities.

the volatility, and less kurtosis and negative skewness.

While this risk reversal example illustrates the potential usefulness of bootstrap option forecasts for problems in risk management and portfolio choice, it understates the full richness of the conditional distribution embodied ORB. Figure 3 shows the distribution of eight contracts against the underlying, ranging from very short dated (21 days) to long dated (294 days), and from deep OTM puts ($m = -1.7$) to deep OTM calls ($m = 1.5$). On August 28, 2015, over 750 different S&P options were trading. The bootstrap delivers the conditional joint distribution of *all* of these, in additional to their joint distribution with the underlying return, VIX, and other surface factors.

4.2 Forecast Evaluation

The previous section described the behavior of bootstrap forecast distributions for a handful of contracts on one particular day. In this section, we conduct a large scale evaluation of the accuracy of forecast distributions across all contracts on all days.

4.2.1 Out-of-Sample Forecast Construction

To avoid any concern that forecast accuracy is due to in-sample overfit or look-ahead bias, we evaluate our model on a purely out-of-sample basis. We do this using a recursive estimation procedure which ensures that any observation being forecasts is strictly excluded from the estimation sample used to construct the forecast.

In particular, consider a forecast of an option price or return at $t + h$ given information as of date t. The h-period ahead out-of-sample forecast estimates the model only using observations in the estimation window $1, ..., t$, thus mimicking the information set that a market participant would have had access to in real time. From the backward-looking window we produce estimates of static model parameters (μ, ρ, etc.), conditional variances (Σ_t, ϕ_t), and the pool of model residuals ($\mathcal{E}$) from which bootstrap samples are drawn. Then, we use these backward-looking estimates to construct forecasts beyond period t.

We construct h-period ahead out-of-sample forecasts recursively. Our initial estimation sample includes the first 1,000 days of data (January 1996 through mid-December 1999). Estimates based on observations 1 through 1,000 produce forecasts for option prices and returns at date 1,000+h. Next, we expand the estimation window by one day to include 1 through 1,001, and use these observations to construct forecasts for date 1,001+h. We iterate this procedure until we have exhausted the full 4,949 days in our sample. As we expand the estimation window, model estimates become more precise and the pool of residuals from which we bootstrap becomes larger.

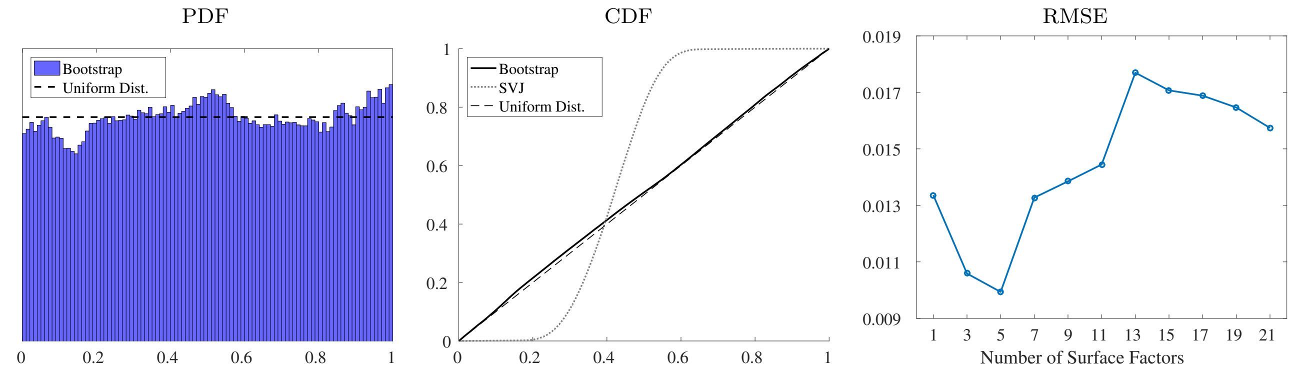
4.3 Specification Choice

The model formulation of Section 3 allows the surface factor vector to include a general number of principal components from the IV surface, in addition to including the S&P 500 index return and log VIX. We choose the number of principal components by searching across specifications to find the model that has the best out-of-sample forecast performance. Optimizing with respect to out-of-sample performance is fundamentally different than optimizing in-sample. In-sample, enriching the model specification generally improves performance in data fit relative to simpler nested specifications. In many settings, the in-sample fit improvement is mechanical and is in fact overfit. This is not the case when fits are evaluated on an out-of-sample basis. A rich specification that overfits the data in-sample will tend to perform poorly out-of-sample. We compare out-of-sample performance across different choices for the number of surface factors. In the simplest specification, we consider a single surface factor—the log VIX. We gradually increase the number of factors by adding PCs of the log IV surface. Whenever we expand the specification, we add one PC each from the put surface and call surface (orthogonalizing each added call PC against the added put PC). Thus the number of factors we consider proceeds in increments of two.

As the focus of our paper is describing conditional distributions of option returns (as opposed to just the conditional mean option return), we use a measure of forecast accuracy motivated by the density forecasting literature. Following Diebold et al. (1998), we compute the probability integral transform (PIT) for each observation in our sample based on the bootstrap distribution from our model. Consider a realized option price (denoted $p_{i,t+1}$) in our data set for contract i on date $t+1$. Given time t information, our model describes a conditional cumulative forecast distribution $\hat{F}_{i,t+1|t}$ of $p_{i,t+1}$ via bootstrap. The probability integral transform evaluates the forecast distribution at the realized data value, $PIT_{i,t+1} = \hat{F}_{i,t+1|t}(p_{i,t+1})$. As shown in Diebold et al. (1998), $PIT_{i,t+1}$ follows an independent standard uniform distribution when $\hat{F}_{i,t+1|t}$ is the CDF of the true data generating process. The logic behind this result is intuitive. A good forecast model will see data realizations fall below the forecasted 1^{st} percentile approximately 1% of the time, below the forecasted 10^{th} percentile approximately 10% of the time, and so forth. And the PIT for an observation will be independent if the model correctly accounts for all conditioning information.

One way that we assess the forecast accuracy of our model is by comparing the distribution of model-based PIT to a uniform distribution. The left panel of Figure 4 shows the empirical PDF histogram of out-of-sample PIT from the bootstrap model based on three surface factors (the log VIX plus two additional surface PCs) pooling all contract-days in our sample. The dashed black line shows the ideal uniform distribution. The middle panel shows the same empirical distribution of PITs from the bootstrap model in the form of a CDF, as well as the ideal uniform CDF. To

Figure 4: Out-of-Sample Forecast Accuracy



Note. The left figure is a histogram of PITs for all contracts in our sample based on the ORB model with five surface factors. The center figure draws the corresponding empirical CDF along with the ideal uniform CDF, and the CDF of PITs based on the SVJ model. The right figure shows the RMSE of PIT distribution relative to the ideal uniform distribution for a variety of model specifications.

give a sense of the forecast accuracy implied by this PIT distribution, we plot the distributions of PITs from the SVJ model. The CDF from the bootstrap model closely tracks the 45-degree line, indicating the model's accuracy in describing the conditional distribution of future option prices. In contrast, the flatness of the SVJ CDF at extreme quantiles indicates that the SVJ model produces a forecast distribution that is too narrow to describe the realized data.

The right panel compares the distance between the bootstrap model CDF and the ideal uniform CDF for different numbers of surface factors. For example, for the bootstrap model with three surface factors shown in the middle panel, the RMSE is calculated by averaging the squared vertical distance between the model CDF and uniform CDF for every data point, then taking the square root. We repeat this for the distribution of PITs for a one-factor surface (only the log VIX) all the way to an 21-factor surface (log VIX plus ten PCs from the call surface and ten from the put surface). The best model—that is, the model whose out-of-sample PIT's have the smallest divergence from uniformity—uses five surface factors. Based on this criterion, we focus the remainder of our analysis on the model with surface factors given by the log VIX plus two additional PCs each from the call and put surfaces.

4.4 Mean Forecasts

We first assess the accuracy of our model's conditional forecast distributions by evaluating the accuracy of mean return forecasts. Throughout the paper we define returns from the perspective of an investor selling the option that dynamically delta-hedges via Black-Scholes each day. In particular, the h-period excess return to selling contract i at $t+h$ is

$$r_{i,t+h} = \frac{1}{S_t}\left(P_{i,t} - P_{i,t+h} + \sum_{j=1}^{h} \Delta_{BS,i,t+j-1}[S_{t+j} - S_{t+j-1}] + r_{f,h,t}P_{i,t}\right)$$

where Δ_{BS} denotes the standard Black-Scholes delta and $r_{f,h,t}$ is the h-period risk-free rate. This return is defined relative to the basis of the initial price of the underlying, rather than with respect to the initial contract price. We do this in order to avoid extreme behavior of returns for deep OTM contracts that have very small initial values. This helps our analyses by ensuring that results throughout the paper are not driven by the return variation of contracts with values near zero. And note that the choice of basis drops out of our Sharpe ratio computations throughout.

For each contract i on date t we have a the full bootstrap conditional distribution for the contract's price (and thus its return) at $t+1$. We calculate the bootstrap conditional mean return forecast as $\hat{E}_t[r_{i,t+1}] \equiv \frac{1}{B}\sum_b \hat{r}^b_{i,t+1}$, which is a pure out-of-sample forecast constructed using only data through date t. Then, we regress the realized return $r_{i,t+1}$ on the bootstrap conditional mean pooling all observations in our sample:

$$r_{i,t+1} = c_0 + c_1\hat{E}_t[r_{i,t+1}] + e_{i,t+1}. \tag{6}$$

Equation (6) is a Mincer and Zarnowitz (1969) regression and is a commonly used forecast evaluation tool. The ideal forecast will produce an intercept of zero, a slope of one, and unforecastable residuals. More broadly, the magnitudes and significance of c_0 and c_1 are informative about predictive content and biasedness of the model-based mean forecast.

We report conditional mean forecasting results in Table 1. Column 1 establishes an agnostic baseline for the extent of predictive information in commonly studied contract-level conditioning

Table 1: Mean Return Forecasting Regressions

	1	2	3	4	5	6	7	8
ORB		0.538		0.552		0.540		0.552
		(11.0)		(11.3)		(17.2)		(17.4)
SVJ			0.008	0.000			0.005	-0.004
			(6.1)	(0.1)			(4.4)	(2.3)
Money	0.007			0.003	-0.008			-0.009
	(1.4)			(0.5)	(2.2)			(2.4)
TTM	-0.002			0.001	0.000			0.003
	(1.9)			(0.5)	(0.3)			(3.5)
Gamma	0.000			0.006	-0.003			0.003
	(0.2)			(3.1)	(2.5)			(1.7)
Vega	0.004			0.003	-0.006			-0.003
	(2.4)			(1.8)	(4.2)			(1.8)
Theta	-0.001			0.003	0.003			0.003
	(0.5)			(0.8)	(0.9)			(0.9)
IV	0.005			0.012	0.023			0.013
	(0.9)			(2.1)	(6.9)			(3.8)
Money*Put	-0.007			-0.008	0.017			0.008
	(1.5)			(1.4)	(4.4)			(1.8)
TTM*Put	0.003			0.000	0.002			0.000
	(2.0)			(0.2)	(1.5)			(0.1)
Gamma*Put	0.001			0.000	0.002			-0.001
	(0.7)			(0.1)	(1.5)			(0.4)
Vega*Put	-0.002			0.000	0.002			0.002
	(1.4)			(0.2)	(1.6)			(1.7)
Theta*Put	-0.003			-0.007	-0.008			-0.007
	(1.0)			(1.8)	(3.2)			(2.4)
IV*Put	-0.005			-0.014	-0.024			-0.018
	(1.0)			(2.7)	(5.7)			(4.4)
R^2 (%)	0.1	5.6	0.2	5.9	0.8	11.5	0.1	12.1
Date FE	No	No	No	No	Yes	Yes	Yes	Yes
N (1000s)	919.3	919.3	919.3	919.3	919.3	919.3	919.3	919.3

Note. Pooled forecasting regressions for delta-hedged returns to selling option contracts. Characteristic coefficients are reported as 100 times the characteristic's standard deviation. ORB and SVJ predictors are means of the models' simulated return distribution for each contract-day. Standard errors are clustered by day.

variables with a linear model, and is motivated by the characteristics frequently used to form portfolios in the options literature. In particular, we regress $r_{i,t+1}$ on time t values of contract moneyness, time-to-maturity, Black-Scholes implied volatility, and Black-Scholes gamma, vega, and theta. We also include the interaction of each characteristic with an indicator for whether the option is a put to allow for differences in the association between contract characteristics and returns for puts versus calls.[15] The explanatory power in this regression is small ($R^2 = 0.1\%$), and only two regressors (vega and maturity$\times$put) are significant at the 5% level (all standard errors in Table 1 are clustered by day).

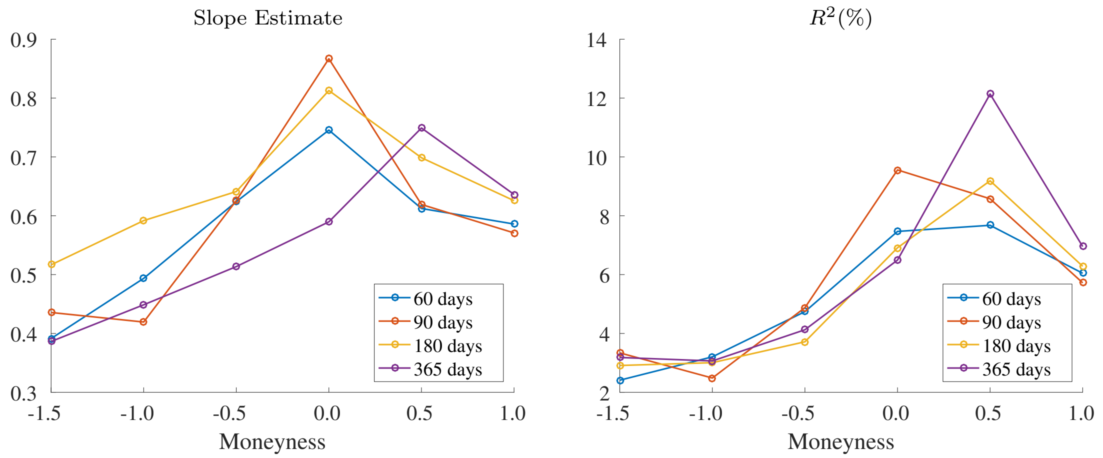
Column 2 reports the Mincer-Zarnowitz regression for forecasts from the bootstrap model. The R^2 of the regression is 5.6%, which is remarkable given that forecasts are made out-of-sample and that the return horizon is one day. The slope coefficient is 0.54 with a t-statistic of 11.0, indicating that the model has strong and highly significant predictive content in a mean forecasting sense. While the slope is statistically far from zero, it is also statistically far from one, and this attenuated slope indicates forecast imperfection likely due to a combination of model misspecification and noisy parameter estimates. The untabulated intercept is 0.000055 ($t = 3.2$), which is nearly the same magnitude as the unconditional mean return and 5% of the unconditional standard deviation of returns in our sample, indicating a small positive and significant forecast bias.

For comparison, we report Mincer-Zarnowitz results for the SVJ model in column 3. The slope coefficient is 0.008 ($t = 6.1$), indicating quantitatively weak but statistically significant predictive content in SVJ mean forecasts (the SVJ intercept is insignificantly different from zero). SVJ predictive power is also weak when viewed in terms of R^2 (0.1%). When the bootstrap mean, SVJ mean, and contract characteristics are included jointly in the regression of column 4, we essentially recover the regression of column 2, indicating that the bootstrap mean subsumes the forecasting content of all predictors considered.

Columns 5–8 repeat these regressions with date fixed effects. This more closely mimics the characteristic-based sorting approaches used in much of the literature. This regression is useful for understanding whether a predictor is useful for explaining cross section variation in returns, even if the predictor is not especially useful in a time series sense. The R^2 we report for fixed effects regressions is an incremental R^2. It describes the fraction of return variation explained by the regressors after removing the explained variation from fixed effects. Indeed, characteristics are

[15] Our option moneyness measure stands in for the Black-Scholes delta. The regressors are nearly collinear when included together with the put indicator. Replacing moneyness with delta has negligible impact on the regression.

Figure 5: Out-of-Sample Mean Forecast Accuracy by Moneyness/Maturity



Note. Forecasting regressions for delta-hedged returns to selling option contracts within moneyness/maturity bins using the mean of the ORB forecast distribution for each contract-day.

comparatively effective in explaining cross section variation in returns, with an R^2 of 0.8%. But the main conclusion from the first four columns is unchanged—the bootstrap forecast essentially subsumes the predictive information among all of the predictors we consider.

Whereas Table 1 pools all contract-days, Figure 5 describes the accuracy of conditional mean forecasts from the bootstrap model in subsamples. We double sort observations into bins with moneyness ranging from -2 to 1 in increments of 0.5 and maturity of 30–60, 61–90, 91–180, and 181–365 days. Then, within each of the 24 resulting moneyness/maturity bins, we estimate regression (6) (with no date fixed effects). The left panel of Figure 5 shows the estimated slope coefficients and the right panel shows regression R^2s. Slopes range from 0.39 to 0.87 and are highly significant with all p-values below 0.0001. The corresponding bin R^2s range from 2.4% to 12.1%. Bootstrap mean forecasts are thus highly informative throughout the moneyness/maturity plane, and are especially strong for ATM as opposed to OTM options, and for calls rather than puts.

4.5 Mean Forecasts: Multi-period Returns

Next, we analyze the model's ability to forecast (cumulative) mean returns over longer holding periods of five and ten trading days. We construct multi-period forecasts by iterating one-period

Table 2: Mean Return Forecasting Regressions: One and Two Weeks Ahead

	One Week				Two Weeks			
	1	2	3	4	5	6	7	8
ORB		0.542		0.587		0.402		0.446
		(7.7)		(8.7)		(8.4)		(9.8)
SVJ			0.022	0.005			0.043	0.027
			(7.6)	(1.2)			(9.8)	(4.4)
Money	0.009			0.001	0.012			0.025
	(1.0)			(0.1)	(1.0)			(1.7)
TTM	-0.009			-0.004	-0.014			-0.013
	(3.8)			(1.6)	(4.6)			(4.4)
Gamma	0.007			0.020	0.015			0.027
	(2.6)			(5.2)	(3.8)			(5.2)
Vega	0.013			0.009	0.023			0.020
	(4.0)			(2.4)	(5.1)			(3.8)
Theta	0.009			0.020	0.015			0.023
	(1.8)			(3.1)	(2.3)			(2.7)
IV	0.009			0.042	0.013			0.070
	(0.7)			(3.6)	(0.9)			(5.0)
Money*Put	-0.008			-0.005	-0.008			-0.026
	(0.8)			(0.4)	(0.6)			(1.5)
TTM*Put	0.007			0.003	0.009			0.007
	(2.4)			(1.0)	(2.9)			(1.8)
Gamma*Put	-0.003			-0.001	-0.005			0.005
	(1.0)			(0.2)	(1.9)			(1.4)
Vega*Put	-0.010			-0.002	-0.019			-0.007
	(4.6)			(0.8)	(6.3)			(2.0)
Theta*Put	-0.020			-0.022	-0.037			-0.031
	(3.9)			(3.8)	(5.2)			(3.6)
IV*Put	-0.005			-0.039	-0.001			-0.065
	(0.4)			(3.6)	(0.1)			(4.6)
R^2 (%)	0.2	5.6	0.4	6.3	0.5	5.8	0.7	6.9
N (1000s)	916.6	916.6	916.6	916.6	914.1	914.1	914.1	914.1

Note. Pooled forecasting regressions for delta-hedged returns to selling option contracts. Characteristic coefficients are reported as 100 times the characteristic's standard deviation. ORB and SVJ predictors are means of the models' simulated return distribution for each contract-day. Standard errors are clustered by day. All regression exclude date fixed effects.

forecasts as described in Section 3.2. In addition, when constructing multi-period returns and return forecasts, we implement a daily delta hedge. The delta hedge helps to purge returns of variation mechanically driven by the underlying spot return and focuses our analysis more narrowly on return variation unique to the options market.[16]

Table 2 reports mean forecast performance for multi-period delta-hedged options returns. One and two week bootstrap forecast performance is qualitatively the same as the performance of one day forecasts. Characteristic and SVJ forecasts improve somewhat over longer horizons with R^2s rising to 0.5% and 0.7%, respectively. Still the bootstrap forecast swamps these in predictive content, with a two week R^2 of 5.8%. The one week bootstrap slope coefficient of 0.54 is nearly identical to the one day slope. At two weeks, however, forecast accuracy begins to deteriorate to some extent as the slope drops to 0.40, but remains highly significant with a t-statistic of 8.4.

4.6 Volatility Forecasts

Next, we analyze our model's conditional forecast distributions by evaluating its accuracy in predicting return volatility. Our analysis mirrors that for mean forecasts. We adapt Mincer-Zarnowitz regressions to assess forecasts of return dispersion by replacing (6) with the regression

$$|r_{i,t+1}| = c_0 + c_1 \hat{E}_t |r_{i,t+1}| + e_{i,t+1},$$

where $\hat{E}_t |r_{i,t+1}| \equiv \frac{1}{B} \sum_b |\hat{r}^b_{i,t+1}|$.

Table 3 reports one day absolute return forecast regressions pooling all contract-days. Column 1 shows that contract characteristics strongly predict return volatility with an R^2 of 24.4%. In column 2, the bootstrap volatility forecast achieves an R^2 of 30.6% with a slope estimate of 0.69. The SVJ forecast in column 3 is also informative, but the magnitudes are substantially less with an R^2 of 4.5% and a slope of 0.04. Combining all predictors in column 4 produce a mild increase in R^2 to 31.9% from the bootstrap-only R^2 of 30.6%.[17]

At horizons of one and two weeks the pattern in volatility prediction is essentially the same. Characteristics and SVJ improve in predictive power, but the bootstrap model remains the single

[16]For a one day forecast, the delta hedge in the data and the forecast coincide as the one day Black-Scholes delta is known at time t. For forecasts beyond one day, we must forecast the delta hedge as well. We do so by calculating the Black-Scholes delta at the the bootstrap forecasted option price (taking into account the corresponding bootstrapped future spot price).

[17]In these regressions, the forecast target is not true volatility but a noisy proxy—the realized absolute return. Noise in the dependent variable mechanically depresses the forecast R^2, understating the volatility prediction power of all models reported in Table 3.

Table 3: Return Volatility Forecasting Regressions

	One Day				One Week				Two Weeks			
ORB		0.689		0.550		0.662		0.451		0.560		0.303
		(17.8)		(9.4)		(18.2)		(9.1)		(15.8)		(7.4)
SVJ			0.039	0.003			0.089	0.016			0.135	0.031
			(21.8)	(1.8)			(23.5)	(4.2)			(25.6)	(5.6)
Money	-0.064			-0.033	-0.137			-0.098	-0.224			-0.206
	(20.5)			(7.0)	(19.4)			(9.9)	(20.0)			(14.4)
TTM	0.013			0.003	0.019			0.009	0.031			0.021
	(15.6)			(2.8)	(11.5)			(4.9)	(12.5)			(8.7)
Gamma	-0.007			-0.006	-0.013			-0.013	-0.014			-0.019
	(6.5)			(4.1)	(6.6)			(4.9)	(4.9)			(4.8)
Vega	-0.010			-0.009	-0.018			-0.024	-0.031			-0.042
	(9.7)			(8.3)	(7.5)			(9.0)	(8.6)			(10.3)
Theta	-0.023			-0.011	-0.027			-0.016	-0.024			-0.016
	(10.4)			(3.6)	(7.9)			(3.4)	(5.1)			(2.5)
IV	0.039			0.006	0.091			0.025	0.125			0.046
	(11.7)			(2.0)	(14.5)			(4.1)	(14.5)			(5.3)
Money*Put	0.080			0.037	0.170			0.110	0.268			0.231
	(24.9)			(7.0)	(22.5)			(10.1)	(22.3)			(14.7)
TTM*Put	0.000			-0.005	0.005			-0.005	-0.002			-0.017
	(0.3)			(4.6)	(3.1)			(2.5)	(0.7)			(5.8)
Gamma*Put	0.004			0.003	0.001			0.000	-0.005			-0.011
	(3.6)			(1.8)	(0.6)			(0.2)	(2.4)			(3.4)
Vega*Put	0.001			0.002	0.000			0.003	0.005			0.006
	(0.5)			(1.9)	(0.2)			(1.3)	(2.2)			(2.1)
Theta*Put	0.009			0.010	-0.010			0.003	-0.022			-0.009
	(3.7)			(3.4)	(2.9)			(0.7)	(4.4)			(1.4)
IV*Put	-0.022			-0.003	-0.060			-0.016	-0.077			-0.021
	(6.4)			(0.8)	(9.4)			(2.2)	(9.2)			(2.2)
R^2 (%)	24.4	30.6	4.5	31.9	25.2	29.8	5.7	31.9	23.6	23.0	6.3	27.6
N (1000s)	919.3	919.3	919.3	919.3	916.6	916.6	916.6	916.6	914.1	914.1	914.1	914.1

Note. Pooled forecasting regressions for the absolute value of delta-hedged returns to selling option contracts. Characteristic coefficients are reported as 100 times the characteristic's standard deviation. ORB and SVJ predictors are means of the models' simulated absolute return distribution for each contract-day. Standard errors are clustered by day.

most powerful forecaster.

Table 4: One Day Conditional Quantile Forecasts

Target Quantile	One Day ORB	One Day SVJ	One Week ORB	One Week SVJ	Two Weeks ORB	Two Weeks SVJ
1	0.9	22.7	0.8	12.1	0.7	6.2
5	4.7	27.7	3.4	18.2	3.2	14.0
10	9.5	30.9	7.8	22.5	6.9	18.4
25	23.4	37.0	22.4	31.8	21.4	28.7
50	49.1	44.6	50.3	44.4	51.1	44.1
75	74.3	51.9	77.0	57.1	79.1	60.0
90	89.1	57.6	91.9	66.7	92.8	71.7
95	94.4	60.7	96.1	71.5	96.7	77.2
99	98.9	65.8	99.1	79.2	99.1	87.8

Note. Frequency with which forecasted quantile from each model exceeds corresponding data realization.

4.7 Quantile Forecasts

Lastly, we evaluate model forecasts in terms of their ability to predict quantiles throughout the distribution. This is our the most demanding assessment of forecast accuracy as it requires the model to precisely describe conditional probabilities of rare events.

For each option contract i on day t, the bootstrap and SVJ models generate a conditional distribution for the option price at time $t+h$. An accurate forecast distribution will see realized $t+h$ prices fall below its 1^{st} percentile in approximately 1% of observations. Likewise, the forecasted medians and 99^{th} percentiles will exceed roughly 50% and 99% of the matched realized prices, respectively. We therefore analyze quantile fits by describing the frequency with which the forecasted quantile exceeds the realized price. This frequency is equal to the cumulative empirical distribution of probability integral transforms, as described in Section 4.3.

In Table 4, we report the bootstrap model's out-of-sample fit of the 1, 5, 10, 25, 50, 75, 90, 95, and 99 percentiles of the future option price distribution. This is the tabular analogue of the CDF's in the second panel of Figure 4. The first column shows the targeted quantile (the CDF of the standard uniform distribution, or the 45-degree line in Figure 4) and remaining quantiles show the exceedence frequency for forecasted quantiles in each model. The first column reports one day ahead quantile fits from the out-of-sample bootstrap model. For example, bootstrap forecasts for the 1^{st} percentile exceed the realized price in 0.9% of the all contract-days in our sample. The

5^{th} and 10^{th} percentile forecasts exceed 4.7% and 9.5% of realizations, respectively. In short, the bootstrap model accurately describes the probability of low option price outcomes one day out. It is similarly accurate in the center of the distribution with fitted values of 23.4%, 49.1%, and 74.3% for the 25^{th}, 50^{th}, and 75^{th} percentiles. Finally, it accurately captures probability of rare events associated with large price realizations, achieving fitted values of 89.1%, 94.4%, and 98.9% for the 90^{th}, 95^{th}, and 99^{th} percentiles.

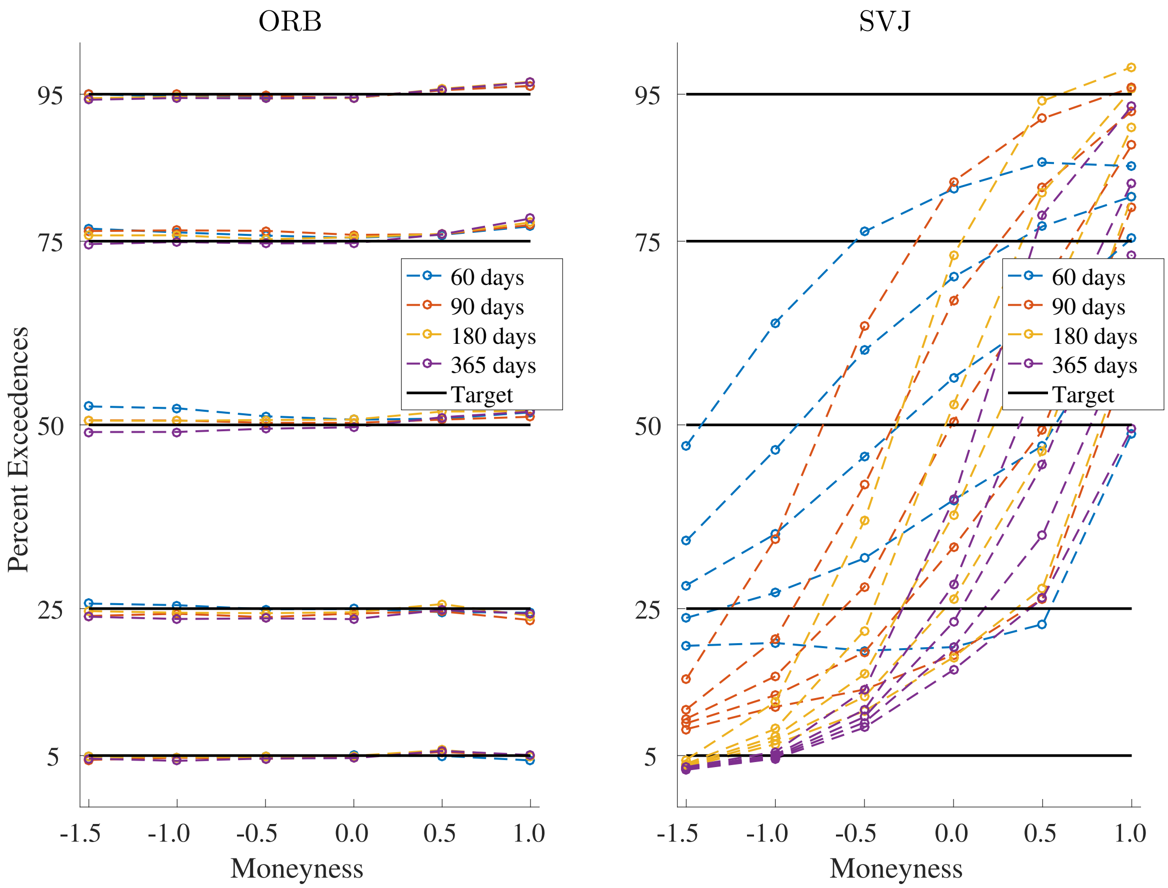
In the SVJ calibration of Broadie et al. (2009), realized one day ahead prices fall below the forecasted 1^{st} percentile 22.7% of the time, and realizations land above the SVJ 99^{th} percentile 34.2% ($1 - 65.8\%$) of the time. These indicate that the SVJ calibration predicts far too narrow of a conditional price distribution compared to that in the data.

Iterated bootstrap quantile forecasts deteriorate slightly at longer forecast horizons. For example, at two weeks the predicted 1^{st} and 10^{th} percentiles exceed realizations 0.7% and 6.9% of the time, indicating that long horizon model forecasts for downside price movements are somewhat more extreme than the data. On the other hand, SVJ forecasts of extreme quantiles improve with the forecast horizon. At two weeks, SVJ 1^{st} percentile forecasts exceeds 6.2% of the data and the SVJ 99^{th} exceeds 87.8% of the data. The SVJ median forecast fails to improve with horizon, with an exceedence frequency of 44% at both one day and two weeks.

The quantile fits in Table 4 pool all contract-days in our sample. In Figure 6, we investigate whether one day ahead quantile forecasts are especially successful or problematic in specific ranges of moneyness and maturity. We double sort observations into the 24 moneyness and maturity bins that we used for Figure 2, then calculate quantile fits within each bin. The left panel of Figure 6 reports results for the bootstrap model. The horizontal axis corresponds to bin moneyness, and line colors differentiate maturity. Some bins experience a mild deterioration in accuracy relative to the pooled sample: Median and upper quantile forecasts for OTM calls, and for median forecasts for short-dated OTM puts. But, as a whole, bootstrap quantile fits are remarkably accurate throughout the moneyness/maturity plane.

As a benchmark, the right panel of Figure 6 gives a full characterization of the distributional forecasting challenges faced by the SVJ model. SVJ forecasts are most accurate for short-dated,

Figure 6: One Day Conditional Quantile Forecasts by Moneyness/Maturity

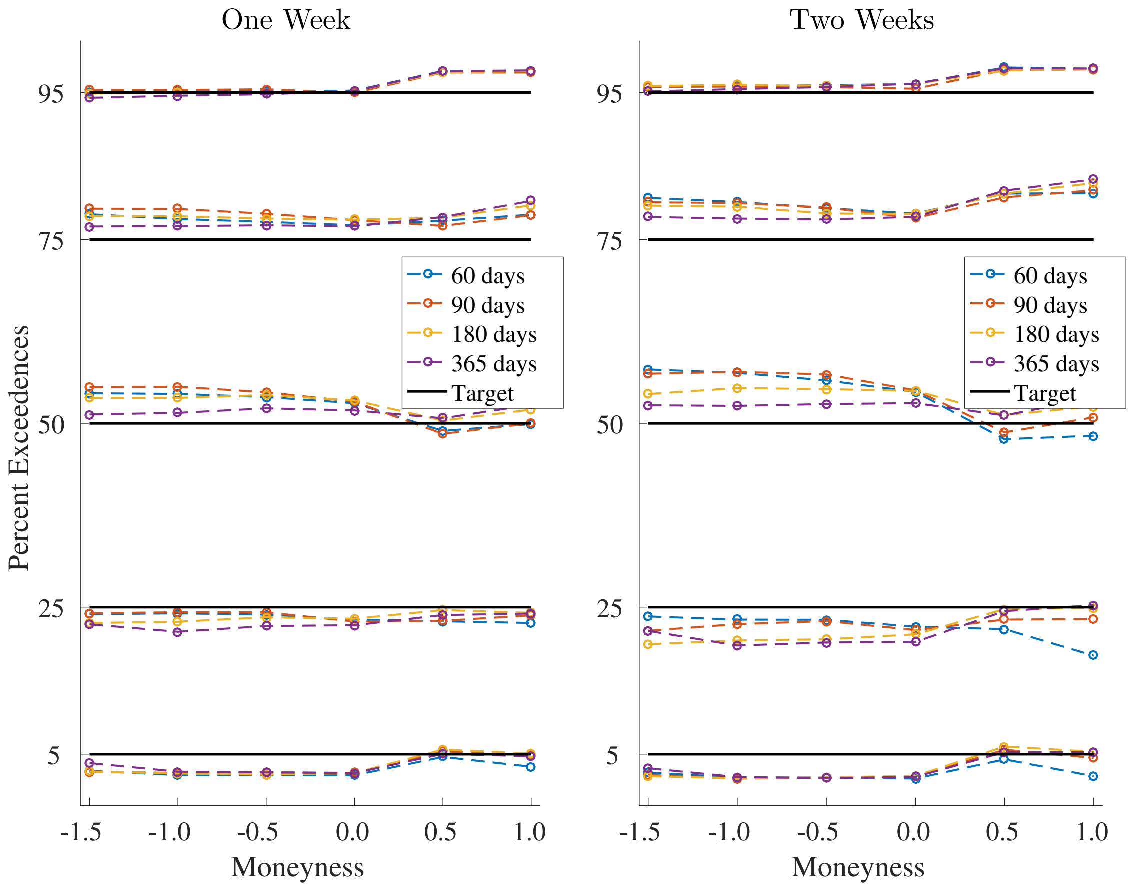


Note. Frequency with which ORB forecasted quantile exceeds corresponding data realization.

ATM options. At all maturities, the predicted distributions cannot generate realistic dispersion in future prices, and this problem worsens at higher maturities and away from the money. Furthermore, SVJ forecast distributions appear severely biased downward for OTM puts and biased upward for OTM calls.

Figure 7 shows the performance of bootstrap quantile forecasts at longer maturities. While our model continues to produce reasonably accurate quantile forecasts at one and two weeks, the figure illustrates that forecasts for extreme quantiles become overly aggressive when iterated

Figure 7: ORB Quantile Forecast Accuracy: One and Two Weeks Ahead



Note. Frequency with which ORB forecasted quantile exceeds corresponding data realization.

multiple periods ahead. There is also evidence of an upward bias in long horizon median forecasts, particularly for short-dated puts.

5 Economic Applications

In this section we study the usefulness of our bootstrap forecasting approach in financial and economic applications including risk management, measuring risk premia for state-contingent market

Table 5: Delta Hedge Regressions

	1	2	3	4	5	6
ORB	0.96				0.66	
	(116.9)				(7.1)	
ORB (IV control)		0.98				1.00
		(160.2)				(2.9)
SVJ			0.93		-0.07	-0.04
			(167.0)		(2.6)	(2.9)
Black-Scholes				0.98	0.39	0.02
				(165.0)	(4.0)	(0.1)
R^2 (%)	91.3	89.7	84.0	89.2	91.6	89.7
N (1000s)	919.3	919.3	919.3	919.3	919.3	919.3

Note. Pooled forecasting regressions of option contract price changes on model-based delta hedges. Standard errors are clustered by day.

exposures, and portfolio optimization.

5.1 Risk Management

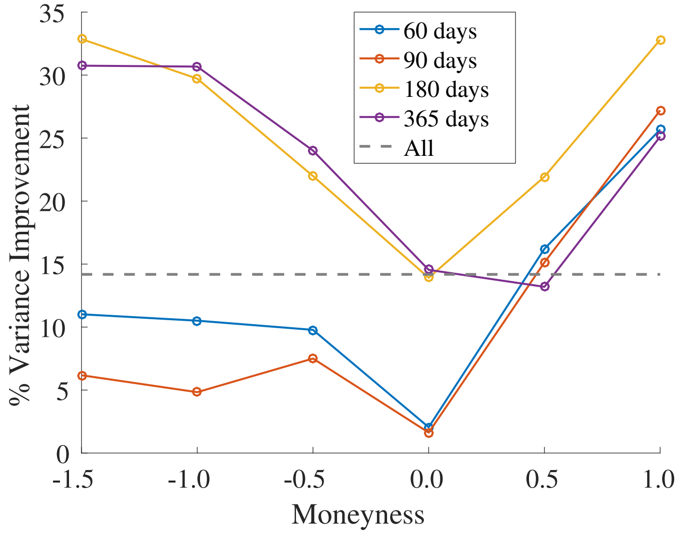
We study two risk management problems using bootstrap forecasts. The first is delta hedging an option position. For many option market participants, and for market makers in particular, delta hedging is the single most important risk management task. It allows the market maker to absolve its option inventory of its predominant source of risk—that associated with fluctuations in the underlying spot price.

The bootstrap model delivers a complete characterization of the joint conditional distribution of the underlying price with each contract price. Thus, the optimal least squares delta hedge is easy to construct via regression. Standing at date t and given a bootstrap sample of next period spot prices ($\hat{S}^b_{t+1}$) and option prices ($\hat{P}^b_{i,t+1}$), we estimate the hedge ratio $\Delta_{i,t}$ by regressing forecasted option prices on spot prices across bootstrap draws b:

$$\hat{P}^b_{i,t+1} - P_{i,t} = \Delta_{i,t}(\hat{S}^b_{t+1} - S_t) + \text{intercept} + \text{controls} + e^b_{i,t}.$$

This regression is estimated separately for each contract i on date t. In congruence with our earlier

Figure 8: ORB Delta Hedge Improvement Versus Black-Scholes



Note. Percentage variance reduction in delta-hedged option price changes by moneyness/maturity bin from using ORB delta versus Black-Scholes delta.

results, $\Delta_{i,t}$ is a purely out-of-sample estimate as the bootstrap draws for $t+1$ are constructed only using data and estimates through date t. The simplest bootstrap delta estimate omits control variables, but we also consider a version that accounts for bootstrapped changes in option implied volatility to potentially produce a purer delta hedge that is closer to the partial derivative represented by the Black-Scholes delta.

Next, to evaluate the performance of delta hedges, we regress realized option price changes on the model-based delta hedge, pooling all contract-days:

$$P_{i,t+1} - P_{i,t} = b_0 + b_1 \cdot \hat{\Delta}_{i,t}(S_{t+1} - S_t) + e_{i,t+1}.$$

We report delta hedge performance for the bootstrap, SVJ, and Black-Scholes models in Table 5. The results show that all models deliver effective hedges, with R^2 ranging from the low end of 84.0% for the SVJ model to a high of 91.3% for ORB. In multiple regressions, the bootstrap delta results in the most beneficial hedge. When the bootstrap delta is constructed controlling for changes in option IV, the ORB hedge subsumes the hedging efficacy of Black-Scholes and SVJ delta.

Table 6: Portfolio Value-at-Risk Forecasts

	Moneyness/Maturity Portfolios				Spread Portfolios	
Target Quantile	Short-dated OTM Put	Short-dated OTM Call	Long-dated OTM Put	Long-dated OTM Call	Term Spread	Risk Reversal
1	0.9	0.8	0.6	0.4	0.2	0.8
5	3.9	4.1	3.7	2.6	1.8	5.4
10	7.7	7.9	8.3	6.1	4.4	11.1
90	89.7	88.4	95.0	91.8	95.3	92.9
95	94.8	93.8	98.0	96.2	98.3	96.9
99	99.1	98.6	99.9	99.4	99.7	99.5

Note. Frequency with which ORB forecasted VaR exceeds corresponding data realization.

To quantify the delta hedging improvement from ORB in economic terms, we compare the variance in hedged option price changes between our model and the standard Black-Scholes hedge. We define the delta-hedged price change for contract i based on model $\mathcal{M}$ as

$$H_{i,t+1,\mathcal{M}} \equiv P_{i,t+1} - P_{i,t} - \Delta_{\mathcal{M}}(S_{t+1} - S_t).$$

Then, following Hull and White (2017), we define the hedging improvement from ORB relative to Black-Scholes as

$$100 \times \left(1 - \frac{Var(H_{i,t+1,ORB})}{Var(H_{i,t+1,BS})}\right).$$

Figure 8 summarizes the results. When variances are computed from all pooled observations, the out-of-sample variance improvement from hedging with ORB is 14.2% (gray dashed line).[18] The figure shows that Black-Scholes is most competitive for short-dated ATM puts, in which case the ORB variance improvement is 1.6%. The largest benefits of bootstrap hedges emerge for options that are long-dated and deep OTM, where the improvement is as large as 32%.[19]

The second risk management problem that we consider is constructing option portfolio value-

[18] For comparison, the variance improvement from ORB hedges versus SVJ is 45.8%.

[19] Our method is more general than the empirical minimum variance delta-hedges studied in Hull and White (2017). Like our procedure, they estimate deltas via regression. Their method first forms option portfolios, then estimates the empirical delta by regressing in the historical sample. Thus their delta estimates are not contract-specific and are not truly conditional. In contrast, we estimate regressions within a data set of simulated future data for the price of an individual contract and the underlying spot. In doing so, our deltas incorporate conditioning information based on the prevailing state of the factor vector and are tailored to the exact moneyness and maturity of the individual contract.

at-risk (VaR). The conditional quantile analysis of Section 4.7 demonstrates the efficacy of our bootstrap model for price distributions on a contract-by-contract basis. Successful portfolio VaR forecasts must describe quantiles in return space and demand that the model not only be successful in describing the tails of marginal return distributions, but must also accurately predict the joint tail of many contracts at once. Because the model generates a complete joint distribution for all contracts, combining bootstrap draws into portfolios allows one to produce the joint distribution for any set of portfolios of individual contracts.

We evaluate VaR forecasts from the ORB model by assessing the accuracy of conditional quantile forecasts of delta-hedged returns to six different option portfolios. The first four are equal-weighted portfolios that sell individual options in different moneyness and maturity bins. We consider short-dated (less than 60 days to maturity) and long-dated (180 to 365 days to maturity) contract bins comprised of either OTM puts (moneyness of -2 to -1) or OTM calls (moneyness of 0 to 1). From these four basic portfolios, we then construct two long-short spread portfolios. The first is a risk reversal portfolio that sells puts by investing one unit in each of the short-dated and long-dated OTM put portfolios and buys calls by investing negative one units of each OTM call portfolio. The second is a term spread portfolio (constructed by investing one unit in each of the short-dated portfolios and negative one unit of the long-dated portfolios).

Portfolio VaR results are reported in Table 6. Because the first four portfolios are selling options, low quantiles represent large losses to a short position, and high quantile describe extreme losses for a long position in the same options. For low quantiles, ORB forecasts are somewhat more accurate for short-dated options than long-dated. At high quantiles, VaR forecasts are very accurate with the exception of long-dated puts, where ORB forecasts tend to be more extreme than what we actually see in the data (thus they are conservative from the perspective of an options investor taking a long position).

VaR forecast exceedence frequencies for the long-short risk reversal strategy are also remarkably accurate at 0.8%, 5.4%, and 11.1% for the 1%, 5%, and 10% quantiles. For high risk reversal quantiles, ORB produces slightly more extreme outcomes than seen in the data. The weakest ORB performance is for the term spread portfolio, where we see overly conservative (too extreme) VaR forecasts for both high and low quantiles. The overall takeaway from the table is that the model produces accurate VaR forecasts for delta-hedged option portfolio returns, indicating that the ORB model capably describes tail dependence among returns to individual contracts.

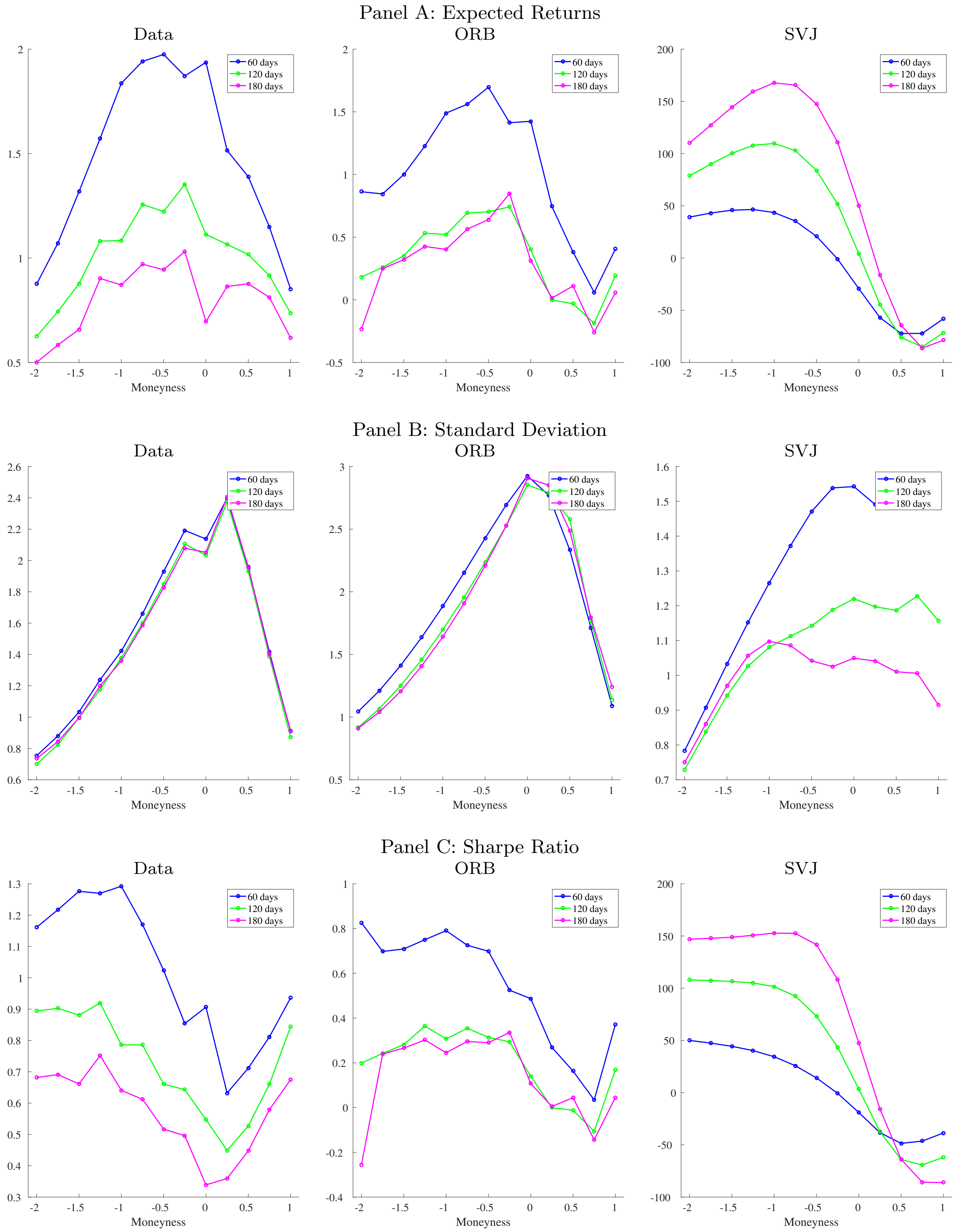
5.2 Risk and Return Surfaces

Options contracts are unique in providing distinct state-contingent payoffs to the aggregate equity claim. They are the among the real world traded securities that come closest to the theoretical state-price securities described by Arrow and Debreu (1954). By better understanding the risk compensation earned by these market-linked securities we improve our understanding of risk premia demanded by the typical investor. In this section, we use the bootstrap forecasting model to construct volatility, expected return, and Sharpe ratio "surfaces." We thereby describe risk compensation as a function of economic states described by option moneyness and maturity—that is, as a function of the future market value of the S&P 500 index at various investment horizons. Perhaps most interestingly, our model allows us to reconstruct risk and compensation surfaces day-by-day, conditional on prevailing economic conditions, which further describes how risk premia interact with state of the economy.

To provide an empirical baseline for understanding risk and return along the moneyness/maturity plane, we first study the unconditional returns of options portfolios. In particular, we form portfolios each day by interpolating realized returns of traded contracts to fixed grid points at maturities of 60, 120, and 180 days and moneyness of -2 to 1 at increments of 0.25. These are genuine portfolio returns, with individual contract weights determined by the distance between the contract and grid point. We then estimate unconditional expected returns, variances, and Sharpe ratios from the time series of portfolio returns at each grid point.

We plot annualized statistics for one day portfolio returns in the first column of Figure 9. Again, the returns we study are those accruing to an option seller and are hedged daily using Black-Scholes delta. In the data, unconditional annual Sharpe ratios are highest for short-dated options and gradually decline with maturity. Sharpe ratios are also higher for OTM versus ATM options, and are especially high for OTM puts. These estimates indicate that investors demand the

Figure 9: Unconditional Moments of One Day Returns



Note. Annualized means, volatilities, and Sharpe ratios for delta-hedge returns to selling options by moneyness/maturity bin. Data averages as well as ORB and SVJ model estimates are reported in left, center, and right columns, respectively.

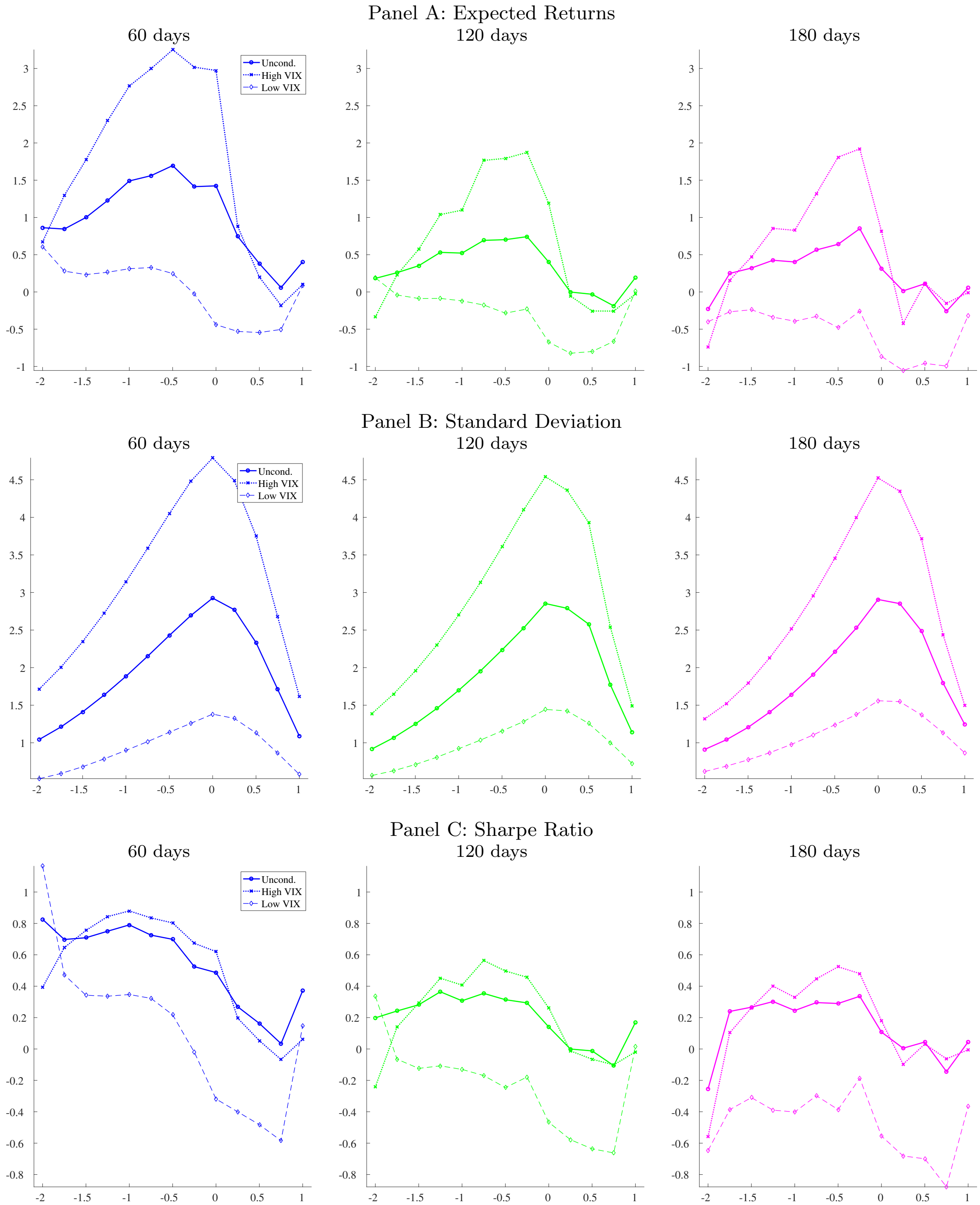
greatest compensation for selling insurance against states of the world with large market declines, and are particularly averse to large market declines occurring in the immediate rather than distant future.

Next, we construct corresponding estimates of unconditional option return moments based on our bootstrap model. To complement each and every realized option return in our sample, our model provides the corresponding (ex ante) bootstrap forecast distribution. Just as we interpolate realized returns to form time series portfolio returns at fixed grid points, we likewise construct forecasted mean and variances for grid point portfolios by interpolating the contract-level forecasted moments day-by-day. The forecasts on a given day correspond to the conditional portfolio return distribution. We estimate unconditional mean and variance by taking time series averages of their conditional values. From the unconditional moments, we compute the portfolio-level unconditional Sharpe ratio estimate from our model. The ORB model statistics are plotted in the second column of Figure 9.

The bootstrap model's estimates of unconditional expected returns, volatilities, and Sharpe ratios match the basic patterns and magnitudes in the data. Compensation per unit of risk is greatest for short-dated contracts, for OTM contracts, and especially for OTM puts. For comparison, column 3 of Figure 9 shows unconditional moment estimates for moneyness/maturity portfolios based on the SVJ model. The pattern in unconditional mean returns, volatility, and Sharpe ratios deviate drastically from the data. Some basic patterns, such as the term structure of Sharpe ratios, are reversed in the SVJ model. The magnitudes of SVJ model-based estimates differ drastically, by an order of magnitude in some places, compared to the data. While OTM puts earn annualized Sharpe ratios between 0.7 and 1.3 in the data, the same range from the SVJ model is 50 to 150.

Figure 9 shows that the bootstrap model specification captures the basic patterns in unconditional risk and return surfaces for the S&P 500. But power of our model is that it can characterize the (ex ante) *conditional* risk and return surface at each point in time. While one can arrive at a reasonable model-free estimate of unconditional option return surfaces by calculating sample moments of portfolio returns over the full time series (as in column 1 of Figure 9)—there is no such model-free analogue when it comes to describing conditional return surfaces. The semi-parametric

Figure 10: Conditional Moments of One Day Returns



Note. Annualized means, volatilities, and Sharpe ratios for delta-hedge returns to selling options by moneyness/maturity bin in the ORB model. Figures show unconditional moments along with conditional moments in high and low volatility regimes.

nature of the ORB model adds enough structure allow us to describe how conditional risks and risk compensation vary along the surface in different economic conditions. And the accuracy of the model's description (especially in comparison to a natural benchmark model like SVJ) lend confidence that the model-based conditional surfaces provide a realistic description of the true conditional return distribution.

Figure 10 draws bootstrap model-based surfaces for conditional expected return, volatility, and Sharpe ratio. We study two different conditioning sets—one representing unusually high volatility states and one for low volatility states. For high volatility states, we study days on which VIX exceeds its 75^{th} percentile of 24.4% (based on the 1996–2015 sample), and low volatility regimes are those when VIX is below its 25^{th} percentile of 15.1%. We estimate the conditional surfaces for these regimes by averaging the model's ex ante one day forecasts for mean and variance of returns, then converting these averages into a conditional Sharpe ratio. Rows report expected return, volatility, and Sharpe ratio, respectively, while columns break out surfaces by maturity.

Throughout the surface, option return volatilities rise by roughly 50% in high volatility states. This is due primarily to a rise in volatility-of-volatility and rise in the likelihood of large price moves (because options positions are delta-hedged). Commensurate with this rise in risk, expected put returns also rise sharply as investors demand additional compensation to provide insurance against market downturns. An interesting finding of the model is that Sharpe ratios conditional on a high risk environment differ little from unconditional Sharpe ratios. They tend to be slightly higher for OTM puts and, depending on maturity, are unchanged or slightly lower for OTM calls.

Conditional Sharpe ratios in low VIX conditions, on the other hand, fall sharply. Sharpe ratios of OTM puts fall by more than half, and become slightly negative for calls and ATM puts. Evidently, investors are especially keen to raise revenues by selling S&P 500 index insurance during low risk regimes, thus signifying their willingness to accept low or negative risk premia in such conditions.[20]

[20] In Appendix C, we report risk and return surfaces for investment horizons up to two weeks.

5.3 Portfolio Choice

By visualizing conditional risk, return, and Sharpe ratio surfaces for S&P 500 options, one can begin to infer the types of positions that a relatively risk tolerant investor might take to exploit differences in risk premia throughout the moneyness and maturity plane. The plots in Figures 9 and 10 suggest for example that, for an investor who is more risk tolerant than the market as a whole, the most attractive compensation comes from selling options on the short end of the maturity spectrum and in the moneyness region associated with severe market downturns (OTM puts). The full formulation of an optimized portfolio, however, relies also on the dependence among contract returns, encoded in their conditional joint distribution and described by the bootstrap model.

In this section, we solve the Markowitz (1952) portfolio choice problem faced by an option market investor given out-of-sample forecasts for the joint distribution of options returns from our bootstrap model. On each day, hundreds of contracts are traded making portfolio choice a high dimension optimization problem. While in principle our bootstrap method can handle even high dimension problems, we simplify the setting in order to make results for the optimal portfolio easier to read and interpret. In particular, before solving for the mean-variance optimal portfolio, we collapse the full set of contracts into a conditional joint distribution for the same four moneyness/maturity portfolios studied in Table 6. We solve for the maximum Sharpe ratio portfolio targeting 1% annualized return volatility.

Table 7 reports the post-formation performance of mean-variance optimized portfolios based on ORB, and represent returns to a strategy constructed on a genuinely out-of-sample basis. We report Sharpe ratio, mean, volatility, skewness and kurtosis of one day, one week, and two week strategies. We also report performance of nine benchmark strategies. The first solves the identical mean-variance portfolio choice problem but instead uses forecast distributions from the SVJ model. Next, we report the four moneyness/maturity portfolios that we use as base assets in the mean-variance problem. The last four benchmarks are common static option strategies. We consider a short-dated risk reversal (investing one unit in the short-dated OTM put portfolio and an opposing one unit position in the short-dated OTM call portfolio), the analogous long-dated risk reversal, and a term structure trade using either OTM puts or OTM call (selling short-dated options and

Table 7: Mean-Variance Optimized Portfolio Performance

	MVP		Moneyness/Maturity Portfolios				Spread Portfolios			
			Short-dated	Short-dated	Long-dated	Long-dated	Risk Reversal		Term Spread	
	ORB	SVJ	OTM Put	OTM Call	OTM Put	OTM Call	Short mat.	Long mat.	OTM Put	OTM Call
					Panel A: One day					
Sharpe ratio	7.12	0.87	1.30	0.75	0.69	0.55	0.01	-0.23	1.19	0.88
Mean	6.86	0.77	1.46	1.44	0.70	1.01	0.02	-0.36	0.83	1.34
Std. dev.	0.96	0.89	1.12	1.93	1.02	1.83	1.54	1.56	0.70	1.53
Skewness	0.46	1.72	-3.91	-1.85	-1.77	-0.84	0.54	0.36	-1.36	-1.50
Kurtosis	10.2	32.9	38.0	21.2	13.5	12.4	17.2	17.2	16.0	15.0
ORB Corr.	-	0.18	0.06	-0.04	0.04	-0.08	0.09	0.12	0.03	0.06
					Panel B: One week					
Sharpe ratio	2.27	0.26	1.22	0.73	0.62	0.53	0.25	-0.21	1.03	0.72
Mean	2.47	0.15	1.44	1.16	0.67	0.85	0.28	-0.21	0.82	1.16
Std. dev.	1.09	0.59	1.18	1.59	1.08	1.60	1.12	1.00	0.79	1.60
Skewness	-5.20	0.53	-4.63	-1.51	-1.98	-0.69	0.26	0.16	-0.79	-0.54
Kurtosis	92.2	11.5	42.7	12.1	14.7	8.1	11.6	11.7	12.4	7.2
ORB Corr.	-	0.16	0.24	0.02	0.16	0.00	0.22	0.18	0.13	0.09
					Panel C: Two weeks					
Sharpe ratio	1.39	0.36	1.09	0.77	0.63	0.53	0.20	-0.19	0.89	0.70
Mean	1.77	0.19	1.43	1.19	0.70	0.84	0.24	-0.18	0.77	1.12
Std. dev.	1.27	0.53	1.31	1.55	1.12	1.59	1.20	0.94	0.86	1.61
Skewness	-4.60	1.31	-5.05	-2.19	-2.25	-0.87	-0.35	0.52	-0.77	-0.51
Kurtosis	48.2	12.8	42.5	18.7	15.4	8.2	15.3	10.5	11.1	6.7
ORB Corr.	-	0.14	0.41	0.04	0.29	0.06	0.39	0.23	0.24	0.10

Note. Moments of option portfolio annualized percentage returns. Mean-variance optimized portfolios (MVP) use four moneyness/maturity option portfolios as base assets. ORB portfolios are based on out-of-sample forecasts.

Table 8: ORB Portfolio Weights

Moneyness Maturity	-2 to -1 20 to 60	0 to 1 20 to 60	-2 to -1 180 to 365	0 to 1 180 to 365
		Panel A: One Day		
Mean	-0.3	0.5	-0.4	0.3
75^{th} Pct.	4.8	4.2	2.4	3.0
25^{th} Pct.	-5.3	-3.5	-3.7	-2.1
		Panel B: One Week		
Mean	-2.0	0.9	-0.2	0.4
75^{th} Pct.	0.1	2.4	1.9	1.8
25^{th} Pct.	-4.0	-0.7	-2.6	-1.0
		Panel C: Two Weeks		
Mean	-2.1	0.7	-0.2	0.4
75^{th} Pct.	-0.7	1.6	1.2	1.3
25^{th} Pct.	-3.2	-0.1	-1.8	-0.6

Note. Time series averages and percentiles of ORB model optimized portfolio weights in percent using four moneyness/maturity option portfolios as base assets.

buying long-dated options with zero net investment).

Panel A shows results for one day returns. Among the four moneyness/maturity portfolios, selling short-dated OTM puts returns the highest Sharpe ratio (1.3), but takes on large negative skewness (-3.9) and high kurtosis (38.0). The static risk reversals do not appear profitable with Sharpe ratios near zero, but they eliminate negative skewness and reduce kurtosis in the short-dated case. The term structure trade produces an annualized Sharpe ratio of 1.2 with puts and 0.9 with calls and somewhat mitigates tail risk relative to naked sales of short-dated options.

The one day ORB optimized portfolio produces higher returns than any of these and with lower risk, with an out-of-sample Sharpe ratio of 7.1. Furthermore, the skewness of the bootstrap portfolio is 0.5, roughly the same as a static risk reversal and improving on skewness of -1.5 for static maturity trades. It also improves kurtosis to 10.2, versus approximately 15–17 for static spread trades. One day return performance is of course before transactions cost, which are known to be large in this market and likely to make trading on model forecasts over such a short horizon infeasible. Nonetheless, it is an apples-to-apples comparison to the benchmark portfolios which are also before transactions costs, quantifying ORB's relative forecasting performance in economic terms.

One and two week portfolio return performance is more representative of what an investor might feasibly achieve (gross of transaction costs). Panel B shows that the same basic patterns hold for one week returns. The annualized bootstrap Sharpe ratio is 2.3, which nearly doubles the Sharpe ratio of short-dated OTM put sales. For two week returns (Panel C), the Sharpe ratio gap narrows as the bootstrap delivers 1.4, versus 1.1 from the short-dated OTM put portfolio. Finally, the table also reports mean-variance optimized portfolio performance based on expected return and covariance inputs from the SVJ model. SVJ produces a portfolio that is typically of similarly low variance as the ORB portfolio, but it performs comparatively poorly in terms of expected returns and thus delivers unambiguously lower Sharpe ratios.

Table 8 shows the time series average, 25^{th} percentile, and 75^{th} percentile of the ORB optimized portfolio weights in the base assets at each investment horizon. For short investment horizons, weights are small on average but have a wide inter-quartile range, as expected returns estimated from the model are typically small but are highly variable (though reliably forecastable). Model-based expectations are comprised of estimated risk premia as well as any mispricings that the model predicts will correct. The combination of frequent changes in the signs of portfolio weights and an out-of-sample one day Sharpe ratio as high as 7.1 suggests short-lived option mispricings (that would be difficult to arbitrage in practice after accounting for transactions costs) make up a large fraction of the one day expected return estimate. In contrast, at horizons of one and two weeks, the weights on each moneyness/maturity bin take on a more consistent sign indicating that a durable risk premium is contributing expected returns over longer holding periods.

At one and two weeks, the max Sharpe ratio ORB portfolio shows substantial negative skewness and excess kurtosis. This fact can also be understood by looking at the typical portfolio weights that it selects. Consistent with the results in Figure 9, the strategy is generally a net seller of options, particularly selling puts and doing so especially aggressively on the short-dated end of the surface. In overweighting short-dated OTM puts, the ORB portfolio inherits the kind of tail risk found in OTM put portfolios.

6 Discussion and Conclusions

We present a statistical model for constructing forecast distributions for option returns. In the S&P 500 options market, our approach has a high degree of predictive accuracy for means, volatilities, and even extreme quantiles of option returns, especially relative to characteristic-based forecasts or forecasts from a traditional no-arbitrage model with stochastic volatility and price jumps. Our method demonstrates promising performance in economic applications such as risk management and portfolio choice. By drawing risk, return, and Sharpe ratio surfaces along the moneyness and maturity plane, we illustrate how our model can be used to understand the conditional risk prices that investors demand for bearing exposures to state-dependent payoffs in the aggregate equity market.

The great advantage of our option return model is its flexibility and ease of use. It requires four basic statistical techniques. The first is interpolation, which we use to build static synthetic options with constant moneyness and maturity. This step allows us to estimate the econometric model on assets with a fixed identity. We also use interpolation to translate forecasts at fixed surface grid points back to the actual contracts at coordinates off of the grid. Second, we use simple time series forecasting models. The backbone of ORB is a linear vector autoregression that captures the joint evolution in the S&P 500 spot price and the factors driving the IV surface. Third, we use GARCH to incorporate conditional volatility dynamics in factor innovations and surface residuals. The fourth technique we rely on is the bootstrap, which allows us to trace out forecast distributions for option prices without making specific parametric assumptions about shock distributions.

Each of these elements can be implemented in any basic statistical software. The main forecasting specification can be easily adjusted and reoptimized by adding or removing grid points to the interpolated surface, changing the number or identities of surface factors in the VAR, adopting essentially any variety of GARCH model for shock volatilities, or modifying the way that bootstrap samples are drawn. The computational burden of the model is minimal. Estimation of the model takes a matter of seconds, which means that the recursive out-of-sample forecasts we report take no more than a few minutes to build for our 20 year sample.

We use our method to draw surfaces for expected returns, volatility, and Sharpe ratios of options throughout the moneyness/maturity plane. In doing so, we provide new insight into the risk and risk premia of state-contingent claims to the aggregate equity market. Our ORB provides an accurate description of unconditional risk compensation, while the traditional no-arbitrage model that we analyze reverses basic patterns along the surface and is sometimes drastically different from the data in terms of Sharpe ratio magnitudes. More importantly, we are able to study *conditional* risk premium surfaces, which cannot be studied without the aid of a model's parametric structure. Our model appears uniquely well suited to describe conditional surfaces based on the evidence that ORB excels in forecasting future option return distributions.

Our approach has limitations as well. Because we rely on non-parametric interpolation of the IV surface, we place few restrictions on the shape of the surface. This flexibility is a virtue for data sets dense with observations throughout the moneyness and maturity plane, but becomes a curse in sparsely populated regions of the surface. Where contracts are few and far between, the interpolation may become inaccurate and any noise that this introduces will filter through model estimates and eventually manifest in the forecast distribution. Our decision to study contracts with moneyness between -2 and 1 and maturity between 20 and 365 days is driven precisely by this consideration. In the early part of the sample, contracts with less than one month to maturity have few traded strikes and longer maturity claims tend to be very widely spaced, thus we choose a moneyness/maturity rectangle that seems to balance breadth of the surface against contract density. An alternative modeling approach that takes a stand on a parametric functional form for the surface can potentially be successful in sparse data settings, but comes at the cost of potential misspecification in the parametric structure.

Our statistical model is a complement to models that strictly impose economic restrictions such as no-arbitrage. For example, by comparing differences in model-implied conditional moments, we can use our statistical framework to better detect and correct misspecifications of the economic restrictions. Similarly, one of the biggest difficulties facing no-arbitrage option models is accurately estimating the $\mathbb{P}$ distribution of the underlying. Incorporating information in option return moments like those we study can help estimate the $\mathbb{P}$ specification of no-arbritrage models by alleviating some of the difficulties of having to infer the $\mathbb{P}$ distribution for index returns alone.

References

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

Amengual, D. and Xiu, D. (2016). Resolution of policy uncertainty and sudden declines in volatility. *Working paper*.

Andersen, T. G., Fusari, N., and Todorov, V. (2015a). Parametric inference and dynamic state recovery from option panels. *Econometrica*, 83(3):1081–1145.

Andersen, T. G., Fusari, N., and Todorov, V. (2015b). The risk premia embedded in index options. *Journal of Financial Economics*, 117(3):558–584.

Arrow, K. J. and Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, pages 265–290.

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

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

Boyer, B. H. and Vorkink, K. (2014). Stock options as lotteries. *The Journal of Finance*, 69(4):1485–1527.

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

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

Cao, J. and Han, B. (2013). Cross section of option returns and idiosyncratic stock volatility. *Journal of Financial Economics*, 108(1):231–249.

Carr, P. and Wu, L. (2016). Analyzing volatility risk and risk premium in option contracts: A new theory. *Journal of Financial Economics*, 120(1):1–20.

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

Constantinides, G. M., Jackwerth, J. C., and Perrakis, S. (2009). Mispricing of s&p 500 index options. *The Review of Financial Studies*, 22(3):1247–1277.

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

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

Daglish, T., Hull, J., and Suo, W. (2007). Volatility surfaces: Theory, rules of thumb, and empirical evidence. *Quantitative Finance*, 7(5):507–524.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, pages 863–883.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376.

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

Fengler, M. R., Härdle, W. K., and Mammen, E. (2007). A semiparametric factor model for implied volatility surface dynamics. *Journal of Financial Econometrics*, 5(2):189–218.

Fengler, M. R., Härdle, W. K., and Villa, C. (2003). The dynamics of implied volatilities: A common principal components approach. *Review of Derivatives Research*, 6(3):179–202.

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

Frazzini, A. and Pedersen, L. H. (2012). Embedded leverage. Technical report, National Bureau of Economic Research.

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

Gatheral, J. (2004). A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. *Presentation at Global Derivatives & Risk Management, Madrid*.

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

Goyal, A. and Saretto, A. (2009). Cross-section of option returns and volatility. *Journal of Financial Economics*, 94(2):310–326.

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

Hu, G. and Jacobs, K. (2016). Volatility and expected option returns. *Working paper*.

Hull, J. and White, A. (2017). Optimal delta hedging for options. *Journal of Banking & Finance*.

Israelov, R. and Tummala, H. (2017). Which index options should you sell. *Working paper*.

Karakaya, M. M. (2014). *Characteristics and expected returns in individual equity options*. University of Chicago, Division of the Social Sciences, Department of Economics.

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

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1-2):125–144.

Mincer, J. A. and Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance*, pages 3–46. NBER.

Ofek, E. and Richardson, M. (2003). Dotcom mania: The rise and fall of internet stock prices. *The Journal of Finance*, 58(3):1113–1137.

Ofek, E., Richardson, M., and Whitelaw, R. F. (2004). Limited arbitrage and short sales restrictions: Evidence from the options markets. *Journal of Financial Economics*, 74(2):305–342.

Vasquez, A. (2016). Equity volatility term structures and the cross-section of option returns. *The Journal of Financial and Quantitative Analysis*.

Appendix to "Forecasting the Distribution of Options Returns"

A SVJ Model Calibration

The SVJ model that we study is from Broadie et al. (2009). The price and stochastic volatility processes under the $\mathbb{P}$ measure are

$$\begin{aligned} dS_t &= (r + \mu - \delta) S_t dt + S_t \sqrt{V_t} dW_t^s(\mathbb{P}) + J_t(\lambda^{\mathbb{P}}, \mu_J^{\mathbb{P}}, \sigma_J^{\mathbb{P}}) \\ dV_t &= \kappa(\theta^{\mathbb{P}} - V_t) dt + \sigma_v \sqrt{V_t} dW_t^v(\mathbb{P}) \end{aligned} \quad \text{(A1)}$$

where $W_t^s(\mathbb{P})$ and $W_t^v(\mathbb{P})$ are Brownian motions with correlation ζ (denoted ρ in their paper, but not to be confused with ρ in our Equation 2). The risk-free rate, equity risk premium, and dividend yield are r, μ, and δ. The term $J_t(\lambda^{\mathbb{P}}, \mu_J^{\mathbb{P}}, \sigma_J^{\mathbb{P}})$ represents a normal mixture price jump with intensity, mean jump size, and jump size volatility of $\lambda^{\mathbb{P}}, \mu_J^{\mathbb{P}}$, and $\sigma_J^{\mathbb{P}}$.

The model under the $\mathbb{Q}$ measure is identical with the exception that all parameters having $\mathbb{P}$ superscript in Equation (A1) are allowed to take different values under $\mathbb{Q}$. Our SVJ simulations use the specific $\mathbb{P}$ parameter values reported in Table 1 and $\mathbb{Q}$ parameters reported in Table 7 (row 1) of their paper.

Table A1: Estimated ρ Matrix

	Index	Log VIX	Put 1	Put 2	Call 1	Call 2
Index Return	0	0	0	0	0	0
VIX	0	0.983	0.009	0.015	0.002	-0.010
Put Factor 1	0	0.055	0.937	-0.142	0.130	0.116
Put Factor 2	0	0.009	-0.026	0.856	0.205	0.178
Call Factor 1	0	0.007	0.040	0.285	0.463	-0.520
Call Factor 2	0	0.002	0.000	0.032	-0.092	0.829

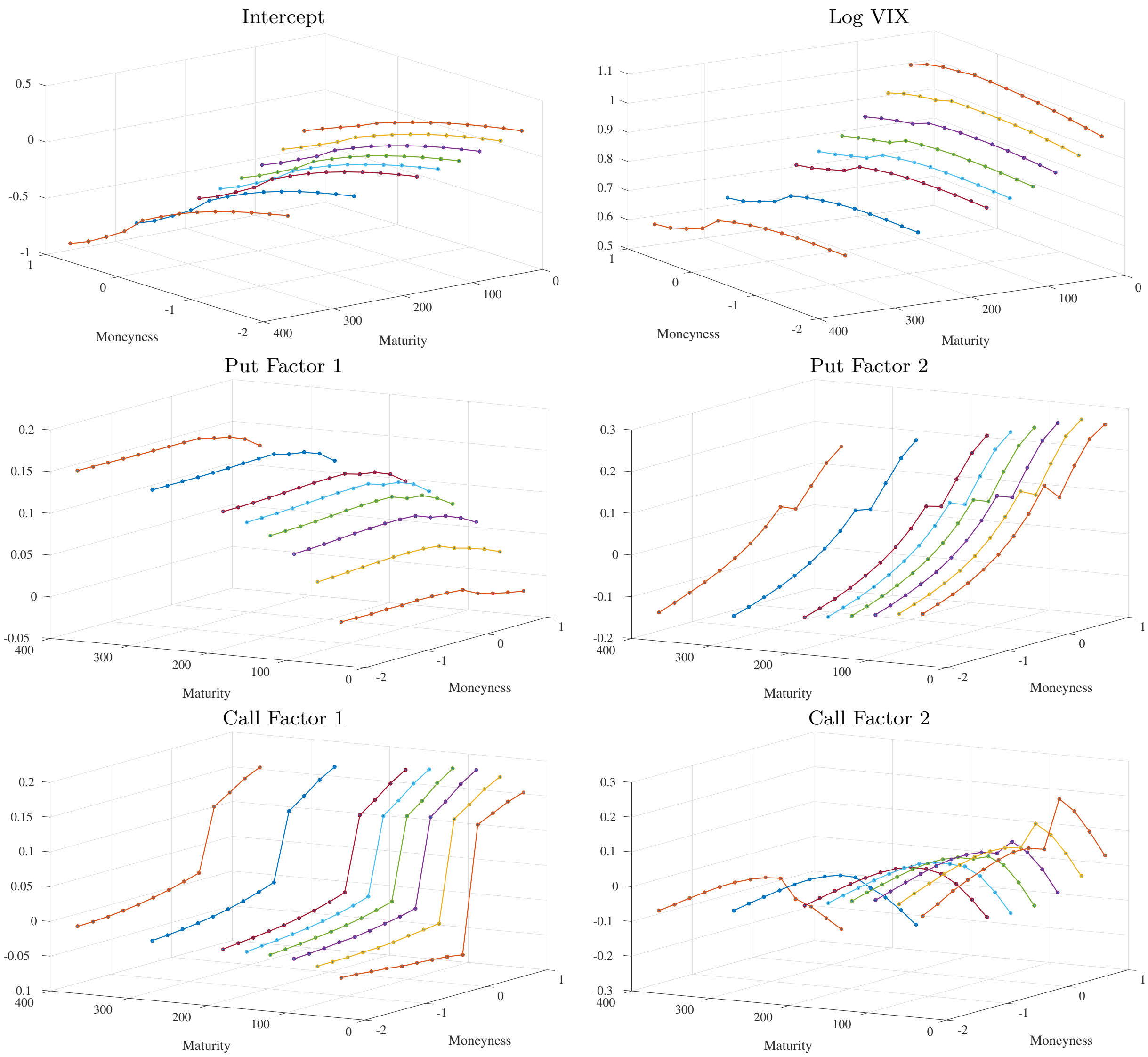
Note. Estimated full sample coefficient matrix for factor vector autoregression.

B Model Parameter Estimates

This appendix describes estimates of model parameters and processes. We focus on the model with five surface factors—the log VIX plus two put surface PCs and two call surface PCs, and we report full sample estimates. Table A1 reports the estimated ρ matrix for the factor vector autoregression. To help reduce parameterization, we impose zero coefficients on the dynamic association of the index return with surface factors. This restriction does not play a large role in our analysis, as the estimated coefficients are close to zero and insignificant when the restriction is lifted.

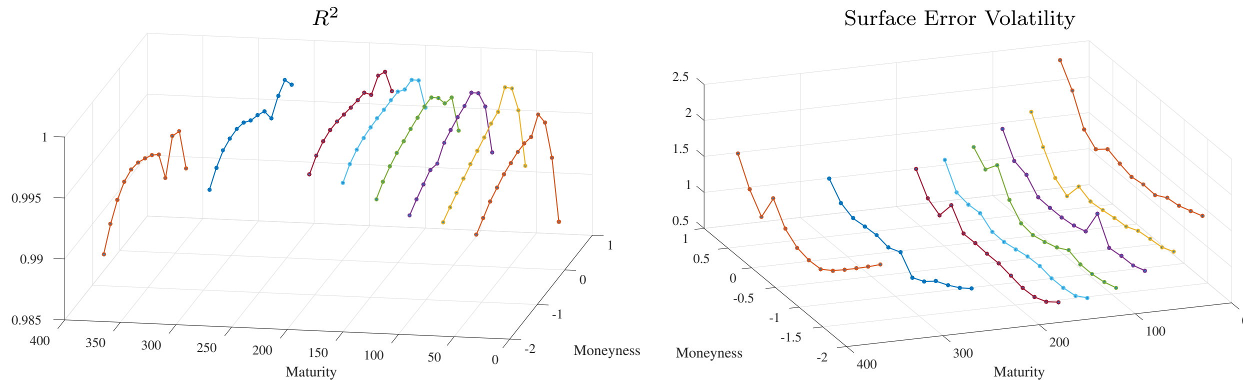
Figure A1 reports factor loading estimates at each grid point. Figure A2 reports the R^2 of the surface factor model at each grid point as well as the time series average GARCH volatility of surface errors. Figure A3 plots the daily time series of surface factors (excluding the VIX). Figure A4 plots the daily time series of GARCH volatilities for each surface factor.

Figure A1: Factor Loading Estimates by Grid Point



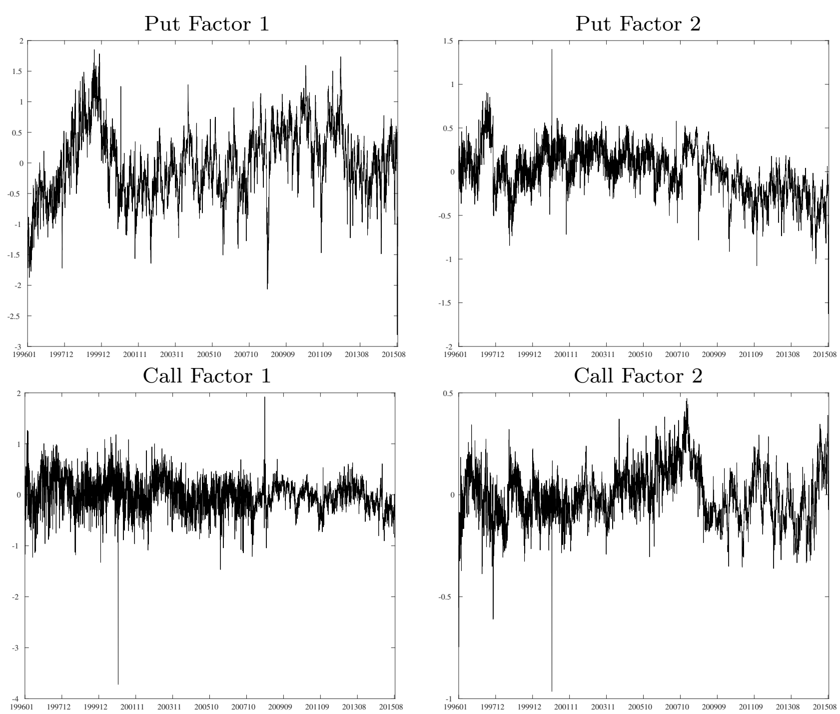
Note. Full sample factor loadings by grid point.

Figure A2: Model R^2 and Surface Error Volatility by Grid Point



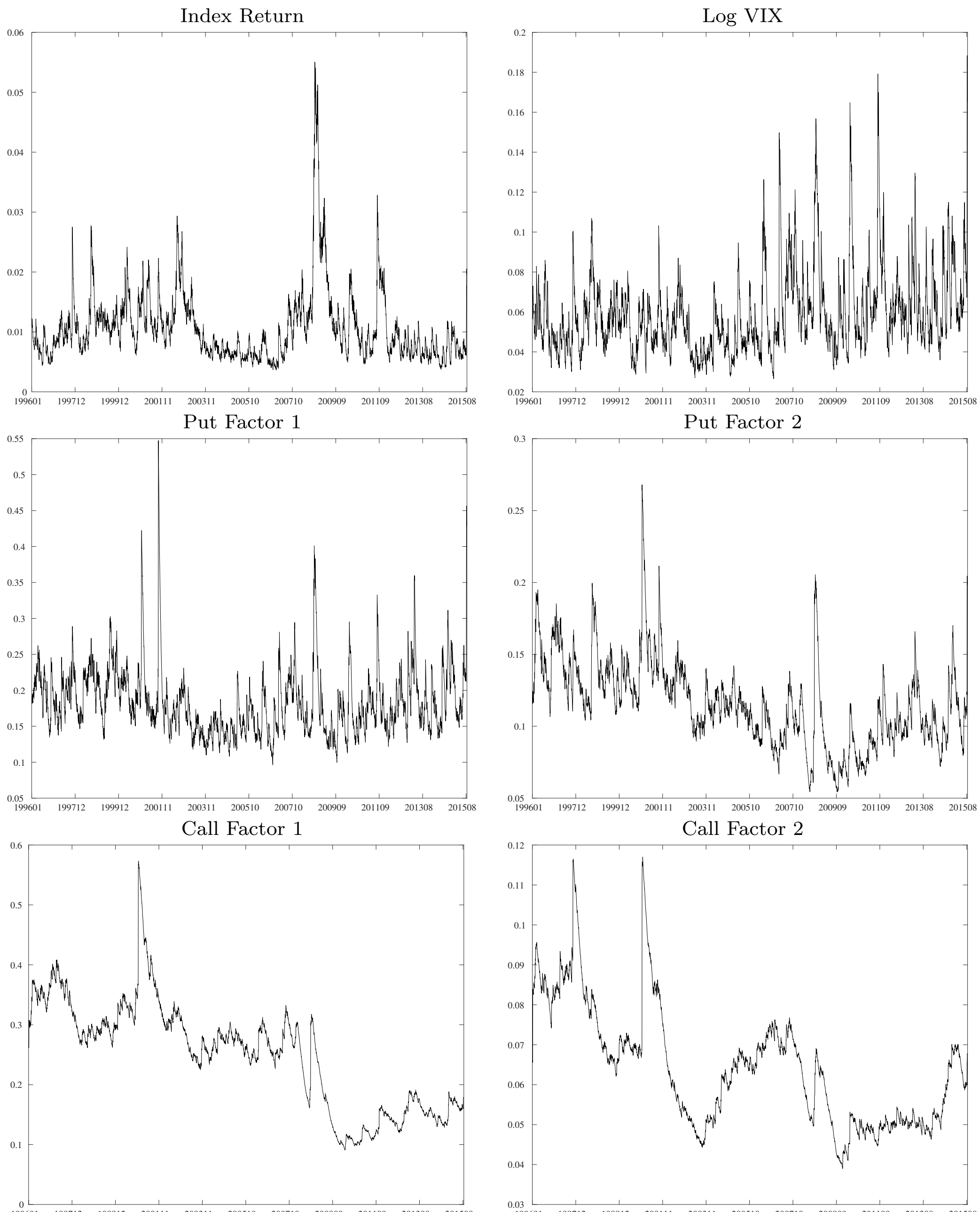
Note. Surface factor model R^2 and residual volatility by grid point.

Figure A3: Surface Factor Time Series



Note. Time series estimates of surface factors (excluding log VIX).

Figure A4: Surface Factor Conditional Volatility Time Series

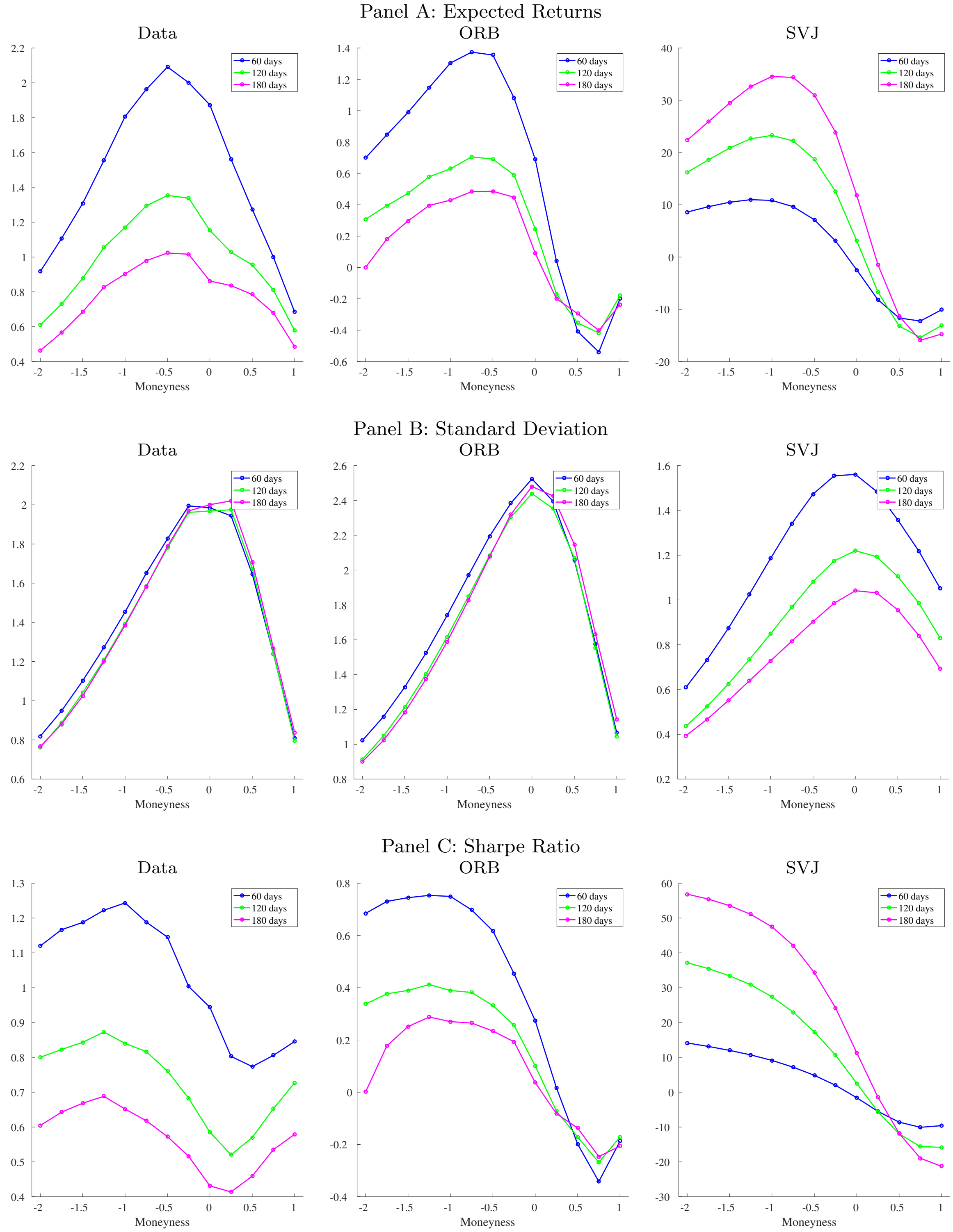


Note. Estimated conditional volatility time series for factor vector innovations.

C Return Surfaces at Longer Horizons

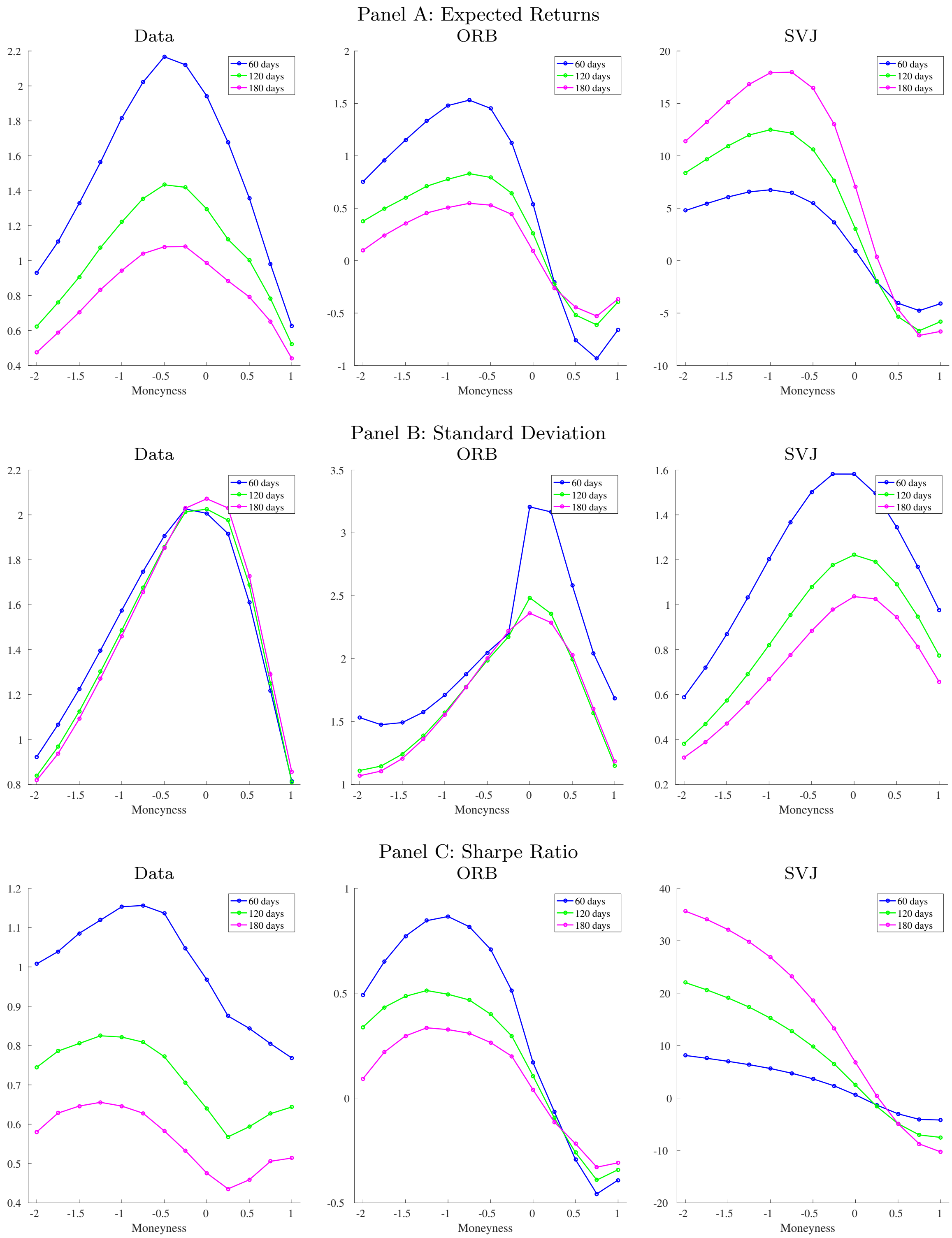
This appendix extends the the analysis of Section 5.2 to longer investment horizons of one and two weeks. Figures A5 and A6 report unconditional surfaces from the data along with the ORB and SVJ models, and Figures A7 and A8 show conditional surfaces in high and low volatility regimes from the ORB model.

Figure A5: Unconditional Moments of One Week Returns



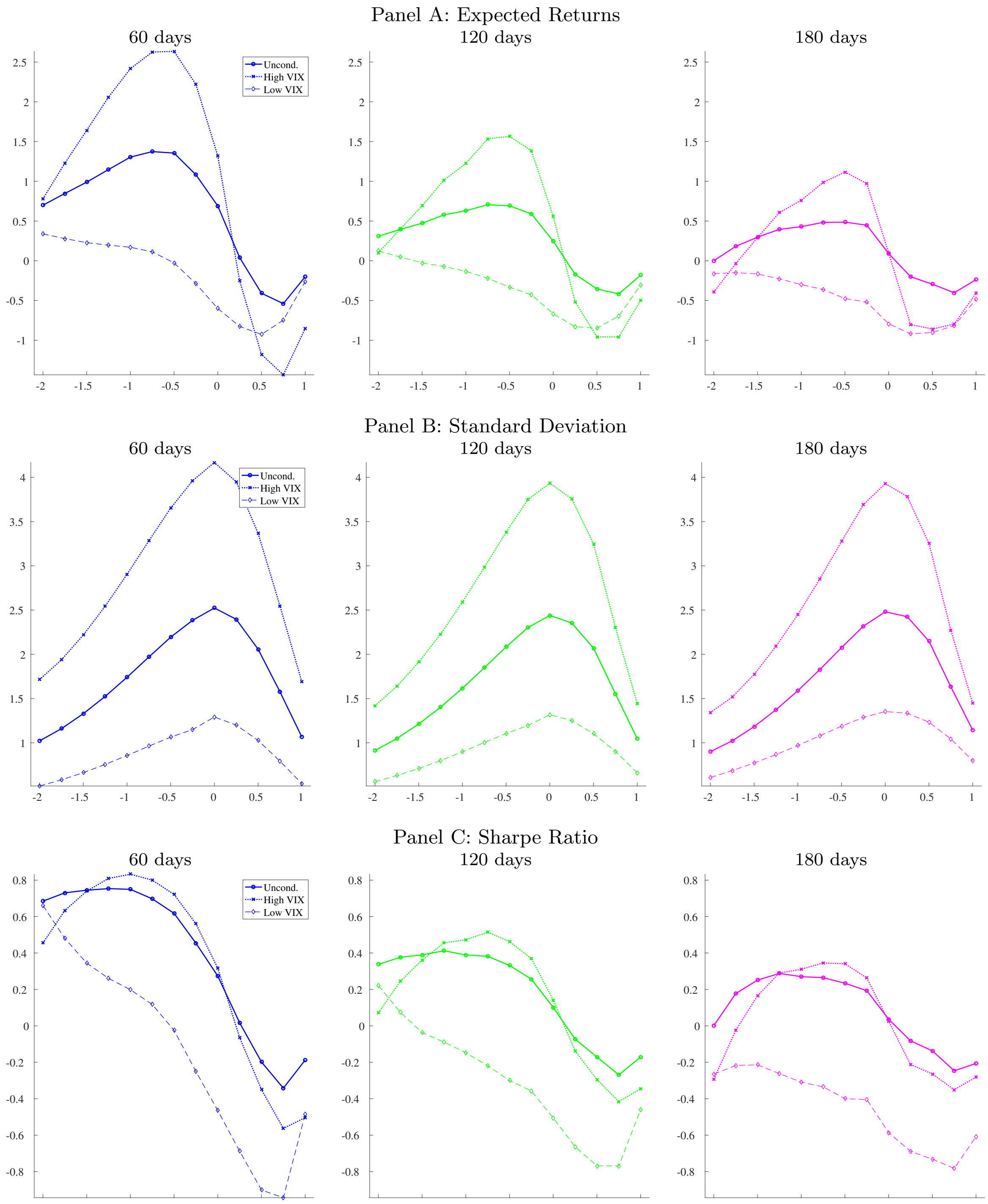
Note. Annualized means, volatilities, and Sharpe ratios for delta-hedge returns to selling options by moneyness/maturity bin. Data averages as well as ORB and SVJ model estimates are reported in left, center, and right columns, respectively.

Figure A6: Unconditional Moments of Two Week Returns



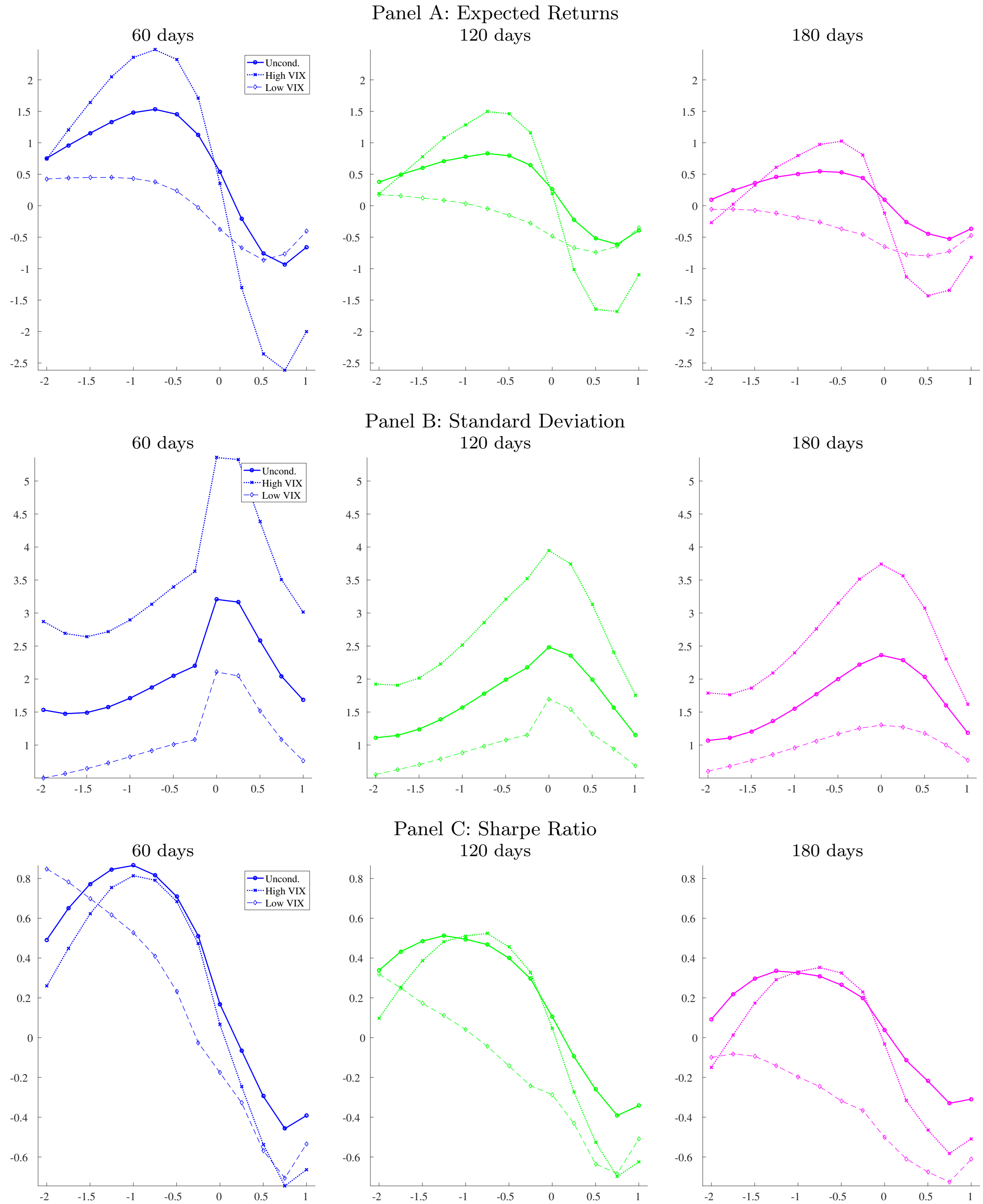
Note. Annualized means, volatilities, and Sharpe ratios for delta-hedge returns to selling options by moneyness/maturity bin. Data averages as well as ORB and SVJ model estimates are reported in left, center, and right columns, respectively.

Figure A7: Conditional Moments of One Week Returns



Note. Annualized means, volatilities, and Sharpe ratios for delta-hedge returns to selling options by moneyness/maturity bin in the ORB model. Figures show unconditional moments along with conditional moments in high and low volatility regimes.

Figure A8: Conditional Moments of Two Week Returns



Note. Annualized means, volatilities, and Sharpe ratios for delta-hedge returns to selling options by moneyness/maturity bin in the ORB model. Figures show unconditional moments along with conditional moments in high and low volatility regimes.