The decline of the variance risk premium: evidence from traded and synthetic options

Ian Dew-Becker and Stefano Giglio*

September 1, 2024

Abstract

Equity index options historically displayed sharply negative returns and CAPM alphas. This could reflect investor risk preferences or intermediary frictions. We document that over the past 15 years, option alphas have become indistinguishable from zero. We also introduce synthetic options, that, under some conditions, reflect risk preferences of the average equity investor, independent of option-market frictions. Synthetic options never, over the last 100 years, had negative alpha, indicating that equity investors never required high compensation for market downturns. An intermediary-based model explains the patterns in both synthetic and traded options, including the recent decline in the variance risk premium.

1 Introduction

Background

*Dew-Becker: Chicago Booth and NBER. Giglio: Yale University and NBER. We appreciate helpful comments from Gurdip Bakshi, Mike Chernov, Itamar Drechsler, Bjorn Eraker, Niels Gormsen, and seminar participants at ITAM, Yale, the University of Connecticut, Northwestern, Johns Hopkins, the University of Miami, INSEAD, HEC Paris, Berkeley, the Canadian Derivatives Institute, Boston University, the SFS Cavalcade, Booth, HBS, SAIF, Kansas University, CFIRM, Rice, SITE, FIU, the NBER, MIT, and Georgetown.

A major empirical fact in financial markets is that equity index options have been overpriced historically relative to simple benchmark models. Investors who purchase options have, on average, earned significant negative returns and negative CAPM alphas.[1] Bates (2022) discusses two classes of explanations for that fact. First, marginal utility for the representative investor might be convex in market returns. Periods with large negative returns have state prices that are higher than would be expected if marginal utility were linear in the market return as in the CAPM. That convexity can be due, for example, to aversion to crashes, aversion to high volatility, time-varying risk aversion, or behavioral factors.[2]

But option prices also have features that are difficult to reconcile with standard utility theory, for example often implying *negative* risk aversion in certain states. A second class of explanations focuses on intermediaries, explaining option overpricing as the result of intermediaries being net short options and charging a premium for their concentrated risk.[3] In that case, option prices reveal the preferences and constraints of the specialist investors that trade in options markets, and *not* necessarily those of the average equity investor.

Understanding which of the two explanations is correct is important because, in addition to being intrinsically interesting, option prices are often used to measure many features of financial markets, including investor expectations of various moments of the conditional

[1]For analyses of returns, see Coval and Shumway (2001), Bakshi and Kapadia (2003), Broadie, Chernov, and Johannes (2007), Constantinides, Jackwerth and Savov (2013), Chambers, Foy, Liebner, and Lu (2014), Dew-Becker et al. (2017), and Muravyev and Ni (2020), among many others. For structural models, see Backus, Chernov, and Martin (2011), Drechsler and Yaron (2011), Gabaix (2012), Drechsler (2013) , Seo and Wachter (2019), and Schreindorfer (2020). Note, though, that those models are almost exclusively endowment economies.

[2]See Cuesdeanu and Jackwerth (2018) for an extensive review of such models and their empirical support, along with a discussion of models of intermediary frictions. On risk aversion, see Ait-Sahalia and Lo (2000), Jackwerth (2000), Rosenberg and Engle (2002), and Schreindorfer and Sichert (2022). On the variance risk premium, see, in addition to work cited elsewhere here, see Lamoreaux and Lastrapes (1993), Coval and Shumway (2001), Bakshi and Kapadia (2003), Du (2011), Dew-Becker et al. (2017), and Ait-Sahalia, Karaman, and Mancini (2020).

[3]See Jackwerth (2000), Bollen and Whaley (2004), Bates (2008), Han (2008), Garleanu, Pedersen, and Poteshman (2008), Jurek and Stafford (2015) , Haddad and Muir (2021), Frazzini and Pedersen (2022), among many others. A related literature has also explored the link between option markets, returns in various asset classes, and the role of risk taking by intermediaries: for example, Brunnermeier, Nagel and Pedersen (2008), Bao, Pan and Wang (2011), Longstaff et al. (2011), Nagel (2012), and Chen, Joslin and Ni (2019).

distribution of returns, investor preferences across market return states, and the drivers of risk premia.[4] They can also reveal potential amplification mechanisms for macroeconomic shocks,[5] and are a key input in understanding the importance of stabilization policy, as optimal policy depends on agents' subjective valuations of different possible states of the world.[6]

Contribution

This paper has theoretical and empirical components. Theoretically, it develops a novel approach to measuring the average investor's risk preferences by studying *synthetic* options – dynamic portfolios that attempt to replicate returns on traded options by dynamically trading the underlying. While that result is totally general, we also illustrate it in a general equilibrium model. Empirically, the paper measures returns on synthetic options over nearly a century of data, compares them to traded option returns, and examines how both have changed over time, along with the frictions that might drive that variation.

Methods and results

The paper first provides a simple theoretical framework to ground the interpretation of option returns – traded and synthetic. The theory gives conditions under which the CAPM alpha of a traded option on the stock market measures curvature in marginal utility with respect to the market return.[7] The CAPM appears as a benchmark here not because marginal utility might be "truly" linear in any sense, but rather to take an empiricist's approach of starting with a linear null and then measuring the degree of nonlinearity.

[4]On preferences over market states, see Ait-Sahalia and Lo (2000), Jackwerth (2000), and Rosenberg and Engle (2002). For conditional moments, see the CBOE VIX index and, among many others, Carr and Madan (1998), Carr and Wu (2009), and Martin (2017). Options have also been used to measure jump risk (e.g. Bollerslev and Todorov (2014)), micro uncertainty (Dew-Becker and Giglio (2020)), and option implied skewness (Kozhan, Neuberger, and Schneider (2013), Dew-Becker (2022). Bollerslev and Todorov (2011) and Beason and Schreindorfer (2022) use option prices to measure the drivers of risk premia.

[5]E.g. He and Krishnamurthy (2013), Hall (2017) , and Muir (2017).

[6]Alvarez and Jermann (2004) is an example of how asset prices can be used to measure the cost of fluctuations. De Paoli and Zabczyk (2013) study optimal policy under time-varying risk aversion.

[7]Importantly, this is a reduced-form statement. Option prices do not measure a *structural* effect of the market return on marginal utility. They measure the average value of marginal utility conditional on a return state.

Since options have payoffs that are convex in the market return, when marginal utility is also convex options have relatively high prices and consequently low returns – negative CAPM alphas. The connection between option alpha and curvature of marginal utility relies on integration between equity and option markets, and is violated when frictions induce segmentation across the two markets.

The paper also provides circumstances under which convexity in marginal utility also implies negative CAPM alphas for *synthetic* options. While the required conditions are relatively strong, the paper provides some empirical evidence in their favor and also gives bounds on the potential magnitude of the bias due to their violation.

Empirically, the paper then examines monthly returns on synthetic and traded options. The synthetic options are constructed back to 1926 using data on the CRSP market return, while monthly traded option returns are available since August, 1987. Empirically, replication works quite well: synthetic options have returns that are over 90 percent correlated with traded option returns and, most importantly, hedge *all realized crashes over the last century* effectively. That does not mean that options could have been synthesized in real time historically, though – trading costs and other frictions would have made that infeasible (and in fact replication frictions are central to our intermediary-based explanation of the results).

The key question, and one of the paper's central results, is what alphas the synthetic options earn. Whereas traded options have strongly negative CAPM alphas, synthetic options have historical alphas that are indistinguishable from zero, with confidence bands that are economically narrow. In the benchmark full-sample results, the lower bound for the confidence bands for the information ratios is -0.2. That result is robust over time, across strikes, across maturities, and to modifying various details in the construction.[8]

[8]That result is consistent with the option pricing literature – the large negative return on delta-hedged returns going back to Bakshi and Kapadia (2003) is equivalent to finding that traded options have more negative returns than synthetic options. The novelty in this dimension is in pointing out that the small alpha for synthetic options extends back to 1926.

On the other hand, the alphas of traded options have *not* been so consistent (see also Bates (2022)). The paper's second key empirical result is that according to various estimation methods, there is a break in the returns somewhere around 2010. In the period since 2010, in fact, *the alphas of the traded options have converged to zero*, consistent with the synthetic options. Since the gap between traded and synthetic option returns is literally a delta-hedged return, another way to state this second result is that the alpha of delta-hedged options has gone to zero. Relatedly, the paper also shows that the CAPM alpha of the variance risk premium has shrunk towards zero.

Note that this paper is not novel for building a replication strategy – every analysis using delta hedging also does so (e.g. Bakshi and Kapadia (2003)). It is novel because it focuses on the delta hedge *by itself,* which allows us to extend the analysis to the beginning of the 20th century (long before traded options become available), and because it documents large variation in the alphas of traded (but not synthetic) options over the recent decades.

The final section of the paper asks what might have caused such a shift. We first develop a general equilibrium model with heterogeneous investors that can simultaneously explain the empirical patterns. It shows that a decline in trading frictions can explain the decline in option overpricing. Intuitively, the idea behind the model is that with investor heterogeneity, when retail investors are unable to sell options – the model's core friction – the equilibrium price of options will be driven by the investors with the greatest demand. But as the frictions decline, overpricing will also, because the investors willing to supply options become free to do so.

The model also delivers additional testable implications. The most important is that the option premium should be related to the amount of risk borne in equilibrium by dealers (similar to Bates (2022) and Garleanu, Pedersen, and Poteshman (2008)). We show that the net S&P 500 gamma exposure of dealers and market makers for Cboe options shifted from being consistently negative to being zero or positive following the financial crisis. Other factors driving hedging costs, including trading frictions and basis risk, also declined, which, according to the model, also contributes to the decline in the traded option alpha.

Broader implications

The paper's basic findings have two additional implications beyond what has been discussed so far. First, the paper's results imply that derivatives prices, up until relatively recently, were distorted away from those implied by the preferences of whoever is the typical investor pricing equities. When estimating and testing a representative-agent model, the results imply that fitting the behavior of syntehtic options is more appropriate than fitting the behavior of traded options. And that can be done not only for the overall equity market – synthetic options can be created on any underlying.

Second, the analysis finds clear evidence of nonstationarity – both in option returns and in positions and other measures of frictions. So when studying traded option returns, attention must be paid to the exact sample being used and how the results may have changed over time. Synthetic option returns, on the other hand, have stable characteristics.

Outline

The remainder of the paper is organized as follows. Section 2 compares traded and synthetic options and discusses the assumptions needed to interpret their CAPM alphas in terms of investor risk preferences. Section 3 shows the empirical results on the time series patterns of traded and synthetic options and their implications for the shape of marginal utility. Section 4 focuses on the decline in alpha of traded options in the last decade and presents a general equilibrium model that can explain the empirical patterns we document, and provides additional testable implications. Section 5 concludes.

2 Traded and synthetic options: theoretical framework

This section introduces the two types of options, and explores under which assumptions average returns on traded and synthetic options can provide a measure of curvature in marginal utility with respect to the market return.

2.1 Definitions and notation

The market return between periods t and $t+j$ is $R^m_{t,t+j}$. The change in marginal utility is $M_{t,t+j}$ (i.e. u'_{t+j}/u'_t, where u is utility over consumption).[9] Since $M_{t,t+j}$ is a ratio of marginal utilities, we immediately have that $M_{t,t+2} = M_{t,t+1}M_{t+1,t+2}$, etc. Since any deviations of investors' subjective probability measure from the truth can also be incorporated into $M_{t,t+j}$, we refer to it more generally as subjective marginal utility, or *SMU*.

Definition 1 *A return $R_{t,t+j}$ is priced by subjective marginal utility, denoted $M_{t,t+j}$, over $t \to t+j$ if*

$$1 = E_t\left[M_{t,t+j}R_{t,t+j}\right] \tag{1}$$

where E_t is the expectation operator under the physical probability measure conditional on information available on date t.

Note that this definition may potentially only hold for certain t and j – i.e. only on some dates or over just some horizons. It does not imply that markets are complete, so marginal utility need not be identical across agents. They only must agree on equation (1) for whichever assets are priced by marginal utility. In addition, it does not require rational expectations – irrational beliefs can be accommodated by $M_{t,t+j}$ as long as they satisfy basic axioms for probability measures.

For simplicity, the theoretical analysis takes the risk-free rate to be zero (equivalently, all returns can be interpreted as on forward contracts). That implies that $E_t M_{t,t+j} = 1$ for all t and j. The analysis is straightforward to recapitulate in the case where the risk-free rate is nonzero, and the empirical analysis accounts for nonzero interest rates.

The paper's goal is to understand how marginal utility varies with the state of the equity market. To that end, define, for an arbitrary variable $X_{t,t+j}$, the nonlinear projection on the

[9] If agents have biased probability measures, then the bias will also be part of $M_{t,t+j}$. We would in that case just require that there are internally consistent in that $M_{t,t+2} = M_{t,t+1}M_{t+1,t+2}$.

market return,

$$\bar{X}_{t,t+j} \equiv E\left[X_{t,t+j} \mid R^m_{t,t+j}\right], \tag{2}$$

$$\text{and } \hat{X}_{t,t+j} = X_{t,t+j} - \bar{X}_{t,t+j} \tag{3}$$

$\bar{X}_{t,t+j}$ is the component of $X_{t,t+j}$ that can be written as a function of the market return and $\hat{X}_{t,t+j}$ is the residual. We say $\bar{X}_{t,t+j}$ is the (nonlinearly) **spanned** part and $\hat{X}_{t,t+j}$ the **unspanned** part.[10]

$\bar{M}_{t,t+j}$ is the paper's primary object of interest – how marginal utility varies with the market return (again, not causally, just in a conditional expectation sense). It is what the past literature on option-implied pricing kernels has focused on, since, as the next section shows, it is what options carry information about.

2.2 Interpreting traded option returns

Define the gross return on some arbitrary option (or derivative) on the market to be $R^O_{t,t+j}$. The part of that return (linearly) correlated with the market can always be subtracted, and we have

$$R^{O\perp}_{t,t+j} \equiv R^O_{t,t+j} - \frac{\text{cov}_t\left(R^O_{t,t+j}, R^m_{t,t+j}\right)}{\text{var}_t\left(R^O_{t,t+j}, R^m_{t,t+j}\right)}\left(R^m_{t,t+j} - 1\right) \tag{4}$$

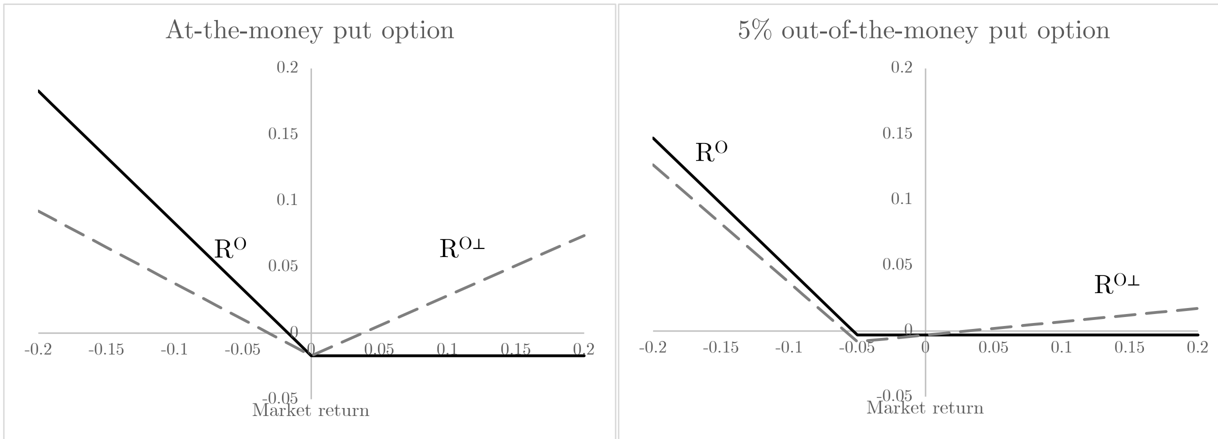
$$\alpha^O_{t,t+j} = E_t\left[R^{O\perp}_{t,t+j} - 1\right] \tag{5}$$

where $\alpha^O_{t,t+j}$ is the CAPM alpha of $R^O_{t,t+j}$. It will become clear in a moment why we create these objects. First, though, note that $R^{O\perp}_{t,t+j}$ is not a dynamically or delta-hedged return; rather, one might say it is beta-hedged: the hedge is a *fixed* position in the underlying, conditioned on date-t information. Since it just adds a static position in the market, $R^{O\perp}_{t,t+j}$ has the usual kinked relationship with $R^m_{t,t+j}$, just tilted compared to $R^O_{t,t+j}$. Figure 1 gives

[10]Throughout the paper, the term "span" is used in the Hilbert space sense of a conditional expectation.

example for an out-of-the-money and at-the-money put option.

Figure 1: Hypothetical option payoffs



Note: Hypothetical net payoffs $R^{O}_{t,t+j}$ and $R^{O\perp}_{t,t+j}$ for two different strikes.

Proposition 2 *If $R^{O}_{t,t+j}$ and $R^{m}_{t,t+j}$ are priced over $t \to t+j$, then $\bar{M}_{t,t+j}$ has the representation,*

$$\bar{M}_{t,t+j} = const. - \frac{E_t\left[R^m_{t,t+j} - 1\right]}{\mathrm{var}_t\left[R^m_{t,t+j}\right]} R^m_{t,t+j} - \frac{\alpha^O_{t,t+j}}{\mathrm{var}_t\left(R^{O\perp}_{t,t+j}\right)} R^{O\perp}_{t,t+j} + resid. \tag{6}$$

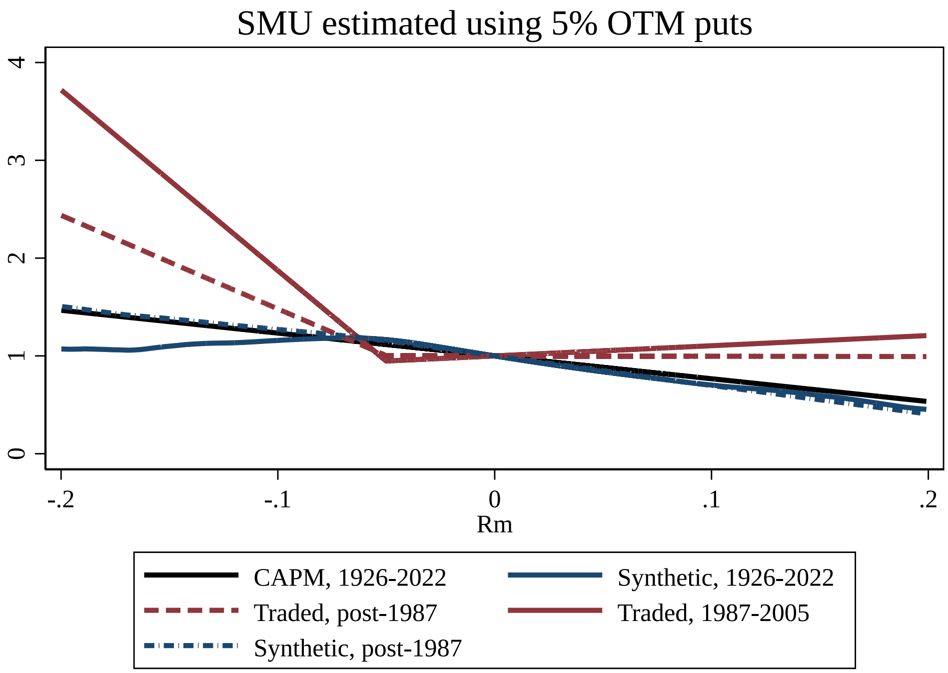
where the residual term is orthogonal to $R^m_{t,t+j}$ and $R^{O\perp}_{t,t+j}$.

The alpha of an option measures nonlinearity in marginal utility relative to the market return. Specifically, equation (6) is a regression of $\bar{M}_{t,t+j}$ on two functions of the market return: a linear term ($R^m_{t,t+1}$) and a nonlinear term ($R^{O\perp}_{t,t+j}$, which is, conditionally, an exact nonlinear function of the market return). The result comes from the fact that the covariances in the numerators of the regression coefficients can be replaced here by the two risk premia – e.g. $E_t\left[R^m_{t,t+j} - 1\right] = -\mathrm{cov}_t\left(R^m_{t,t+j}, M_{t,t+j}\right)$.

$\frac{-E_t\left[R^m_{t,t+j}-1\right]}{\mathrm{var}_t\left[R^m_{t,t+j}\right]}$ therefore measures the average slope of marginal utility with respect to the market return. $R^{O\perp}_{t,t+j}$ is a piecewise linear function of the market return, so its coefficient, $\frac{-\alpha^O_{t,t+j}}{\mathrm{var}_t\left(R^{O\perp}_{t,t+j}\right)}$, measures how the slope of $\bar{M}_{t,t+j}$ changes across the strike.

Figure 2 illustrates that idea, plotting SMU, normalized to equal 1 for $R^m_{t,t+j} = 1$, relative to the market return. Under the CAPM (the black line), SMU is linear in the market return, all alphas are zero, there is no convexity, and the slope is recovered simply as $\frac{-E_t\left[R^m_{t,t+j}-1\right]}{\text{var}_t\left[R^m_{t,t+j}\right]}$.

Figure 2: SMU estimated using exchange-traded and synthetic puts



Note: The figure shows estimated SMU under different models and estimated in different samples. The solid black line reports the estimated SMU as a function of the market alone (as in the CAPM). The other lines model SMU as a function of the market and the orthogonalized returns on traded and synthetic options in various samples.

The dashed red line plots the SMU implied by the alphas observed for 5% out-of-the-money listed S&P 500 puts between 1987 and 2022. Historical put returns imply that effective risk aversion – as measured by the slope of SMU – is significantly higher when the market falls. The non-monotonicity here is a typical, if surprising, empirical finding.

2.3 Interpreting synthetic option returns

It is well known that option returns can be approximated through dynamic trading in the underlying asset – the market return in this case. Do the approximated returns then also approximately measure nonlinearity in $\bar{M}_{t,t+j}$?

By *synthetic* option return, we simply mean a return that is replicated via dynamic trading in the underlying. Denote the weight on the underlying each day by δ_t^S (which depends only on information up to date t, ensuring feasibility, at least if trading is frictionless). In the Black–Scholes (1973) replication, for example, δ_t^S is exactly the delta of the option being replicated, which depends on the level of the market index and its volatility.

The return on the synthetic option from t to $t+j$ is then

$$R_{t,t+j}^S \equiv \sum_{s=t}^{t+j-1} \delta_s^S \left(R_{s,s+1}^m - 1\right) + 1 \tag{7}$$

Synthetic options will earn a CAPM alpha if they productively time the market – i.e. if δ_s^S is correlated with variation in the market risk premium (see section 3.3.3).

Note that in general $R_{t,t+j}^S \neq R_{t,t+j}^O$ and the replication of an option will not be perfect. Nevertheless, we have the following result:

Proposition 3 *If $R_{t,t+1}^m$ is priced by SMU for all $s \to s+1$ for $t \leq s < t+j$, then*

$$\bar{M}_{t,t+j} = const. - \frac{E_t\left[R_{t,t+j}^m - 1\right]}{\mathrm{var}_t\left[R_{t,t+j}^m\right]} R_{t,t+j}^m - \frac{\left(\alpha_{t,t+j}^S + \mathrm{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}_{t,t+j}^S\right)\right)}{\mathrm{var}_t\left(R_{t,t+j}^{S\perp}\right)} \bar{R}_{t,t+j}^{S\perp} + \ resid. \tag{8}$$

where $\alpha_{t,t+j}^S$ is the CAPM alpha of $R_{t,t+j}^S$ and the residual is orthogonal to $R_{t,t+j}^m$ and $\bar{R}_{t,t+j}^{S\perp}$.[11]

There is again an expression for SMU in terms of two returns, with coefficients depending on their risk premia. Proposition 3 gives three conditions under which $\alpha_{t,t+j}^S$ can be used to measure convexity in SMU:

1. $\bar{R}_{t,t+j}^S$ is a convex function of the market return
2. $R_{t,t+1}^m$ is priced by SMU for all $t \to t+1$
3. $\mathrm{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}_{t,t+j}^S\right)$ is zero

[11] Recall that overbars denote components spanned by the market return and hats the unspanned components.

The first condition just says that synthetic options have returns that are convex in the market, which can be checked empirically. The second and third conditions are harder to evaluate because they are statements about SMU, which is not directly observable. The pricing condition might fail at the daily level if there are frictions allowing market prices to deviate from their fundamental values at high frequency (see section 3.6.2).

The third condition requires that the unspanned part of synthetic option returns not be priced. Section 3.4 discusses it extensively, both looking at what variables $\hat{R}^{S}_{t,t+j}$ is correlated with and also using the method of Cochrane and Saa-Requejo (2000) to bound the covariance.[12]

Those three conditions are the key point of the general theoretical analysis. They show what is required in order to use synthetic options to measure nonlinearity in marginal utility.

2.3.1 Feasibility

A surprising feature of the three conditions for interpreting synthetic option returns is that none of them directly requires that options synthesis be feasible in practice. The requirement that the market is priced correctly every day, for example, does not say that every investor can trade every day. In many models, the market is correctly priced even though there is no trade in equilibrium (this also holds in the general equilibrium model of section 4.2). There is no question that for the vast majority of investors over the vast majority of the empirical sample, replicating options dynamically would have been expensive and time-consuming. And in fact frictions associated with the replication are ultimately central to the paper's theoretical explanation. But frictions also prevent many investors from freely trading

[12]Condition 3 holds in any model where $R^{m}_{t,t+j}$ is a sufficient statistic for SMU (so that $\hat{M}_{t,t+j} = 0$); for example if marginal utility is a function of current wealth and that wealth returns are perfectly correlated with stock market returns (which can hold under Epstein–Zin preferences). If the exact path of the market return matters or if there are other state variables that affect marginal utility and are independent of market returns, they will appear in $\hat{M}_{t,t+j}$. Condition 3 also holds when nonlinear payoffs on the market can be replicated via dynamic trading, e.g. when the market follows a binomial tree, so that $\hat{R}^{S}_{t,t+j} = 0$. In reality neither of those conditions holds literally, and the question becomes how large the bias from $\operatorname{cov}\left(\hat{M}_{t,t+j}, \hat{R}^{S}_{t,t+j}\right)$ is.

options, which violates the assumptions needed to use traded options to infer marginal utility.

What is important in the end is not whether agents can trade continuously or perfectly replicate options. Rather, the question is whether the price of the stock market – or that of traded options – is close to "fair" – in the sense of definition 1. The paper will return to this issue throughout the analysis.

3 Empirical analysis: traded and synthetic option returns

This section documents the properties of returns of traded and synthetic options over time, as well as their difference. It also derives the SMU that are implied by the CAPM alphas of the two instruments.

3.1 Data

Option synthesis Throughout the analysis, t is taken to be a day. The weight δ_t^S is equal to the delta of the option – the partial derivative of the value with respect to the price of the underlying. Different models give different exact expressions for delta. The main analysis uses a method from Hull and White (2017) that corrects the Black–Scholes delta for the leverage effect, but this is largely immaterial – the unadjusted Black–Scholes delta delivers similar results (see section 3.6).

Section 3.6.2 discusses potential effects of market microstructure biases. To avoid the effects of stale prices (see Bates (2012)), δ_t^S is constructed based only on information lagged by a day in the main analysis.

The return volatility needed to calculate delta is obtained from a heterogeneous autoregressive model (Corsi (2009)) that forecasts 1-month volatility as a function of past two-week volatility and the past three months of volatility (with the lags chosen based on the Bayesian information criterion). The model is estimated on an expanding window, so that when the delta is computed only past information is used. Robustness to the various choices here is examined in section 3.6.

The market return is measured as the (daily) CRSP value-weighted stock market return and the risk-free rate is the one-month Treasury bill rate from Kenneth French's website.

Traded options The dataset for traded options splices together CME futures options for the period 1987–1995 with CBOE SPX options from Optionmetrics for 1996–2022. Following Broadie, Chernov, and Johannes (2009), we study a monthly rolling strategy, where options are purchased on the third Friday of every month and then held to their maturity on the following month's third Friday.

In parts of the analysis that involve direct comparisons of synthetic and traded options, we align the returns – comparing returns over the same third-Friday-to-third-Friday period. However, when looking at univariate statistics, the analysis uses 21-day overlapping windows for the synthetic options to maximize statistical power (since there is no need to only use a single return per month).

For both the synthetic and traded options, excess returns are scaled with the price of the underlying in the denominator, rather than the price of the option, as in Büchner and Kelly (2022). The scaling is the return perceived by an investor who is buying options in proportion to the underlying. It is a payoff per unit of insurance, rather than per unit of the insurance premium that is paid. See appendix B for further discussion.

3.2 The relationship between $R^S_{t,t+j}$ and $R^m_{t,t+j}$

The top panel of Figure 3 plots $R^S_{t,t+j}$ for put options against $R^m_{t,t+j}$, where $j = 21$ days and the strike used to construct δ^S_t is 95% of the initial level of the market (corresponding to approximately a unit standard deviation decline). The plot for call options with the same strike is identical but rotated 45 degrees counterclockwise.

There is clearly significant nonlinearity – for values of the market return above the strike the slope is near zero, while for values below it the slope is approximately -1, consistent with the fact that $R^S_{t,t+j}$ is constructed to mimic a put option. The red line plots the nonparametric estimates of $\bar{R}^S_{t,t+j} \equiv E\left[R^S_{t,t+j} \mid R^m_{t,t+j}\right]$.[13] They formally quantify the relevant nonlinearity, showing that $\bar{R}^S_{t,t+j}$ is close to piecewise linear in the market return.

Importantly, $\bar{R}^S_{t,t+j}$ rises with a consistent slope as $R^m_{t,t+j}$ falls, regardless of how large the decline is. If it was not possible to span large declines in the market with time-varying weights, e.g. due to large jumps, $\bar{R}^S_{t,t+j}$ would flatten out for the very negative values of $R^m_{t,t+j}$. Figure A.1 replicates figure 3 across strikes and shows the results are highly similar.

The table below reports the most extreme 21-day returns in the five most extreme events in the US stock market in the sample. As a benchmark, the 5% OTM synthetic puts would ideally generate a return that is 5% lower than the negative of the market return, minus the initial cost of the position. In all five cases, the synthetic puts have highly positive returns, providing economically meaningful insurance against these crashes.

Returns on the market and synthetic puts, five most extreme events

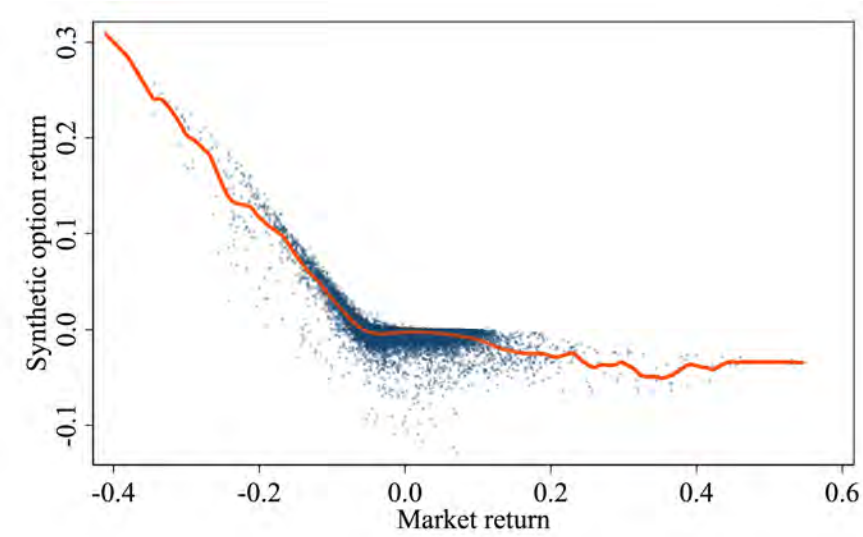
	$R^m_{t,t+j}$	$\bar{R}^S_{t,t+j}$	$\bar{R}^S_{t,t+j}$ (no lag in δ^S_t)	Ideal return
Nov. 1929	-41%	31%	32%	35%
Mar. 2020	-33%	25%	25%	28%
Oct. 2008	-31%	22%	22%	24%
Oct. 1987	-30%	17%	23%	24%
Oct. 1931	-29%	22%	22%	24%

Recall that in the benchmark results, δ^S_t depends on data only up to date $t-1$ in order to avoid microstructure biases. The third column in the table shows that when δ^S_t uses date-t information, the hedge becomes noticeably better, particularly in 1987.[14] This shows that

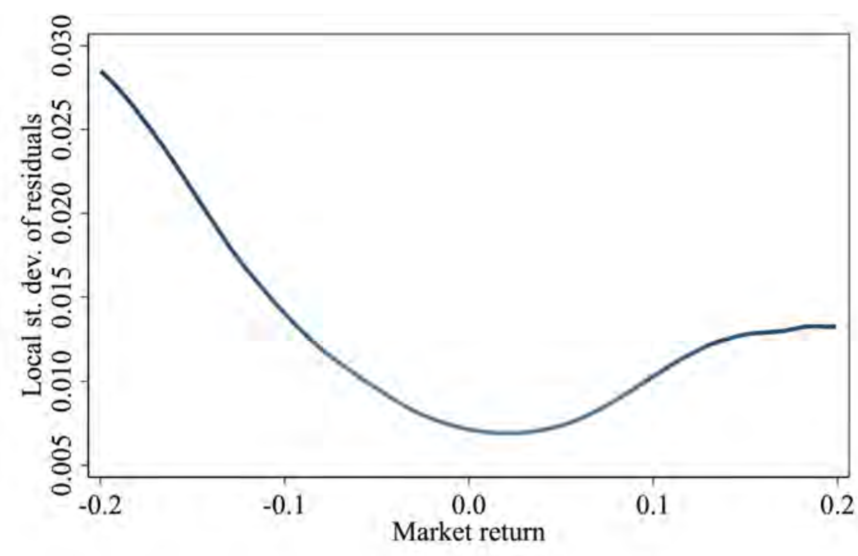
[13] Conditional expectations are calculated via a local linear regression on $R^m_{t,t+j}$ with a Gaussian kernel and the bandwidth set to 0.01.

[14] Using date t information means that the investment held during day $t+1$ (e.g. Wednesday) is chosen based on information at the end of date t (Tuesday afternoon; as opposed to the benchmark using info from

Figure 3: Synthetic put returns as a function of the market



(a) R^S vs. R^m



(b) Local standard deviation of redsiduals ($\hat{R}^S_{t,t+j}$)

Note: Panel (a) plots returns on synthetic options, R^S, against the returns of the market, R^m. The red line is a kernel estimate of the local mean. Panel (b) plots the local standard deviation of the residuals.

using lagged information for hedging is conservative for fit.

Even though the synthetic options fit well, the claim is not that the replication was implementable. The results just show that the hypothetical returns are nonlinear in the market, so that the first required condition from section 2.3 is satisfied empirically. In addition, even though the synthetic strategy did a good job of hedging the crashes that have occurred in US data over the last century, it is always possible that investors worry about a disaster that has not been realized yet in which the replication would perform worse.

To begin to evaluate the third condition from section 2.3, that the residual risk is small, the bottom panel of figure 3 plots the conditional standard deviation of the residuals $\hat{R}^S_{t,t+j} \equiv R^S_{t,t+j} - \bar{R}^S_{t,t+j}$.[15] That standard deviation is always less than 3%, and in many cases less than 1%. In addition, we have the following variance decomposition:

$$\underbrace{\text{var}\left(R^S_{t,t+j}\right)}_{3.1\times10^{-4}} = \underbrace{\text{var}\left(\bar{R}^S_{t,t+j}\right)}_{2.5\times10^{-4}} + \underbrace{\text{var}\left(\hat{R}^S_{t,t+j}\right)}_{0.6\times10^{-4}} \tag{10}$$

21% of the variation in synthetic option returns in this case are unspanned by the market return. The amount of residual risk measured here is again a conservative estimate due to the fact that δ^S_t is constructed using only lagged information. If δ^S_t uses date-t information, the fraction of the variance of $R^S_{t,t+j}$ from $\hat{R}^S_{t,t+j}$ falls to 15%. The more important question, though, is whether that variation is correlated with a typical investor's marginal utility, which we return to in section 3.4.

Finally, to directly compare synthetic and traded option returns, figure A.2 compares standard deviations and betas for traded and synthetic options across strikes and finds they are highly similar. Figure A.9 plots synthetic against traded option returns for different

Monday afternoon).

[15] The local volatility is estimated from a second kernel regression,

$$\eta^2_{t,t+j} = h\left(R^m_{t,t+j}\right) + residual \tag{9}$$

where for the function h we set the bandwidth to 0.05 due to there being greater variation in the squared residuals around the fitted value.

strikes and reports pairwise correlations, which range between 0.85 and 1.00.[16]

3.3 Risk premia

3.3.1 Varying strikes at the monthly maturity

Figure 4 reports the paper's key results for long-run average option risk premia. It includes results for three different periods: the full sample available for synthetic returns (1926–2022), the full sample available for both synthetic and traded returns (1987–2022), and the 1987–2005 sample used by Broadie, Chernov, and Johannes (BCJ; 2009), who report an extensive analysis of the performance of traded options. In all cases, the figure gives results for put options. For alphas and information ratios, results for puts and calls are guaranteed to be identical for synthetic options at a given strike, and they are highly similar for traded options (in both cases due to put-call parity).[17]

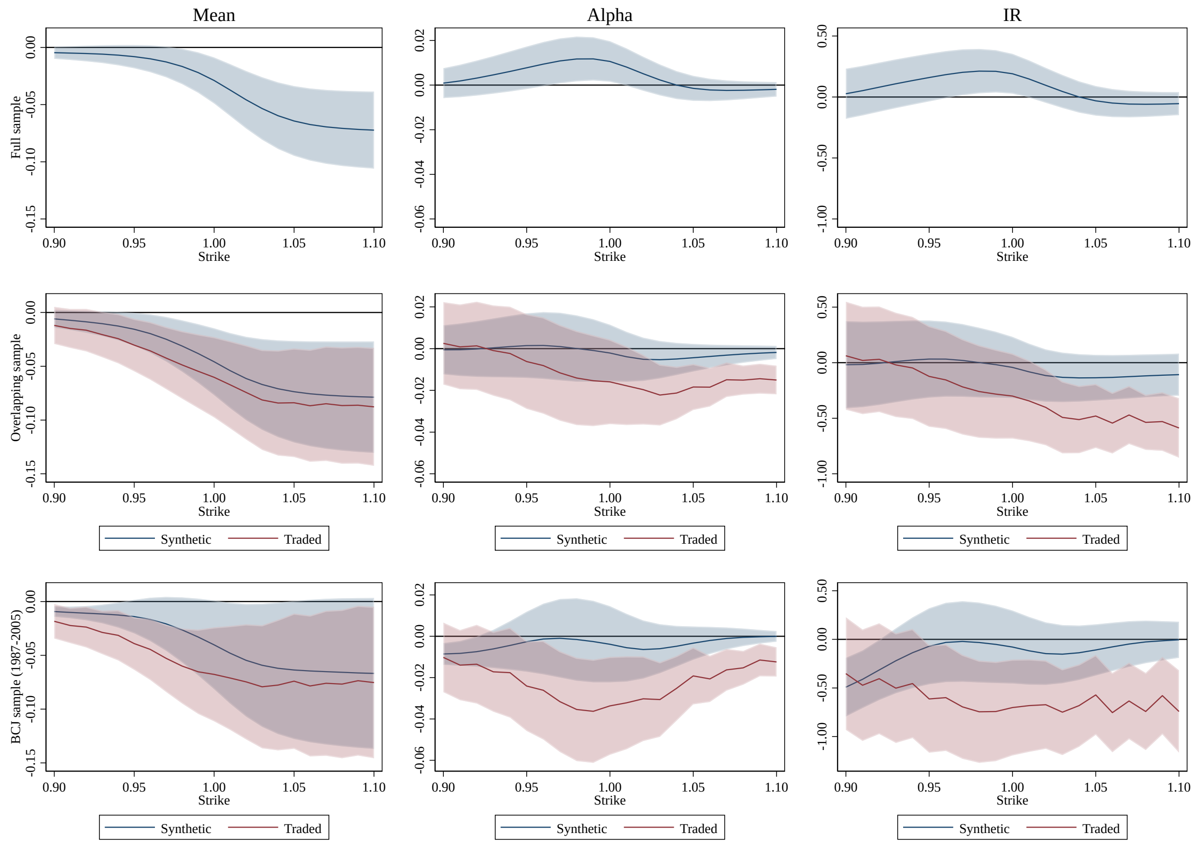
The left-hand column of Figure 4 plots average returns. In all three panels, returns decline as the strike rises, which is to be expected as the betas also become more negative for higher strikes. In the periods of overlap between synthetic and traded options, traded options always have lower average returns than synthetic.

The middle column reports CAPM alphas, which are the paper's key object of interest based on the theoretical analysis. The bottom two panels show that the estimated alphas of traded options are negative across all strikes in both the 1987–2022 and BCJ samples. The statistical evidence for the alphas being negative is stronger in the earlier BCJ sample, with the magnitudes falling by half in the longer sample.

[16] For all results that study returns for a range of strikes above and below the spot, we compute the returns of the traded options at that strike by averaging the returns of the actual put (whether in the money or out of the money) with the return from the put implied by the traded call, as implied by put-call parity. This averages out some of the noise due to the fact that put-call parity doesn't hold exactly with traded options. All the results are similar if using only puts for strikes below the spot and only calls for strikes above the spot.

[17] For traded options, results are not numerically identical between puts and calls because put-call parity doesn't hold exactly with traded options. To mitigate noise, we average the return of the put and that obtained from the call via put-call parity, with minor effects on the results.

Figure 4: Average option returns across strikes



Note: Means, CAPM alphas, and CAPM information ratios across various samples for traded and synthetic put options. Shaded regions represent 95-percent confidence intervals. In the top panel, for the full post-1926 period, only synthetic options are available. The overlapping sample in the middle row is 1987–2022.

In the same post-1987 period, though, and regardless of whether the post-2005 period is included, synthetic options have estimated alphas very close to zero, with no evidence of mispricing relative to the CAPM. The top panel shows that the same result holds in the full sample. The paper's claim that synthetic options have been fairly priced historically relative to the CAPM is based on the results reported here for alphas.

The right-hand column of Figure 4 plots information ratios – the Sharpe ratio of the part of option returns uncorrelated with the market. Again, they are generally statistically insignificant – some are if anything weakly positive. The confidence bands are also economically tight in that they reject information ratios below -0.2 at the 95% confidence level for all strikes, and for the strikes with most of the put volume (0.95 to 1), they reject negative information ratios entirely. The traded options, on the other hand, have information ratios as negative as -0.75 in the BCJ sample, which is larger than the Sharpe ratio of the overall stock market.

While the confidence bands overlap slightly in some cases, the alphas and information ratios are in general statistically significantly different from each other, especially in the BCJ sample (and, again, it is an important part of the paper's story that the difference actually weakens after that sample). Figure A.3 in the appendix shows that result formally. The reason that the difference is well estimated even though the individual confidence bands overlap somewhat is that the two series are very highly positively correlated.

The figures therefore provide economically and statistically significant evidence that synthetic options have had much less negative alphas and information ratios than traded options.

Figure A.4 replicates 4, but varying the maturity instead of the strike price, figure A.5 scales the moneyness in volatility units, and figure A.6 puts the price of the option in the denominator instead of the level of the underlying. The results are similar to the baseline in all three cases.

3.3.2 Cumulative alphas and variation in risk premia over time

The top panel of figure 5 plots cumulative CAPM alphas for synthetic 5% OTM options over the period 1926–2022 and for traded put options over the period 1987–2022. For readability, the cumulative returns are normalized to zero in July 1987 when the data for the traded options begins.

For synthetic options, the figure reinforces the result that over the full sample the alphas have been slightly positive. Covid jumps out in 2020 as a large positive innovation, due to the significant decline in the level of the stock market. The fact that the synthetic portfolio captures that gives clear evidence that it is able to capture economically significant large declines in the market. The bottom panel of figure 5 plots information ratios over rolling

10-year windows and again shows that the returns on synthetic options have been stable, with a brief period in the 1930's when the returns were statistically significantly positive. In no period were they significantly negative.

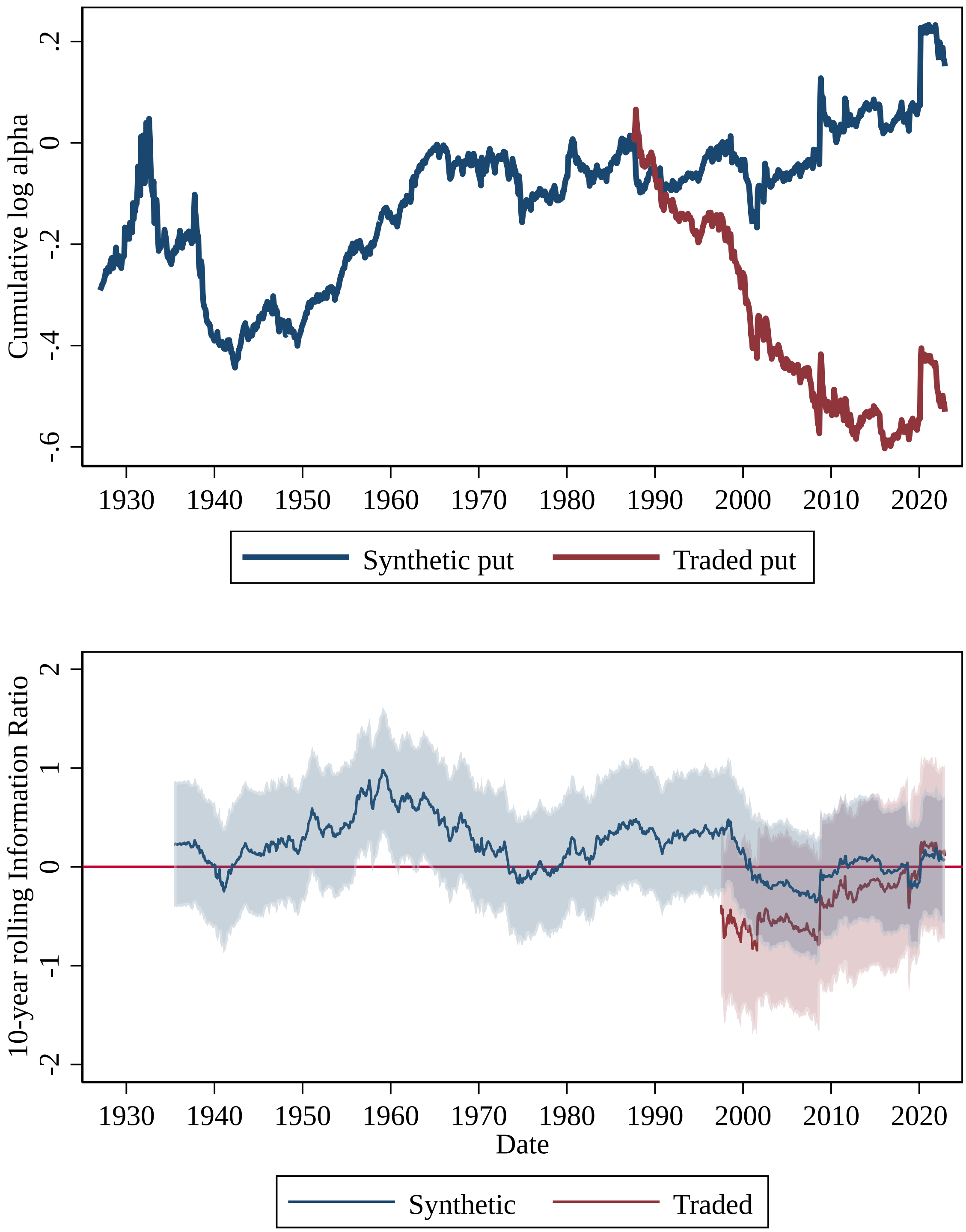
Traded put returns have two striking features. On the one hand, the month-to-month variation appears very similar to that for the synthetic options, consistent with the results presented so far. On the other hand, the average return is drastically different. The returns are highly negative, especially in the period up to 2010. The returns were roughly flat from then until the large market decline with Covid. In fact, the overall cumulative return on traded puts is zero between March, 2009 and the end of the sample in December, 2022. The bottom panel of figure 5 shows how, over ten-year windows, the return on traded puts actually turned positive at the end of the sample.

Since a synthetic put is a delta hedge, the difference between the returns on the traded and synthetic put returns is exactly the return on a delta-hedged put, which is a version of the variance risk premium (Bakshi and Kapadia (2003)). To see that, figure A.11 in the appendix plots cumulative alphas for delta-hedged puts and straddles. They have been approximately zero since 2010**.** Section 4 revisits this point in more detail and reports formal tests for a structural break.

3.3.3 A conditional CAPM interpretation

Since synthetic options are created by trading the market dynamically, any monthly CAPM alpha they earn has to come from timing the market risk premium. For both synthetic puts and calls, the investment in the market, δ_t^S, declines – becoming more negative for a put and less positive for a call – when the market declines and rises when the market rises. Synthetic options are therefore bets on momentum. To get a negative alpha (which would be consistent with traded option returns) would require mean reversion in returns.

Figure 5: Cumulative returns



Note: The top panel plots cumulative CAPM alphas for traded and synthetic -5% put options. The lines are constructed to equal zero in July, 1987, when the true put options become available. The bottom panel plots 10-year rolling CAPM information ratios, with the shaded regions representing 95-percent confidence intervals.

Formally, one can derive from results in Lewellen and Nagel (2006) that

$$\alpha_{t,t+j}^{S} \approx \text{cov}\left(\delta_t^S, \left[E_t\left[R_{t,t+1}^m\right] - E\left[R_{t,t+1}^m\right]\right] - \frac{E\left[R_{t,t+1}^m - 1\right]}{\text{var}\left(R_{t,t+1}^m\right)}\left[\text{var}_t\left(R_{t,t+1}^m\right) - \text{var}\left(R_{t,t+1}^m\right)\right]\right) \tag{11}$$

The first part of the covariance is the usual conditional CAPM intuition, which says that if δ_t^S covaries positively with the market risk premium, then $\alpha_{t,t+j}^S$ will be positive. The second part is a contribution from the comovement of δ_t^S with conditional volatility – the movement of deltas with volatility is second-order (and its sign is ambiguous), so this term is small quantitatively. The equation can also be interpreted in the opposite direction: if synthetic puts earn a negative alpha then (ignoring second-order volatility effects) expected returns must be countercyclical. That is, convexity in SMU implies countercyclical risk premia.[18]

3.4 The effect of unspanned variation – $\hat{R}_{t,t+j}^S$ and $\hat{M}_{t,t+j}$

Recall from section 2.3 that the curvature of marginal utility is determined by $\alpha_t^S + \text{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}_{t,t+j}^S\right)$ where the variables with hats are the components unexplained by any nonlinear function of the market return. So far the analysis has ignored the covariance term. The questions are whether that covariance is zero and if not, how large it might be. This section examines what risk factors $\hat{R}_{t,t+j}^S$ might be correlated with and then bounds the magnitude of $\text{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}_{t,t+j}^S\right)$ using the method of Cochrane and Saa-Requejo (2000).

[18] How does time-varying volatility affect this analysis? If the CAPM holds period-by-period with constant risk aversion (in the sense of the pricing kernel being linear in $R_{t,t+1}^m$), then $E_t\left[R_{t,t+1}^m - 1\right] \propto \text{var}_t\left(R_{t,t+1}^m\right)$, which would imply that the covariance in (11) is identically zero. If risk aversion is countercyclical, as with a convex pricing kernel, then even if volatility rises when the market falls, $E_t\left[R_{t,t+1}^m - 1\right]$ will rise by enough to offset that effect, so that the covariance term is negative and $\alpha_{t,t+j}^S < 0$. In other words, if both volatility and risk aversion are countercyclical, equation (11) implies that $\alpha_{t,t+j}^S$ is negative.

3.4.1 Relationship of $\hat{R}^S_{t,t+j}$ with risk factors

Table 1 reports correlations between the unspanned part of synthetic option returns, $\hat{R}^S_{t,t+j}$, and statistical innovations in prominent macro and financial variables.[19] Since $\hat{R}^S_{t,t+j}$ is orthogonal to the market return by construction, we also orthogonalize the innovations in all of the other macro and financial time series with respect to the market return.

Table 1: Correlation of residuals with macro variables

	All data	Excluding 2020
Unemployment	-0.08	-0.06
Ind. Pro. Growth	0.06	0.07
Employment growth	0.07	0.06
FFR	-0.01	-0.01
Term Spread	0.02	0.02
Default Spread	-0.01	-0.01
EBP	-0.11	-0.14
VIX	-0.21	-0.21
VXO	-0.14	-0.12
rv	-0.35	-0.38
Maximal corr	0.35	0.41

Note: Table reports the correlations between the residuals of the nonlinear fit of R^S onto the market and various macroeconomic variables: unemployment, industrial production growth, employment growth, the federal funds rate, the term spread (10 year minus 1 year), the default spread (BAA-AAA spread), the excess bond premium (EBP) from Gilchrist et al. (2021), the VIX, the VXO, and realized volatility. All variables are orthogonalized to the market. The last row reports the maximal correlation between any linear combination of these variables and the residuals. The second column replicates the results excluding 2020.

Among the macro time series, the correlations are all economically small and statistically insignificant. The only notable correlations are for financial series: the excess bond premium (EBP), the VIX, and realized volatility. In months in which shocks to these financial series, after orthogonalizing with respect to the market return, are unexpectedly high, $\hat{R}^S_{t,t+j}$ tends

[19] For $\hat{R}^S_{t,t+j}$ we take the return from the beginning to the end of a month. For the variables that are measured at a fixed point in time, we take the statistical innovation in the value at the end of the month relative to the lags of the variable (information available at the beginning of the month). For the other variables, we take the statistical innovation in the monthly value relative to data available in the previous month. The different time series are available for different time periods and each correlation is computed using the longest period available for that variable.

to be low.[20] If those are bad states of the world, that would make $\hat{R}^S_{t,t+j}$ risky, which proposition 3 shows would imply more convexity in SMU than implied by $\alpha^S_{t,t+j}$.

The results suggest that options do not look risky to an investor whose marginal utility depends on either the level of the stock market or to macroeconomic variables. They *do* look risky, though, to an investor who cares about the *path* that the market return takes, suggesting that intermediaries might be relevant for pricing.

The bottom row of table 1 reports the maximum correlation of $\hat{R}^S_{t,t+j}$ with any linear combination of the innovations. It is less than 0.5, so we take that as an upper end for a reasonable estimate of the correlation of $\hat{R}^S_{t,t+j}$ with $\hat{M}_{t,t+j}$, but we also examine results with the correlation set to 1.

3.4.2 Robust uncertainty bands

If $\text{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}^S_{t,t+j}\right)$ is not equal to zero, how large of a bias does it create in measuring curvature in marginal utility? To bound the magnitude of $\text{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}^S_{t,t+j}\right)$, following Cochrane and Saa-Requejo (2000), start from the identity,

$$\left|\text{cov}_t\left(\hat{R}^S_{t,t+j}, \hat{M}_{t,t+j}\right)\right| = \left|corr_t\left(\hat{R}^S_{t,t+j}, \hat{M}_{t,t+j}\right)\right| std_t\left(\hat{R}^S_{t,t+j}\right) std_t\left(\hat{M}_{t,t+j}\right) \tag{12}$$

$std\left(\hat{R}^S_{t,t+j}\right)$ can be estimated based on the empirical time-series of $\hat{R}^S_{t,t+j}$. $\left|corr\left(\hat{R}^S_{t,t+j}, \hat{M}_{t,t+j}\right)\right|$ is not observable, but the results in the previous section imply 0.5 as an estimate, and the upper bound is 1. Finally, to get $std\left(\hat{M}_{t,t+j}\right)$ we assume that the volatility of the unspanned part of SMU, $\hat{M}_{t,t+j}$, is no greater than that from the part of SMU spanned by the market, $\bar{M}_{t,t+j}$. That implies that

$$std\left(\hat{M}_{t,t+j}\right) \leq E\left[R^m_{t,t+j} - 1\right] / std\left(R^m_{t,t+j}\right) \tag{13}$$

[20]The correlations of the unspanned option returns with the SMB and HML factors are also very small – 0.02 and -0.05, respectively.

Intuitively, that restriction says that the Sharpe ratio available from any investment independent of the market return can be no greater than that of the market itself, similar to Cochrane and Saa-Requejo (2000) (see also references therein).[21]

The parameter of interest, which measures convexity in SMU, is

$$\alpha_{t,t+j}^{S,adjusted} \equiv \alpha_{t,t+j}^{S} + \text{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}_{t,t+j}^{S}\right) \tag{14}$$

It has two sources of uncertainty: estimation uncertainty in $\alpha_{t,t+j}^{S}$ and the unobservable value of $\text{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}_{t,t+j}^{S}\right)$. Appendix C shows how those two sources of uncertainty can be combined geometrically, essentially treating $\text{cov}\left(\hat{R}_{t,t+j}^{S}, \hat{M}_{t,t+j}\right)$ as another Gaussian source of error.

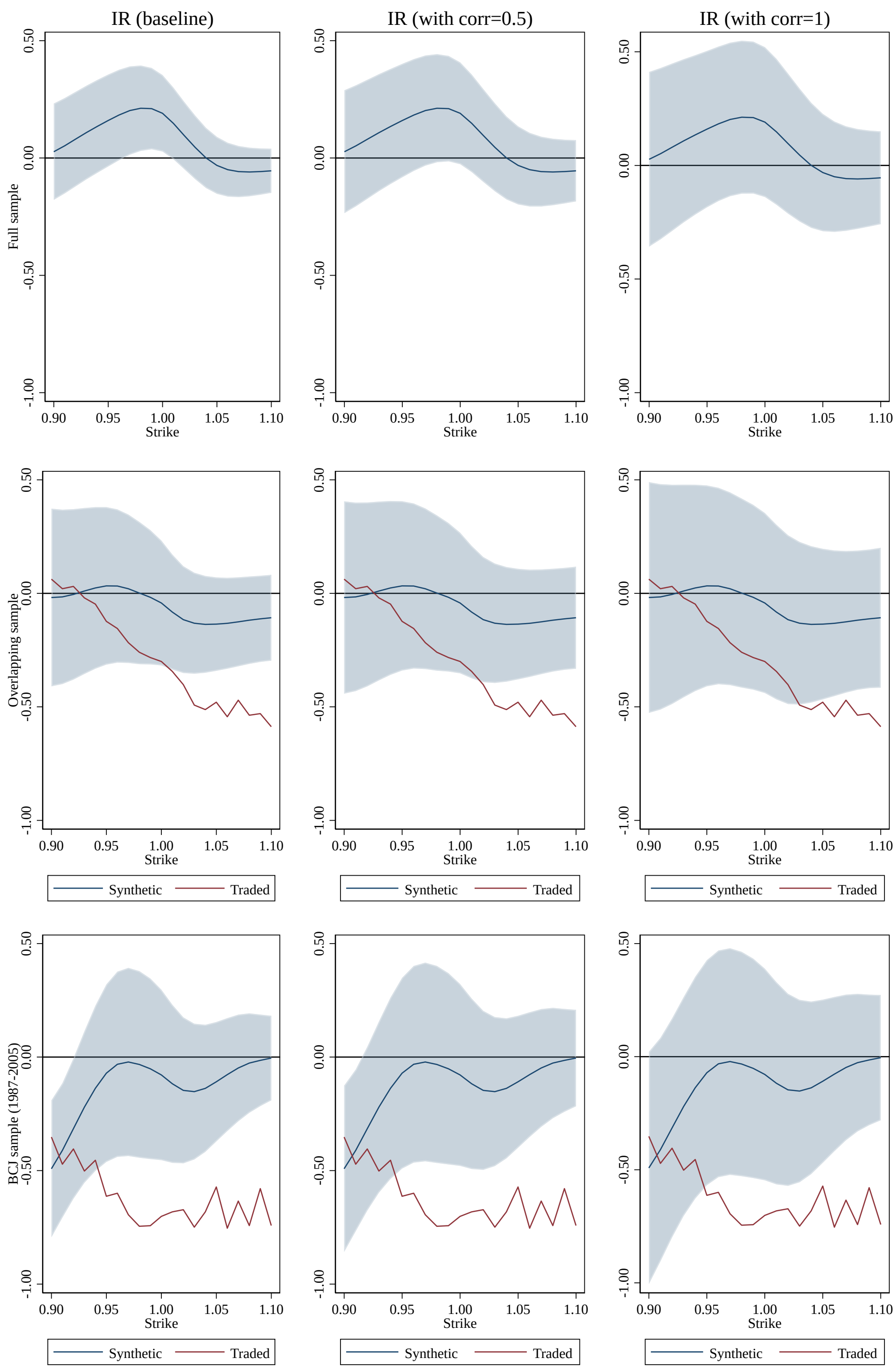
Figure 6 reports an alternative version of figure 3 that now incorporates these robust uncertainty bands instead of the original confidence bands based only on estimation error. The left-hand panels assume that $corr\left(\hat{R}_{t,t+j}^{S}, \hat{M}_{t,t+j}\right)$ based on the results in the previous section, while the right-hand panels use the most conservative possible value of 1.[22]

The uncertainty bands in figure 6 are guaranteed to be wider than in the baseline. However, they can still reject information ratios of -0.5 in all but a few cases with the shortest sample. In the top-center panel, which is the most powerful case, using the longest sample and $corr\left(\hat{R}_{t,t+j}^{S}, \hat{M}_{t,t+j}\right) = 0.5$, the bound can reject even small negative information ratios. So even when accounting for unspanned risk, the curvature of SMU implied by synthetic options remains small. That said, naturally these bounds are dependent on what one takes to be the maximum available Sharpe ratio. There is always a price for the residual risk such that any curvature of SMU may lie within the confidence bands.

[21] The 1 here represents the gross-risk-free rate. Again, in the empirical analysis the actual risk-free rate is used.

[22] Equivalently, the right-hand panel can be taken as treating the market Sharpe ratio as 1 instead of 0.5 (and leaving the correlation at 0.5).

Figure 6: Average option returns with robust uncertainty intervals



Note: These graphs replicate the main results, but the shaded uncertainty intervals here incorporate the bound on the effect from potential pricing of unspanned risk in the synthetic option returns.

3.5 Implications for marginal utility

Figure 2 plots, in blue, the shape of SMU implied by the synthetic option returns for a strike of -5% at maturity of one month for various samples. In all cases, the synthetic options imply that marginal utility is if anything weakly concave, consistent with the positive measured alphas and implying risk aversion falls slightly in bad times. While there are no confidence bands plotted in figure 2, the change in the slope across the strike is measured by the alphas reported above, and so the confidence bands apply here.

As discussed above, the red lines corresponding to traded options imply significant convexity due to the large negative estimated alphas, so that effective risk aversion rises strongly as market returns fall.

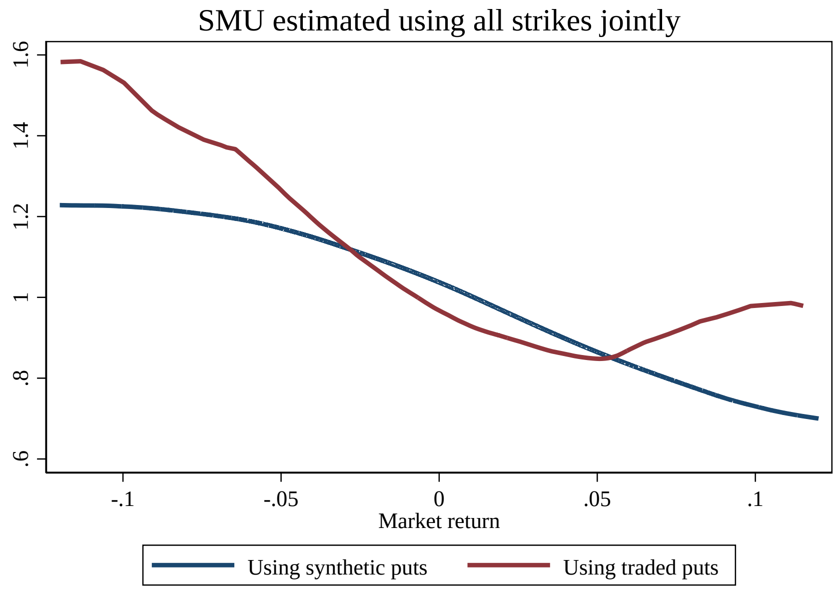
The analysis in section 2 of how to estimate SMU based on the returns on the market and a single option is a special case of the minimum-variance SDF of Hansen and Jagannathan (1991) and naturally extends to using multiple options simultaneously.

Figure 7 plots the minimum-variance SMU for the traded and synthetic options separately using strikes at 5% moneyness intervals from 90 to 110 percent (for 1926–2022 for synthetic options and 1987–2022 for traded). The mean-variance optimal portfolio is calculated based on the full-sample estimates of mean returns and covariances, with both a lasso- and ridge-type adjustment for robustness.[23]

As in the benchmark case, estimated SMU is convex for the traded options and concave for the synthetic options. For the synthetic options, the concavity is fairly consistent, though it may change signs at large positive strikes. For the traded options, the convexity appears strongest around market returns of zero. While traded options again imply non-monotone marginal utility, the synthetic options do not.

[23] Mean-variance optimization requires inverting the covariance matrix. We make two modifications to ensure that the inversion is well behaved. First, we inflate the main diagonal of the covariance matrix by 10%, which corresponds to a ridge-type adjustment. Second, in constructing the inverse, we drop any eigenvalue smaller than 0.01 times the largest (in practice this eliminates one eigenvalue). These adjustments only affect the weights in the tangency portfolio.

Figure 7: Estimated SMU using all options jointly



Note: The figure reports the estimated SMU obtained by using options of all strikes jointly. The blue line uses synthetic puts (sample 1926–2022), and the red line uses traded puts (sample 1987–2022).

3.6 Robustness

The baseline results use deltas from the method of Hull and White (2017). The left column of figure A.10 shows (for different sample periods) that the results are highly similar simply using the standard Black–Scholes delta.

Constructing the weights for the synthesis requires a forecast of volatility. The benchmark analysis uses a recursively estimated HAR model (Corsi (2009). The middle column of figure A.10 shows that the alphas are very similar to those in the baseline if the volatility used to calculate the hedge weights is simply set to 0.15 on all dates. That shows that the results are driven by how δ_t^S depends on the level of the market, rather than its volatility.

The benchmark analysis uses Hansen–Hodrick asymptotic standard errors. Since option returns can be highly non-normal, convergence to the asymptotic distribution might be slow, which can be addressed via bootstrapping. Results with block-bootstrapped standard errors with block length equal to two months are reported in right column of figure A.10, and they are highly similar to the baseline.

3.6.1 Effects of conditioning on betas

In the theoretical analysis, all alphas and betas are conditional, and hence potentially time-varying.[24] Figure A.7 examines possible ways of accounting for time-variation in conditional betas. The left-hand column of panels plots the baseline results. In the middle column, betas are estimated from a rolling three-month window. The right-hand panels model conditional beta as a function of lagged (i.e., end of previous month) variables: the market return volatility forecast (which is most important), the market return itself, industrial production growth, and the corporate bond default spread. In both cases, the results are highly similar to the benchmark qualitatively and quantitatively.

3.6.2 Effects of daily mispricing

Recall the second condition for using synthetic options to measure curvature in marginal utility from section 2.3 that the market is priced correctly every day. Obviously no such condition is literally true, so the question is how pricing errors might bias the results. Appendix D shows that if pricing errors create spurious serial correlation in returns, e.g. via stale prices or bid-ask bounce, they will bias the results. The bias can be eliminated, though, by using information lagged beyond the order of autocorrelation in returns.

To examine this effect in the data, the top panel of figure A.8 reports the autocorrelations of daily returns in the full sample and pre- and post-1973 separately. In the pre-1973 sample, there is clear evidence of one-day positive serial correlation, consistent with the presence of stale prices. The two-day autocorrelation, on the other hand, is negative, which is one reason why the analysis only uses a single-day lag in constructing the weights.

To see the effects of different choices for the information set for δ_t^S, the bottom panel

[24] Note that in the theoretical analysis, all conditioning for an option return between dates t and $t+j$ is taken as of date t. That is, for monthly returns, we must condition on information available at the *beginning* of the month, not over the course of the month.

of figure A.8 plots three versions of the cumulative alphas for synthetic options: with the baseline one-day lag, with no lag, and with a two-day lag. Switching from the baseline to no lag causes a large increase in alphas, entirely due to the positive autocorrelation, which is why it shows up only in the first half of the sample. Going from a one-day to a two-day lag also increases the alphas (by a lower amount). None of the lag choices leads to negative average alphas.

4 The decline of option overpricing and the role of intermediaries

This section shows that over the last 10-15 years the negative alpha associated with traded options has essentially disappeared and the risk premium of traded options has converged to that of synthetic options. It then develops a simple general equilibrium model that can explain that pattern as well as those documented in the previous section. Last, it shows that the decline in overpricing was contemporaneous with a decline in the risk borne by financial intermediaries, consistent with the model.

4.1 The decline of option overpricing: empirical evidence

Figure 5, presented in the previous section, already showed some suggestive evidence that the returns on traded options may have trended towards zero in recent years. Figure 8 examines that behavior in greater detail. The left-hand panel of figure 8 plots information ratios over rolling 10-year windows for traded and synthetic puts along with their difference for the 1987–2022 sample.[25] The synthetic options consistently had zero or positive returns, while the traded options consistently had strongly negative returns until 2009, when they begin to

[25] The main results use the overlapping 21-day returns for synthetic option. In this section, all results use synthetic option returns that match exactly the roll dates of the traded options, in order to ensure comparability. This has only minor effects, and in any case they run against the main conclusions of this section.

trend up, eventually turning starkly positive in 2020. In the early periods, the information ratio for the traded options was very large – about equal to the size of the market risk premium itself.

The difference between the information ratios on traded and synthetic options, plotted in the middle panel of figure 8, has also nearly converged to zero over the same period. The confidence bands show that the change in the information ratio appears to be highly statistically significant, something that is tested more formally below, in section 4.1.1.

Figure 8: Changes in premia over time

Note: The left-hand panel plots 10-year rolling CAPM information ratios for traded and synthetic 5The middle panel plots the difference between those two series for each 10-year window along with a 95-percent confidence band. The right-hand panel plots information ratios for three other related measures of the difference between traded and synthetic options.

The difference in the information ratios between traded and synthetic options is very closely related to the return on delta-hedged options, which have been studied widely in the past literature and used as a proxy for the variance risk premium (e.g. Bakshi and Kapadia (2003)). To examine that idea more directly, the right-hand panel of figure 8 plots the rolling 10-year information ratios for delta-hedged 5% OTM put options and at-the-money straddles. Both have risen over time, converging to zero in 2020 (with a complete convergence for the ATM straddles coming even before Covid). As an even simpler test, the green line in the same plot proxies for the payoff of a variance swap simply as the gap between realized variance and the squared VIX (Carr and Wu (2009)). The information ratio of RV-minus-VIX behaves highly similarly to the other series, also converging to zero over the sample.[26]

Again, most or all of the convergence occurred *before* the large market declines due to Covid in March, 2020. As of January, 2020, the difference in rolling information ratios in the middle panel of figure 8 is no longer significant.

The results here *do not* mean that the variance risk premium is zero by the end of the sample. There is still a premium for variance risk, but the results imply that the premium is no larger than what would be expected from the CAPM beta. That is, the variance risk premium can be decomposed into parts coming from CAPM alpha and beta, and the results here just show that the alpha contribution has shrunk.

Variance rises when the market falls, so it has a negative beta and carries a negative premium. In the past, the premium from trading variance via delta-hedged options was even larger than is implied by the CAPM beta (i.e. there was also a negative alpha), but by the end of the sample that is no longer true. These results are consistent with those in Heston, Jacobs, and Kim (2022), who also find that there is a negative variance risk premium, but that it cannot be distinguished from simple market (beta) risk.

4.1.1 Statistical tests

We now test statistically that the information ratios have changed between the early part of the sample and the late part of the sample. To maximize power, we divide the sample in two parts. The general equilibrium model in the next section predicts that the risk premium of traded options, relative to synthetic options, should depend on the total gamma (i.e., hedging) risk borne by intermediaries – when the intermediaries have to hold a lot of this risk, they require extra compensation from option-buyers. As the next section will show, intermediaries switched from being net short to having zero or slightly positive net exposure

[26] Specifically, define a (pseudo-) return $R_t^{RV} = \left(RV_t - VIX_{t-1}^2\right)/VIX_{t-1}^2$, where RV_t is (annualized) realized variance in month t and VIX_{t-1} is the level of the VIX at the end of month $t-1$. The information ratio is then the CAPM alpha of R_t^{RV} divided by the residual standard deviation.

around 2012 (specifically, a break test identifies 2012m5 as the break date). We therefore use this date to distinguish the "early" and "late" part of our sample. That said, the results of this section are very similar for the other break dates around 2010.

Table 2: Test for change in risk premia over time

IR (traded)			IR (synthetic)			IR (traded-synth)		
Before	After	p-value	Before	After	p-value	Before	After	p-value
-0.60	0.09	0.06	0.06	0.35	0.46	-0.66	-0.25	0.02

Note: Table reports time-series tests of whether the average information ratio of traded (left), synthetic (middle), and their difference (right) are different before 2012m5 and after, with the p-value for their difference. Full sample is 1987-2022.

Table 2 reports, respectively for traded options (left), synthetic options (middle), and for their difference (i.e., the delta-hedged options, right), the information ratio before and after that break date, as well as a p-value for their difference. As the table shows, the information ratio of traded options changed not only economically (from -0.6 to 0.09) but also statistically significantly over this time period. Similarly, the difference between traded and synthetic options also showed a significant change, both economically and statistically. But synthetic options didn't see a significant change in the information ratio during this period. Therefore, consistent with the graphs showed in the previous section, the gap between traded and synthetic options has fallen significantly over time, by traded options converging to synthetic options, rather than the opposite.

4.2 A general equilibrium model of the options market

This section presents a simple equilibrium model that can qualitatively match the results presented so far – that synthetic options have earned zero CAPM alpha while traded options earned a large negative premium that has converged to zero. In the model, that shift is driven by a decline in frictions inhibiting investors (other than dealers) from selling – but not buying – options.

In order to model frictions in trade, we consider an overlapping-generations model in which agents can only trade equity and options at "birth". The "lives" of the investors will be calibrated to be equal to the maturity of options. In that sense, these investors can be thought of as simply myopic portfolio decisions and rebalancing once a month. The goal of the model is to understand the circumstances under which synthetic and traded options do and do not earn CAPM alphas. Note that since investors can only trade equities once, the model captures the idea that typical investors cannot themselves dynamically replicate options.

The agents in the model represent those who face investment frictions. They can be thought of as retail investors, but more generally they are agents who are not market-makers or dealers and hence not fully integrated with financial markets.

4.2.1 Model setup

There are two types of agents, indexed by i. Agents live for $J+1$ periods. They consume in each period and also have a bequest motive. The agents differ in how their effective risk aversion varies over time.

Budget constraints. For an agent who was born on date t, and hence has J periods remaining to live,

$$C_{i,J,t} + B_{i,J,t} + P_t^X X_{i,t} + \underbrace{P_t^O O_{i,t}}_{\text{If options are tradable}} = P_t^X \tag{15}$$

where $C_{i,j,t}$ is consumption on date t of an agent of type i with j periods remaining to live, $B_{i,j,t}$ is the same for the holding of the riskless bonds, and $P_t^X X_{i,t}$ and $P_t^O O_{i,t}$ are the purchases of equity and J-period options, respectively, for agents of type i born on date t (these do not require j subscripts because equity and options can ony be traded at birth). The right-hand side represents each agent's endowment – they are each born with a unit allocation of equity.

Agents with $0 < j < J$ periods left to live receive dividends, trade bonds, and consume:

$$C_{i,j,t} + B_{i,j,t} = D_t X_{i,t-(J-j)} + R_t^B B_{i,j+1,t-1} \tag{16}$$

where R_t^B is the risk-free rate from date $t-1$ to t and D_t is the dividend paid by equity on date t.

Terminal wealth is

$$W_{i,t} = X_{i,t-J}\left(P_t^X + D_t\right) + O_{i,t-J} X_t^O + R_t^B B_{i,1,t-1} \tag{17}$$

where X_t^O is the payoff of a J-period option on date t.

Objective. Agents have log utility over consumption. On the day they are born their objective is

$$\max E_t^i \left[\log C_{i,J,t} + \sum_{k=1}^{J-1} \beta^k \log C_{i,J-k,t+k} + \beta^J \log\left(C_{i,0,t+J}\right)\right] \tag{18}$$

where E_t^i is agent i's expectation on date t.

Since the agent has log utility, the consumption-wealth ratio is constant. We therefore model terminal consumption as

$$C_{i,0,t} = \frac{C_{i,J,t}}{P_t^X} W_{i,t} \tag{19}$$

where P_t^X is the wealth of an agent born on date t, so that $C_{i,J,t}/P_t^X$ is their consumption/wealth ratio. Intuitively, this is just a way to capture a bequest motive, or, more realistically, the marginal utility of wealth at date $t+J$ when the agent is able to reoptimize.

In general log utility does not generate realistically large risk premia. To do so, we modify the expectation operator to make agents pessimistic over the distribution of shocks. That pessimism can be thought of as a reduced-form for risk or ambiguity aversion, or simply as behavioral. To generate demand for options, we additionally assume that the pessimism (which determines effective risk aversion) varies over time.[27] Specifically, in the model there will be a single fundamental shock $\varepsilon_t \sim N(0,1)$, and on date t, agent i believes that

$$\varepsilon_{t+1} \sim N(-\mu_{i,t}, 1) \tag{20}$$

$$\mu_{i,t} = \phi\mu_{i,t-1} + (1-\phi)\bar{\mu} + \kappa_i \varepsilon_t \tag{21}$$

$\bar{\mu}$ determines average pessimism. Pessimism is procyclical when $\kappa_i > 0$ and countercyclical for $\kappa_i < 0$.

Option specification. For both tractability and simplicity, we model "options" as quadratic contracts on equity, with payoff

$$X_t^O = \left(\prod_{j=0}^{J-1} R_{t-j} - 1\right)^2 \tag{22}$$

$$\text{where } R_t = \frac{P_t^X + D_t}{P_{t-1}^X} \tag{23}$$

A quadratic contract is equivalent to a particular portfolio of options (Bakshi and Madan (2000)). On any given date, its exposure to the underlying is proportional to the cumulative return since inception. And that is exactly why it plays a role for investors in the model. Following positive equity returns, agents with countercyclical pessimism are relatively more optimistic about future returns, increasing their desired allocation to equities. Since they cannot change their exposure after their first period of life, the option is valuable to them for inducing that dynamic reallocation automatically. Agents with countercyclical pessimsm will thus tend to demand options, while those with procyclical pessimism (intuitively, agents who think returns display momentum, rather than mean reversion) will tend to supply options. Agents here thus have a desire to buy and sell options *precisely* because they cannot replicate them dynamically.

[27]Similar to the habit formation of Campbell and Cochrane (1999), or the endogenous time-varying pessimism in Bidder and Dew-Becker (2016).

4.2.2 Calibration and solution.

The equilibrium concept is standard:

Definition 4 *An equilibrium is a set of processes for prices,* $\left\{P_t^X, P_t^O, R_t^B\right\}$*, and the agents' demands,* $\{C_{i,j,t}, B_{i,j,t}, X_{i,t}, O_{i,t}\}$*, such that markets clear,*

$$\sum_i X_{i,t} = \sum_i 1, \quad \sum_i O_{i,t} = 0, \quad \sum_i \sum_{j=1}^J B_{i,j,t} = 0$$

and agents maximize (18).

The model is calibrated to the weekly frequency. Since the model is meant to match behavior of monthly options, that implies $J = 3$. We calibrate the volatility of dividends in order to generate equity returns with similar volatility to that of the aggregate stock market, so we set

$$\log D_t = \log D_{t-1} + \frac{15\%}{\sqrt{52}}\varepsilon_t \tag{24}$$

$$\varepsilon_t \sim N(0,1) \tag{25}$$

We set $\bar{\mu} = -\frac{1}{2}\sqrt{1/52}$ to generate an annualized Sharpe ratio for equities of about 1/2 (the additional risk aversion from log utility makes it slightly higher). We set $\phi = 0.79$, so that pessimism has a half-life of one month. κ for the procyclical and countercyclical agent are set to $\pm\bar{\mu}\sqrt{1-\phi^2}$, which implies that the unconditional standard deviation of pessimism is equal to its mean.

Finally, the rate of time preference is calibrated to 5% per year in order to generate a plausible risk-free interest rate.

We numerically approximate and solve the model via a fourth-order perturbation using Dynare.

4.2.3 Results

To analyze the model, we consider two periods – one in which all agents may buy but not sell options, and a second in which they are free to both buy and sell. This change is meant to capture the decline in frictions to trading options. In the early period, when agents can buy but not sell options, the market clearing price is the one such that the maximum option demand across all agents is equal to zero (some agents may have negative demand – they would like to sell but can't). When they can both buy and sell, we look for the standard market-clearing price.

Table 3 reports simulation results for Sharpe and information ratios. The first pair of columns report results for the heterogeneous-agent version of the model where the pessimism of one agent is countercyclical and the other procyclical. Column 1 corresponds to the early and column 2 the late period. Across the two columns, the Sharpe ratio of equities is the same, since options do not change the total equity risk that must be borne. In the early period, traded options earn significant negative Sharpe and information ratios. Synthetic options earn no premium, though, and therefore delta-hedged option returns are also significantly negative. The correlation between traded and synthetic options, at 0.81, is large but also meaningfully different from 1. In the late period, the results are similar except that traded options, with or without delta-hedging, no longer earn a negative premium.

Why do synthetic options earn no premium? In this model, average effective risk aversion – coming from the combination of curvature in utility and the tilt in the expectation operator – is constant. Following negative shocks, when one agent becomes more risk averse, the other's risk aversion declines by exactly the same amount. So marginal utility averaged across agents, which ultimately determines the market SDF, has a constant slope with respect to market returns. Synthetic options therefore correctly capture that marginal utility is effectively linear – the convexity for half the agents is canceled out by the concavity of the other half.

Another way to look at it is that the agents who are able to trade equities (those who have just been born) have the same average risk aversion as those who cannot trade. So even though not everybody trades all the time, equity is priced *as if* all agents trade. Going back to section 2, *equities* here are priced by a representative – average – investor each day – and therefore, *synthetic options* recover the preferences of the representative investor. On the contrary, *traded options* are not priced by the average investor in the period when their trade is restricted, and thus do not measure the preferences of the average agent: instead, they represent the preferences of just the subset of agent who desire to *buy* options (whose demand drives prices and who price options on the margin).

Overall, the first two columns qualitatively match the empirical results reported so far: there is an early period in which traded options (both alone and delta-hedged) earn significant negative risk premia and a later period in which that premium disappears, while synthetic options never earn a premium. In addition, since the model matches the behavior of both synthetic and traded options, it also matches the behavior of delta-hedged options, whose returns are just the difference between traded and synthetic options.

The second pair of columns reports results for a version of the model with homogeneous agents with countercyclical pessimism. Since the agents are identical, it is irrelevant whether they can trade or not, so the two columns are identical. As in the heterogeneous case with restricted trade, options again earn negative returns and with an magnitude. Intuitively, this is because the agents who effectively drive the price of options in the first column – the countercyclical agents with high option demand – are identical to the agents in the homogeneous case. The difference here, though, is that now synthetic options also earn large negative premia – which also means that delta-hedged options earn no premium. Just like before, synthetic options successfully recover the preferences of the representative agent. In the homogeneous case, average effective risk aversion is countercyclical, so are risk premia, and so the momentum strategy embodied in a synthetic option has negative market timing. The result then is that while the homogeneous-agent model generates a large options premium, it fails to match data on synthetic options and on delta-hedged option returns.

It is worth stressing that in *all cases*, synthetic options successfully recover the preferences of the average investor even if no investor can actually build the synthetic trading strategy – i.e., cost of trading for any investor after the period they are born is infinite.

Table 3: Model results

	Heterogeneous		Homogeneous	
	Early	Late	Early	Late
Equity SR	0.65	0.65	0.67	0.67
Traded option SR	-0.22	0.00	-0.22	-0.22
Synthetic option SR	0.00	0.00	-0.28	-0.28
Traded option IR	-0.29	-0.06	-0.29	-0.29
Synthetic option IR	-0.07	-0.07	-0.35	-0.35
Delta-hedged option IR	-0.37	0.01	0.09	0.09
Corr(synthetic,traded)	0.81	0.81	0.77	0.77

Note: Table reports various statistics from the calibrated model. SR is the Sharpe ratio and IR is the information ratio.

4.3 Empirical implications and tests of the model

4.3.1 Dealer gamma exposure and option returns

The analysis above takes price in the frictional period to be the greatest willingness to pay of any agent. In the real world that would be the market clearing price if there was a very small set of agents who could sell options – i.e. with risk-bearing capacity far smaller than the demand of retail investors.

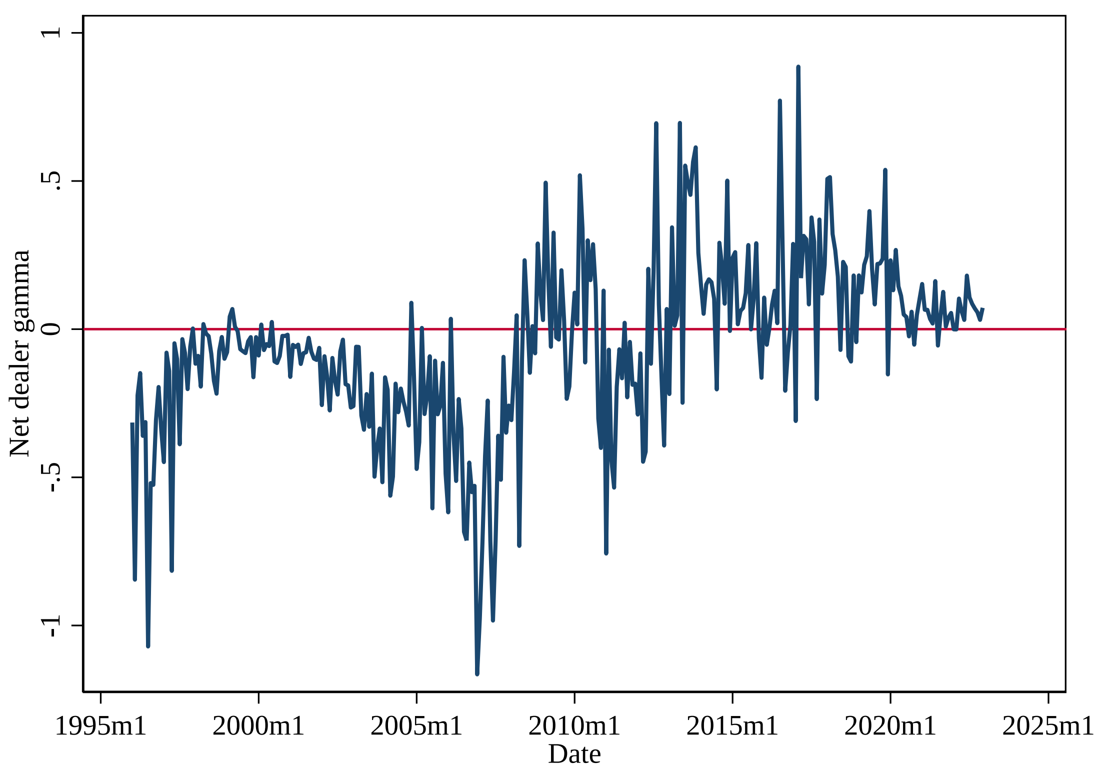
That is often how intermediaries are modeled – they are small and take exogenous demand from retail investors as given, for example in Garleanu, Pedersen, and Poteshman (GPP; 2009).[28] A core feature of GPP's model is that the correct measure of the risk that

[28] For example, GPP include a motivating quote from a derivatives trader, "The number of players in the skew market is limited. ... there's a huge imbalance between what clients want and what professionals can provide."

intermediaries must bear is measured by their unhedgeable risk. Up to second order, that is measured by gamma. The size of the option premium then should be related to dealer gamma.

In the context of the previous section's model, that gamma should have declined to zero once retail investors are able to both buy and sell options, since dealers no longer need to bear asymmetric demand. **This section tests that hypothesis – has dealer gamma exposure gone from negative to zero?** It also tests the core mechanism in the model – that the price of gamma risk, i.e. the difference in the risk premium of traded and synthetic options, should have declined at the same time as gamma exposure went to zero.

Figure 9: Net dealer gamma over time



Note: Figure plots the net gamma of intermediaries over time, based on CBOE open-close and Optionmetrics data.

Figure 9 plots net dealer purchases and sales of gamma in S&P 500 options over the period

1996-2022, constructed using the CBOE open-close data.[29] Dealer gamma was significantly negative until around the Global Financial Crisis, which has been noted widely in previous work. However, since the Global Financial Crisis, net dealer gamma has trended to zero or even positive. The timing of that switch is highly similar to the timing of the shift in the premium on options.

Table 4: Dealer gamma and option risk premia

	Traded	Synthetic	Difference
Market maker net gamma	0.78	-0.28	0.98***
	(0.69)	(0.66)	(0.25)
Spot-future basis volatility	-1.22*	0.04	-1.30***
	(0.68)	(0.65)	(0.23)

Note: Table reports the coefficient of a regression of traded options, synthetic options, and their difference on the lagged net gamma of market markers and on the lagged volatility of the spot-future basis. Both are filtered through an exponentially-weighted moving average. Sample is 1987-2022.

To further understand that change over time, table 4 reports the coefficients of a regression of the information ratios onto lagged dealer gamma (exponentially-weighted to remove the very high-frequency fluctuations). To reduce the influence of outliers, the estimation is performed via maximum likelihood with a student-t error distribution, but results are similar using OLS regression. The table shows that the difference between traded options and synthetic options loads positively on the intermediaries' net gamma, as predicted by the model: as intermediaries started bearing less risk over time (gamma became less negative), the delta-hedged alpha became less negative as well (generating a positive coefficient in this regression). This table therefore provides direct evidence in support of the model presented in this section.

[29] This is the same dataset used in the previous literature on option frictions, e.g. GPP, as well as Chen, Joslin and Ni (2019) and Constantinides and Lian (2015). The dataset classifies, for each option, the total daily buy and sell orders by type of entity (customer, firm, and broker dealer). We compute the total gamma bought and sold each day by intermediaries (defined, as in GPP, as entities that are neither customer or firms) by combining this data with the gamma of each option from Optionmetrics. We use options with 10 to 180 days maturity.

4.3.2 Additional potential drivers of option premia

In order to generate further implications of excess retail options demand that we can take to the data, appendix E.3 extends GPP's model to add three additional realistic frictions that dealers may face: unspanned risk, basis risk, and hedging costs (see proposition 7). Unspanned risk is any risk left after discrete hedging with the underlying (e.g. jump risk and unspanned volatility). Basis risk represents the deviation between the hedging instrument – e.g. S&P 500 futures – and the actual underlying index. And hedging costs represent the cost to dealers due to the actual cost of synthesizing a hedge, such as transaction costs or price pressure. Figure A.12 examines how each of these may have changed over time.

The top two panels of figure A.12 show that trading costs, measured both by posted and effective spreads (Roll (1984)) declined as option premia shrunk. So, in addition to dealers bearing less option risk over time, they also pay smaller costs to hedge the risk they do bear.

We measure basis risk empirically from the gap between the level of the S&P 500 index and the futures price. The middle-left panel of figure A.12 plots the three-month rolling standard deviation of that gap. The y-axis is again on a $\log_{10}$ scale. Over time, basis risk has fallen by about an order of magnitude. While there is a large decline early in the sample, similar to trading frictions, basis risk seems to settle at its current level around the early 2000's.

Finally, figure A.12 plots three measures of unhedgeable risk. The first is the standard deviation of the daily delta-hedged options return, which measures the gap between the return on the option and the daily rebalanced hedge. It shows no clear trend, particularly once the 1987 crash is no longer included in the moving average. Second, following Bollerslev et al. (2009), figure A.12 plots the difference between quadratic and bipower variation, which is a measure of realized jump variation. It rose during the 2008 financial crisis, and has been lower subsequently, but again does not have a clear trend. Finally, unhedgeable risk is, more broadly, driven by higher moments in returns, so the bottom-right panel of figure A.12 plots the measure of S&P 500 return skewness developed in Neuberger (2012). Realized skewness has, over time, trended consistently more negative (implied skewness does the same; see the CBOE's SKEW index and Dew-Becker (2022)). So if higher moments drove the options premium, it should have grown instead of shrunk over time. The decline that we observe in the overpricing of options cannot be driven by declines in jump or higher moment risk, as measured here, since none of those sources of risk have declined. Instead, consistent with the model, the decline in the premium can be driven by a reduction in how much of this risk is borne by intermediaries, and in the frictions they face.

Overall, figure A.12 shows that hedging costs and basis risk have declined – indicating a reduction in frictions faced by intermediaries – while unhedgable risk does not appear to have shrunk.

The second row of table 4 shows that information ratios of traded options and of delta-hedged options moved negatively relative to basis risk, measured as the lagged exponentially-weighted average of the standard deviation of the basis pictured above: as the basis on average became less volatile over time, traded and delta-hedged information ratio became less negative, again in line with the predictions of the theory.

4.3.3 Summary

Over time, dealers have borne a smaller amount of unhedgeable S&P 500 risk, as measured by their gamma exposure. That shift correlates well with the decline in the observed premium on options. Additionally, other frictions that they face, including trading costs and basis risk have also shrunk. All of these factors can help explain why the options premium has converged to zero, in light of the predictions of the model.

The model is stark, in that it has just two regimes: no option selling and free selling. What would have allowed dealers to take on less gamma? While some of it may very well be retail investors having an easier time selling options themselves, there is also the rise of alternative investors who can bear that risk. For example, Jurek and Stafford (2015) show that hedge fund returns are highly similar to those from writing S&P 500 puts. As the hedge fund sector has grown over the past three decades, it may have contributed to the supply of options, increasing risk-bearing capacity and reducing premia. Similarly, Calvet et al. (2024) show using micro evidence that increase access to structured products directly decreased option prices. What is important here is not necessarily who exactly is selling options, but rather that the capacity to do so has grown, driving their price down.

5 Conclusion

This paper studies returns on traded and synthetic options on the S&P 500. The key empirical result is that while their returns used to be drastically different – with synthetic options having CAPM alphas near zero for a century while traded options had significantly negative alphas – that has changed: returns on traded options have converged up to those on synthetic options.

Theoretically, the paper gives conditions under which synthetic options can measure the shape of marginal utility with resppect to market returns, and it develops a general equilibrium model that can fit the data. The model's core mechanism is that in the early period when traded options were overpriced, that was due to the fact that while many could buy options, it was hard for many to sell, causing financial intermediaries to bear a large amount of risk and price that risk accordingly. Then, more recently, as constraints relax so that more agents can sell options, the risk is borne by a greater number of agents – not just dealers anymore – and its price falls. Consistent with the model, dealer net gamma has gone from being consistently negative prior to 2012 to being centered around zero for the past decade.

There are two additional implications of the results that we would like to highlight. First, they show that it is important to take nonstationarity in options returns into account in empirical analyses. The results that one obtains for options premia are very sensitive to the sample used, and research using the full sample currently available through 2023 will get answers that are both economically and statistically significantly different from earlier work using data only through the 1990's.

Second, if synthetic options premia are the superior measure of the preferences of the typical investor – with derivatives markets being segmented – that suggests in the future calibrating and estimating structural models not based on the returns on traded options, but on those for synthetic options. The synthetic options have the added benefits of being available over the same period that equity returns are available and being able to be constructed not only for the total stock market, but also for bonds, anomaly portfolios, or any other asset class.

References

Aıt-Sahalia, Yacine and Andrew W Lo, "Nonparametric risk management and implied risk aversion," *Journal of econometrics*, 2000, *94* (1-2), 9–51.

Aït-Sahalia, Yacine, Mustafa Karaman, and Loriano Mancini, "The term structure of equity and variance risk premia," *Journal of Econometrics*, 2020, *219* (2), 204–230.

Alvarez, Fernando and Urban J. Jermann, "Using Asset Prices to Measure the Cost of Business Cycles," *Journal of Political Economy*, 2004, *112(6)*, 1223–1256.

Backus, David, Mikhail Chernov, and Ian Martin, "Disasters implied by equity index options," *The Journal of Finance*, 2011, *66* (6), 1969–2012.

Bakshi, Gurdip and Nikunj Kapadia, "Delta-Hedge Gains and the Negative Market Volatility Risk Premium," *The Review of Financial Studies*, 2003, *16(2)*, 527–566.

Bakshi, Gurdip and Dilip Madan, "Spanning and derivative-security valuation," *The Journal of Financial Economics*, 2000, *55(2)*, 205–238.

Bao, Jack, Jun Pan, and Jiang Wang, "The illiquidity of corporate bonds," *The Journal of Finance*, 2011, *66* (3), 911–946.

Bates, David S, "The market for crash risk," *Journal of Economic Dynamics and Control*, 2008, *32* (7), 2291–2321.

—— , "US stock market crash risk, 1926–2010," *Journal of Financial Economics*, 2012, *105* (2), 229–259.

—— , "Empirical Option Pricing Models," *Annual Review of Financial Economics*, 2022, *14*, 369–389.

Beason, Tyler and David Schreindorfer, "Dissecting the equity premium," 2022.

Bidder, Rhys and Ian Dew-Becker, "Long-Run Risk is the Worst-Case Scenario," *The American Economic Review*, September 2016, *106* (9), 2494–2527.

Black, Fischer and Myron Scholes, "The Pricing of Options and Corporate Liabilities," *The Journal of Political Economy*, 1973, *81* (3), 637–654.

Bollerslev, Tim and Viktor Todorov, "Tails, fears, and risk premia," *Journal of Finance*, 2011, *66(6)*, 2165–2211.

—— , Uta Kretschmer, Christian Pigorsch, and George Tauchen, "A discrete-time model for daily S & P500 returns and realized variations: Jumps and leverage effects," *Journal of Econometrics*, 2009, *150* (2), 151–166.

Broadie, Mark, Mikhail Chernov, and Michael Johannes, "Model Specification and Risk Premia: Evidence from Futures Options," *The Journal of Finance*, 2007, *62(3)*, 1453–1490.

—— , —— , and —— , "Understanding Index Option Returns," *The Review of Financial Studies*, 2009, *22(11)*, 4493–4529.

Brunnermeier, Markus K, Stefan Nagel, and Lasse H Pedersen, "Carry trades and currency crashes," *NBER macroeconomics annual*, 2008, *23* (1), 313–348.

Büchner, Matthias and Bryan Kelly, "A factor model for option returns," *Journal of Financial Economics*, 2022, *143* (3), 1140–1161.

Calvet, Laurent, Claire Celerier, Gordon Liao, and Boris Vallee, "Household Preferences, Security Design, and Volatility Prices," Technical Report 2024.

Campbell, John Y. and John H. Cochrane, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 1999, *107* (2), 205–251.

Carr, Peter and Dilips B. Madan, *Towards a Theory of Volatility Trading*, London: Risk Books,

___ **and Liuren Wu**, "Variance Risk Premiums," *Review of Financial Studies*, 2009, *22(3)*, 1311–1341.

Chambers, Donald R, Matthew Foy, Jeffrey Liebner, and Qin Lu, "Index option returns: Still puzzling," *The Review of Financial Studies*, 2014, *27* (6), 1915–1928.

Chen, Hui, Scott Joslin, and Sophie Xiaoyan Ni, "Demand for crash insurance, intermediary constraints, and risk premia in financial markets," *The Review of Financial Studies*, 2019, *32* (1), 228–265.

Cochrane, John H and Jesus Saa-Requejo, "Beyond arbitrage: Good-deal asset price bounds in incomplete markets," *Journal of political economy*, 2000, *108* (1), 79–119.

Constantinides, George M and Lei Lian, "The supply and demand of S&P 500 put options," Technical Report, National Bureau of Economic Research 2015.

——, **Jens Carsten Jackwerth, and Alexi Savov**, "The puzzle of index option returns," *Review of Asset Pricing Studies*, 2013, *3* (2), 229–257.

Corsi, Fulvio, "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, 2009, *7* (2), 174–196.

Coval, Joshua D. and Tyler Shumway, "Expected Option Returns," *The Journal of Finance*, 2001, *56(3)*, 983–1009.

Cuesdeanu, Horatio and Jens Carsten Jackwerth, "The pricing kernel puzzle: Survey and outlook," *Annals of Finance*, 2018, *14*, 289–329.

Dew-Becker, Ian, "Real-time forward-looking skewness over the business cycle," 2022. Working paper.

—— **and Stefano Giglio**, "Cross-sectional uncertainty and the business cycle: evidence from 40 years of options data," 2020. Working paper.

——, ——, **Anh Le, and Marius Rodriguez**, "The price of variance risk," *Journal of Financial Economics*, 2017, *123* (2), 225 – 250.

Drechsler, Itamar, "Uncertainty, Time-Varying Fear, and Asset Prices," *The Journal of Finance*, 2013, *68* (5), 1843–1889.

—— **and Amir Yaron**, "What's Vol Got to Do with it?," *The Review of Financial Studies*, 2011, *24(1)*, 1–45.

Du, Du, "General equilibrium pricing of options with habit formation and event risks," *Journal of Financial Economics*, 2011, *99*, 400–426.

Frazzini, Andrea and Lasse Heje Pedersen, "Embedded leverage," *The Review of Asset Pricing Studies*, 2022, *12* (1), 1–52.

Gabaix, Xavier, "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *Quarterly Journal of Economics*, 2012, *127(2)*, 645–700.

Garleanu, Nicolae, Lasse Heje Pedersen, and Allen M Poteshman, "Demand-based option pricing," *The Review of Financial Studies*, 2008, *22* (10), 4259–4299.

Haddad, Valentin and Tyler Muir, "Do intermediaries matter for aggregate asset prices?," *The Journal of Finance*, 2021, *76* (6), 2719–2761.

Hall, Robert E, "High discounts and high unemployment," *American Economic Review*, 2017, *107* (2), 305–30.

Han, Bing, "Investor sentiment and option prices," *The Review of Financial Studies*, 2008, *21* (1), 387–414.

Hansen, Lars Peter and Ravi Jagannathan, "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 1991, *99(2)*, 225–262.

___ **and Robert J Hodrick**, "Forward exchange rates as optimal predictors of future spot rates: An econometric analysis," *Journal of Political Economy*, 1980, *88* (5), 829–853.

He, Zhiguo and Arvind Krishnamurthy, "Intermediary asset pricing," *American Economic Review*, 2013, *103* (2), 732–770.

Heston, Steven, Kris Jacobs, and Hyung Joo Kim, "Exploring Risk Premia, Pricing Kernels, and No-ArbitrageRestrictions in Option Pricing Models." Working paper.

Hull, John and Alan White, "Optimal delta hedging for options," *Journal of Banking & Finance*, 2017, *82*, 180–190.

Jackwerth, Jens Carsten, "Recovering risk aversion from option prices and realized returns," *The Review of Financial Studies*, 2000, *13* (2), 433–451.

Jegadeesh, Narasimhan and Sheridan Titman, "Short-horizon return reversals and the bid-ask spread," *Journal of Financial Intermediation*, 1995, *4* (2), 116–132.

Jurek, Jakub W and Erik Stafford, "The cost of capital for alternative investments," *The Journal of Finance*, 2015, *70* (5), 2185–2226.

Kozhan, Roman, Anthony Neuberger, and Paul Schneider, "The skew risk premium in the equity index market," *The Review of Financial Studies*, 2013, *26* (9), 2174–2203.

Lamoureux, Christopher G and William D Lastrapes, "Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities," *The Review of Financial Studies*, 1993, *6* (2), 293–326.

Lewellen, Jonathan and Stefan Nagel, "The Conditional CAPM does not Explain Asset-Pricing Anomalies," *Journal of Financial Economics*, 2006, *82* (2), 289–314.

Longstaff, Francis A, Jun Pan, Lasse H Pedersen, and Kenneth J Singleton, "How sovereign is sovereign credit risk?," *American Economic Journal: Macroeconomics*, 2011, *3* (2), 75–103.

Martin, Ian, "What is the Expected Return on the Market?," *The Quarterly Journal of Economics*, 2017, *132* (1), 367–433.

Muir, Tyler, "Financial crises and risk premia," *The Quarterly Journal of Economics*, 2017, *132* (2), 765–809.

Muravyev, Dmitriy and Xuechuan Charles Ni, "Why do option returns change sign from day to night?," *Journal of Financial Economics*, 2020, *136* (1), 219–238.

Nagel, Stefan, "Evaporating liquidity," *The Review of Financial Studies*, 2012, *25* (7), 2005–2039.

Neuberger, Anthony, "Realized skewness," *The Review of Financial Studies*, 2012, *25* (11), 3423–3455.

Paoli, Bianca De and Pawel Zabczyk, "Cyclical risk aversion, precautionary saving, and monetary policy," *Journal of Money, Credit and Banking*, 2013, *45* (1), 1–36.

Roll, Richard, "A simple implicit measure of the effective bid-ask spread in an efficient market," *The Journal of finance*, 1984, *39* (4), 1127–1139.

Rosenberg, Joshua V and Robert F Engle, "Empirical pricing kernels," *Journal of Financial Economics*, 2002, *64* (3), 341–372.

Schreindorfer, David, "Macroeconomic tail risks and asset prices," *The Review of Financial Studies*, 2020, *33* (8), 3541–3582.

___ **and Tobias Sichert**, "Volatility and the Pricing Kernel." Working paper.

Seo, Sang Byung and Jessica A Wachter, "Option prices in a model with stochastic disaster risk," *Management Science*, 2019, *65* (8), 3449–3469.

A Proofs from section 2

A.1 Proposition 2

Consider a regression of $M_{t,t+j}$ on $R^m_{t,t+j}$ and $R^{O\perp}_{t,t+j}$. Since $R^m_{t,t+j}$ and $R^{O\perp}_{t,t+j}$ are, by construction, conditionally uncorrelated with each other, we have

$$M_{t,t+j} = \text{const.} + \frac{\text{cov}_t\left(R^m_{t,t+j}, M_{t,t+j}\right)}{\text{var}_t\left[R^m_{t,t+j}\right]} R^m_{t,t+j} + \frac{\text{cov}_t\left(R^{O\perp}_{t,t+j}, M_{t,t+j}\right)}{\text{var}_t\left(R^{O\perp}_{t,t+j}\right)} R^{O\perp}_{t,t+j} + \text{resid.} \tag{26}$$

Additionally, $R^m_{t,t+j}$ is conditionally uncorrelated with $\hat{M}_{t,t+j}$ by construction, so that

$$\text{cov}\left(R^m_{t,t+j}, M_{t,t+j}\right) = \text{cov}\left(R^m_{t,t+j}, \bar{M}_{t,t+j}\right) \tag{27}$$

The same fact works for $R^{O\perp}_{t,t+j}$ since, conditional on date-t information, $R^{O\perp}_{t,t+j}$ is a (nonlinear) function of $R^m_{t,t+j}$. We then can replace $M_{t,t+j}$ on the left-hand side above with $\bar{M}_{t,t+j}$.

Under the assumption that $R^m_{t,t+j}$ and $R^{O\perp}_{t,t+j}$ are priced,

$$\text{cov}\left(R^m_{t,t+j}, M_{t,t+j}\right) = -E_t\left[R^m_{t,t+j} - 1\right] \tag{28}$$

$$\text{cov}\left(R^{O\perp}_{t,t+j}, M_{t,t+j}\right) = -E_t\left[R^{O\perp}_{t,t+j} - 1\right] \tag{29}$$

But since $R^{O\perp}_{t,t+j}$ is uncorrelated with $R^m_{t,t+j}$, its (conditional) CAPM beta is zero by construction, and hence

$$E_t\left[R^{O\perp}_{t,t+j} - 1\right] = \alpha^O_{t,t+j} \tag{30}$$

The result then follows.

A.2 Proposition 3

The proof follows similar lines to that for proposition 2. The only required adjustment is to note that

$$\mathrm{cov}_t\left(R^S_{t,t+j}, M_{t,t+j}\right) = -E_t\left[R^S_{t,t+j} - 1\right] = -\alpha^S_{t,t+j} \tag{31}$$

and

$$\begin{aligned}
\mathrm{cov}_t\left(R^S_{t,t+j}, \bar{M}_{t,t+j}\right) &= \mathrm{cov}_t\left(R^S_{t,t+j}, M_{t,t+j}\right) - \mathrm{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}^S_{t,t+j}\right) &(32)\\
&= -\left(\alpha^S_{t,t+j} + \mathrm{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}^S_{t,t+j}\right)\right) &(33)
\end{aligned}$$

B Synthetic puts and replicating portfolio

B.1 Standard delta-hedging

Consider an option with price P_t, and the underlying with price S_t. The payoff from holding the option from time 1 to T is:

$$\Pi^{option}_{1,T} = P_T - P_1 = \sum_{t=1}^{T-1}(P_{t+1} - P_t)$$

The Black-Scholes-Merton replication portfolio is a dynamic strategy that buys a time-varying number of shares of the underlying, Δ_t, and invests a time-varying amount B_t in the risk free rate. These numbers are chosen so that, in the original setup of Black and Scholes, they guarantee an exact replication of the option value as it evolves over time: equivalently, they guarantee that at maturity the payoff of the option and the replicating portfolio are equal, while cash flows are zero in every period except the first and the last.

To achieve this, Δ_t is chosen as the Black-Scholes delta, and B_t is chosen as the difference between the BS option value P_t and the cost of buying the underlying $\Delta_t S_t$:

$$B_t = P_t - \Delta_t S_t$$

This portfolio costs $B_t + \Delta_t S_t$ to buy at time t, and has a value of $B_t(1 + \frac{r_t}{365}) + \Delta_t S_{t+1}$ at time $t + 1$, where r_t is the annualized interest rate at time t. At time $t + 1$, the strategy requires buying the new portfolio $B_{t+1} + \Delta_{t+1} S_{t+1}$, and so on. In every period between the first and the last one, this strategy generates an *intermediate cash flow* of:

$$ICF_{t+1} = (B_t(1 + \frac{r_t}{365}) + \Delta_t S_{t+1}) - (B_{t+1} + \Delta_{t+1} S_{t+1})$$

and at the last period T, it generates a final cash flow of $B_{T-1}(1 + \frac{r_{T-1}}{365}) + \Delta_{T-1} S_T$ which is equal to the option payoff. In the BS model, the portfolio is self-financing, so $ICF_{t+1} = 0$.

Therefore, the total P&L that this replication portfolio generates is:

$$\Pi_{1,T}^{hedge} = \underbrace{\sum_{t=1}^{T-2} \Big[(B_t(1 + \frac{r_t}{365}) + \Delta_t S_{t+1}) - (B_{t+1} + \Delta_{t+1} S_{t+1})\Big]}_{=0\, under\, BS} + \underbrace{\Big[B_{0T-1}(1 + \frac{r_{T-1}}{365}) + \Delta_{T-1} S_T\Big]}_{final\, cash\, flow} - \underbrace{[B_1 + \Delta_1 S_1]}_{initial\, cost} \tag{34}$$

This can be also rewritten in the following way:

$$\Pi_{1,T}^{hedge} = \sum_{t=1}^{T-1} \Big[(B_t(1 + \frac{r_t}{365}) + \Delta_t S_{t+1}) - (B_t + \Delta_t S_t)\Big] = \sum_{t=1}^{T-1} \Delta_t (S_{t+1} - S_t) + \sum_{t=1}^{T-1} B_t \frac{r_t}{365} \tag{35}$$

which corresponds to the formula of Bakshi and Kapadia (2003) and Buchner and Kelly (2022) when setting $B_t = P_t - \Delta_t S_t$. When the assumptions of Black-Scholes do not hold, the rebalancing of the portfolio generates intermediate cash flows (i.e., CF_t is not zero), which is accounted for by the formula above.

Finally, we examine *excess* returns of the delta-hedged portfolio. Both $\Pi_{1,T}^{option}$ and $\Pi_{1,T}^{hedge}$ are dollar payoffs that correspond to initial investments of P_1 and $B_1 + \Delta_1 S_1$, respectively.

If the objective is to compute delta-hedged returns, then one can compute:

$$\Pi_{1,T}^{option} - \Pi_{1,T}^{hedge} = \sum_{t=1}^{T-1}(P_{t+1} - P_t) - \sum_{t=1}^{T-1}\Delta_t(S_{t+1} - S_t) - \sum_{t=1}^{T-1} B_t \frac{r_t}{365} \tag{36}$$

and this is an excess return: the cost of buying the option is P_1, the income from shorting the hedge portfolio is $B_1 + \Delta_1 S_1 = (P_1 - \Delta_1 S_1) + \Delta_1 S_1 = P_1$. This is because the hedge portfolio is designed to borrow exactly an amount B_1 that fully matches the price of the option P_1 (which, in this case, is observed). The hedge strategy updates B_t over time by always setting $B_t = P_t - \Delta_t S_t$, using the observed price of the option P_t at each point in time. Therefore, the delta-hedging strategy described above (eq. 36) is an excess return whether P_t conforms or not to the BS prices. The only difference is, if the BS model is correct, then the ICF_t and the last period cash flow will all be zero.

B.2 Synthetic options: P&L of zero-cost strategy

Next, we consider the case in which we do not observe the option price P_t. In that case, we cannot use it as an input for computing B_t and Δ_t. We also cannot obviously compute $\Pi_{1,T}^{option}$. However, we note that equation (34) still describes the total P&L of *any* dynamic trading strategy that at each point in time buys Δ_t units of the underlying and invests B_t in the risk free rate, whether or not those are chosen as per the BS model. Therefore, we can still compute $\Pi_{1,T}^{hedge}$ for a choice of P_t and Δ_t[30] (and hence B_t). In particular, we determine the Black-Scholes price of an option, P_t, using as input the current underlying S_t and an estimated value for σ_t^2, and then build the hedging portfolio for that idealized option. The fact that that option is not directly tradable is irrelevant, in the sense that the P&L we build as described above is an actual return of a portfolio that is just a dynamic portfolio of the market.

[30] Δ_t corresponds to $\delta_{t,t+j}^S$ in the main text. To evaluate Δ_t, we follow Hull and White (2017); specifically, we use equation (5) of the referenced paper and select $a = -0.25, b = -0.4, c = -0.5$.

While the delta-hedged P&L, $\Pi_{1,T}^{option} - \Pi_{1,T}^{hedge}$, is, as described above, the P&L of a zero-cost portfolio, $\Pi_{1,T}^{hedge}$ is not. So we next describe the P&L of hedge portfolios that yield $\Pi_{1,T}^{option}$ and $\Pi_{1,T}^{hedge}$ separately but are funded at the risk-free rate at inception. For the option (i.e. when we do observe P_t), we have:

$$\Pi_{1,T}^{option,exc} = P_T - P_1(1+\frac{r_1}{365})^T$$

Funding the hedge portfolio requires borrowing $B_1+\Delta_1 S_1 = P_t$, where P_t is the theoretical BS price of an option. So the total P&L can be written as:

$$\Pi_{1,T}^{hedge,exc} = \underbrace{\sum_{t=1}^{T-2}\left[(B_t(1+\frac{r_t}{365}) + \Delta_t S_{t+1}) - (B_{t+1}+\Delta_{t+1}S_{t+1})\right]}_{=0\,under\,BS} + \left[B_{T-1}(1+\frac{r_{T-1}}{365}) + \Delta_{T-1}S_T\right] - \left[B_1+\Delta_1 S_1\right](1+\frac{r_1}{365})^T$$

Note that in these formulas the intermediate cash flows are assumed not to be reinvested. One can also reinvest them to obtain:

$$\Pi_{1,T}^{hedge,exc} = \sum_{t=1}^{T-2}(1+r_{t+1})^{T-t-1}\left[(B_t(1+\frac{r_t}{365}) + \Delta_t S_{t+1}) - (B_{t+1}+\Delta_{t+1}S_{t+1})\right]$$

$$+\left[B_{T-1}(1+\frac{r_{T-1}}{365}) + \Delta_{T-1}S_T\right] - \left[B_1+\Delta_1 S_1\right](1+\frac{r_1}{365})^T$$

An alternative procedure is to get the same exposure Δ_t every period, but entirely fund it at the risk-free rate every period. The P&L of this zero-cost portfolio is:

$$\tilde{\Pi}_{1,T}^{hedge,exc} = \sum_{t=1}^{T-1}\Delta_t(S_{t+1} - S_t(1+\frac{r_t}{365})) = \sum_{t=1}^{T-1}\Delta_t(S_{t+1}-S_t) - \sum_{t=1}^{T-1}\Delta_t S_t\frac{r_t}{365}$$

Note that this relates closely to $\Pi_{1,T}^{hedge,exc}$, since

$$\Pi_{1,T}^{hedge,exc} = \sum_{t=1}^{T-1}\Delta_t(S_{t+1}-S_t) + \sum_{t=1}^{T-1}(P_t - \Delta_t S_t)\frac{r_t}{365} - P_1[(1+\frac{r_1}{365})^T - 1]$$

where the latter term is the total interest paid on the original loan of P_1. So we can write:

$$\Pi_{1,T}^{hedge,exc} - \tilde{\Pi}_{1,T}^{hedge,exc} = \sum_{t=1}^{T-1} P_t \frac{r_t}{365} - P_1[(1 + \frac{r_1}{365})^T - 1]$$

The difference is effectively only coming from the different timing of the borrowing (every period vs. at the beginning of the month), and is unlikely to make any substantial difference empirically.

In fact, we can also consider funding the original hedging strategy, $\Pi_{1,T}^{hedge}$ day by day instead of once at the very beginning. Modifying eq. (35):

$$\hat{\Pi}_{1,T}^{hedge} = \sum_{t=1}^{T-1} \left[(B_t(1 + \frac{r_t}{365}) + \Delta_t S_{t+1}) - (B_t + \Delta_t S_t)(1 + \frac{r_t}{365}) \right] = \sum_{t=1}^{T-1} \Delta_t (S_{t+1} - S_t(1 + \frac{r_t}{365}))$$

So that:

$$\hat{\Pi}_{1,T}^{hedge,exc} = \tilde{\Pi}_{1,T}^{hedge,exc} \simeq \Pi_{1,T}^{hedge,exc}$$

B.3 P&L and returns

The P&Ls described above ($\tilde{\Pi}_{1,T}^{hedge,exc}$ and $\Pi_{1,T}^{hedge,exc}$) correspond to strategies that have zero cost. Therefore, they also represent excess returns. Scaling that excess return by any time-1 quantity is also an excess return. Just like Buchner and Kelly (2022), we scale P&Ls by the underlying at time 1:

$$R_{1,T}^{hedge,exc} = \frac{\Pi_{1,T}^{hedge,exc}}{S_1}$$

and

$$\tilde{R}_{1,T}^{hedge,exc} = \frac{\tilde{\Pi}_{1,T}^{hedge,exc}}{S_1}$$

Finally, we consider another related zero-cost trading strategy. Instead of scaling by S_1, we scale the position of the strategy that funds every day ($\tilde{\Pi}^{hedge}_{1,T}$) by S_t every day:

$$\tilde{R}^{hedge,scaled}_{1,T} = \sum_{t=1}^{T-1} \Delta_t \frac{S_{t+1} - S_t(1 + \frac{r_t}{365})}{S_t}$$

Defining $R^M_{t+1} = \frac{S_{t+1}-S_t}{S_t}$ we obtain:

$$\frac{S_{t+1} - S_t(1 + \frac{r_t}{365})}{S_t} = \frac{S_{t+1} - S_t}{S_t} - (1 + \frac{r_t}{365}) = R^M_{t+1} - R^f_{t+1}$$

and therefore

$$\tilde{R}^{hedge,scaled}_{1,T} = \sum_{t=1}^{T-1} \Delta_t (R^M_{t+1} - R^f_{t+1})$$

A final point concerns dividends. While dividends make a small difference over short time horizons, we can incorporate them easily in our analysis since we are not trying to hedge a traded option. In other words, we consider a synthetic option that aims to hedge a value S_t that tracks the value of an investment in the underlying that reinvests all the dividends. In that case, R^M is the one-day gross return (including dividends) of the market.

B.4 Comparison of the different approaches

In this section, we compare our baseline excess returns ($R^{hedge,exc}_{1,T}$, with reinvested intermediate cash flow) to the one obtained by funding the position each day, $\tilde{R}^{hedge,scaled}_{1,T}$. The table below reports, for different combinations of maturity M and strike K, the correlation between $R^{hedge,exc}_{1,T}$ and $\tilde{R}^{hedge,scaled}_{1,T}$ and the information ratio of each of them.

C Robust confidence bands for alpha

To combine statistical uncertainty with uncertainty from $\text{cov}_t\left(\hat{R}^S_{t,t+j}, \hat{M}_{t,t+j}\right)$, we treat them as two independent sources of error. Specifically, suppose one starts with a diffuse prior for

$$\alpha^{S,adjusted}_{t,t+j} \equiv \alpha^S_{t,t+j} + \text{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}^S_{t,t+j}\right) \tag{37}$$

Denote by $\alpha^{S,estimated}_{t,t+j}$ the empirical estimate. Asymptotically, $\alpha^{S,estimated}_{t,t+j} \sim N\left(\alpha^S_{t,t+j}, SE^2\right)$, where SE is the standard error for the estimate. We also treat the second term as though it is drawn from the distribution,

$$\text{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}^S_{t,t+j}\right) \sim N\left(0, \left(\frac{1}{2} \times 0.5 \times std_t\left(\hat{R}^S_{t,t+j}\right) \frac{E\left[R^m_{t,t+j} - 1\right]}{std\left(R^m_{t,t+j}\right)}\right)^2\right) \tag{38}$$

Recall from the main text that we take $0.5 \times std_t\left(\hat{R}^S_{t,t+j}\right) \frac{E\left[R^m_{t,t+j}-1\right]}{std\left(R^m_{t,t+j}\right)}$ as an upper bound for $\text{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}^S_{t,t+j}\right)$. To incorporate that with the estimation uncertainty, we treat $\text{cov}_t\left(\hat{M}_{t,t+j}, \hat{R}^S_{t,t+j}\right)$ as though it has a standard deviation of $0.5 \times std_t\left(\hat{R}^S_{t,t+j}\right) \frac{E\left[R^m_{t,t+j}-1\right]}{std\left(R^m_{t,t+j}\right)}$, so that the upper bound is two standard deviations from the prior mean – i.e. at the edge of a ± 2 standard deviation interval.

Given those two assumptions along with the diffuse prior, we then have

$$\alpha^{S,adjusted}_{t,t+j} \sim N\left(\alpha^{S,estimated}_{t,t+j}, SE^2 + \left(\frac{1}{2} \times 0.5 \times std_t\left(\hat{R}^S_{t,t+j}\right) \frac{E\left[R^m_{t,t+j} - 1\right]}{std\left(R^m_{t,t+j}\right)}\right)^2\right) \tag{39}$$

D Effects of daily mispricing

Suppose that the "true" market return that satisfies $1 = E_t\left[M_{t,t+1} R^m_{t,t+1}\right]$ every day, $R^m_{t,t+1}$, is unobservable and instead we can only see some contaminated return $R^{m*}_{t,t+1}$, with

$$R^{m*}_{t,t+1} = R^m_{t,t+1} + \varepsilon_{t+1} \tag{40}$$

where ε_{t+1} is the contamination. Depending on the properties of ε_{t+1}, it may be that $1 \neq E_t\left[M_{t,t+1}R^{m*}_{t,t+1}\right]$ even if $R^m_{t,t+1}$ is in fact priced each day.

If we define $R^{S*}_{t,t+j}$ to be the return on an option synthesized from the contaminated market return, $R^{m*}_{t,t+1}$, then

$$E\left[R^{S*}_{t,t+j}\right] = E\left[R^{S}_{t,t+j}\right] + \sum_{s=0}^{j-1} E\left[\delta^S_{t+s}\varepsilon_{t+s+1}\right] \tag{41}$$

In order for the contamination, ε_{t+1}, to not affect measured risk premia, it must have zero mean and be uncorrelated with past values of the weights δ^S_t. On the other hand, the errors need not be i.i.d., for example, or necessarily independent of anything else.

As discussed in Bates (2012), observed positive serial correlation in daily market index returns is evidence of stale prices.[31,32] When there is a positive return in the underlying, the hedge weight δ^S_t rises, both for puts and calls. That leads to the result that $\operatorname{cov}\left(\varepsilon_{t+1}, \delta^S_t\right) > 0$ when ε_t is positively serially correlated, which would lead to a positive bias, $E\left[R^S_{t,t+j}\right] > E\left[R^{S*}_{t,t+j}\right]$. That is why the main results use lagged information, so that δ^S_t does not depend on R^{m*}_t.[33]

[31] For example, suppose on each day 50% of stock prices are updated. If there is good news on date t, that will be impounded into half of stock prices on date t, 1/4 on t+1, 1/8 on t+2, etc., inducing positive serial correlation in the "errors" ε_t relative to the "true" market return that would be observed if all stock prices were updated every day.

[32] Another type of error that can arise is bid-ask bounce, which, as discussed in Jegadeesh and Titman (1995), creates negative serial correlation in returns. In contrast to stale prices, that would bias the estimate of synthetic option returns *downward*, implying the true alphas for synthetic options are even more positive and in even stronger conflict with the alphas on traded options than is suggested by the baseline results.

[33] One might also ask about the effect of mispricing on estimated betas. First, as an empirical matter, recall that the data shows that the betas of synthetic and traded options are highly similar, implying that there is not a severe bias. Second, while stale prices and bid-ask bounce can affect betas at high frequencies, those effects tend to shrink at lower frequencies, hence the more common focus on, say, monthly data in studies of equity returns.

E Theoretical results for intermediary model

This section presents our version of the GPP model in more detail. The vast majority of the content is due to them; the only change is the addition of the trading friction and index-futures basis.

E.1 Basic setup

There is a constant gross risk-free rate R_f. The underlying index has an exogenous excess return R^I_{t+1}. We consider a simplified version of the model where there is a single option traded that has some price P_t. Its excess return is then $R^O_{t+1} = P_{t+1} - R_f P_t$.

Dealers/intermediaries maximize discounted utility over consumption C_t,

$$E_t \sum_{j=0}^{\infty} \rho^j \left(-\gamma^{-1}\right) \exp\left(-\gamma C_t\right) \tag{42}$$

subject to a transversality condition and budget constraint, which is

$$W_{t+1} = \left(W_t - C_t\right) R_f + D_t R^O_{t+1} + F_t R^F_{t+1} - \frac{\kappa}{2} F_t^2 \tag{43}$$

where W_t is wealth. The intermediaries optimize over D_t, F_t, and C_t subject to the budget constraint and taking the returns as given.

As described in the text, the term $\frac{\kappa}{2} F_t^2$ is a deviation from GPP, as is the distinction between R^F and R^I.

It is assumed that the futures contract on the underlying that the dealers trade is available in infinite supply. For the options, there is some exogenous demand from outside investors, d_t, and the market clearing condition is $D_t + d_t = 0$.

Lemma 5 *In this model, assets are priced under a probability measure d which is equal to* *the measure P multiplied by the factor $\frac{\exp(-k(W_{t+1}+G(d_{t+1},X_{t+1})))}{E_t[\exp(-k(W_{t+1}+G(d_{t+1},X_{t+1})))]}$. In addition,*

$$\kappa F_t = E_t^d \left[R_{t+1}^F\right] \tag{44}$$

$$P_t = R_f^{-1} E_t^d P_{t+1} \tag{45}$$

where P_t is the price of the option (equivalently, $0 = E_t^d R_{t+1}^O$).

Proof. The value function and budget constraint satisfy

$$V_t = \max_{C_t, D_t, F_t} -\gamma^{-1} \exp\left(-\gamma C_t\right) + \rho E_t V_{t+1} \tag{46}$$

$$W_{t+1} = \left(W_t - C_t\right) R_f + D_t \left(P_{t+1} - R_f P_t\right) + F_t R_{t+1}^F - \frac{\kappa}{2} F_t^2 \tag{47}$$

Now guess that

$$V_t = -k^{-1} \exp\left(-k\left(W_t + G_t\right)\right)$$

for some variable G_t that is exogenous to the dealers, and where

$$k = \gamma \frac{R_f - 1}{R_f} \tag{48}$$

We have

$$\frac{\partial}{\partial W_t} V_t = -k V_t \tag{49}$$

$$\text{and } \frac{dW_{t+1}}{dC_t} = -R_f \tag{50}$$

So then the FOC for consumption under this guess is

$$0 = \exp\left(-\gamma C_t\right) + k R_f \rho E_t V_{t+1} \tag{51}$$

Noting that

$$
\begin{aligned}
V_t &= -\gamma^{-1}\exp\left(-\gamma C_t\right) + \rho E_t V_{t+1} && (52)\\
\rho E_t V_{t+1} &= V_t + \gamma^{-1}\exp\left(-\gamma C_t\right) && (53)
\end{aligned}
$$

We have

$$
\exp\left(-\gamma C_t\right) = \exp\left(-k\left(W_t + G_t\right)\right) \tag{54}
$$

Now consider the FOC with respect to F_t. First,

$$
\frac{dW_{t+1}}{dF_t} = R^F_{t+1} - \kappa F_t \tag{55}
$$

And hence the FOC is

$$
\begin{aligned}
0 &= \rho E_t\left[\exp\left(-k\left(W_{t+1} + G_{t+1}\right)\right)\left(R^F_{t+1} - \kappa F_t\right)\right] && (56)\\
\kappa F_t &= E^d_t\left[R^F_{t+1}\right] && (57)
\end{aligned}
$$

where E^d is the expectation under the risk-neutral measure, which is the physical measure distorted by the factor

$$
\frac{\exp\left(-k\left(W_{t+1} + G_{t+1}\right)\right)}{E_t\left[\exp\left(-k\left(W_{t+1} + G_{t+1}\right)\right)\right]} \tag{58}
$$

Next, for D_t,

$$
\frac{dW_{t+1}}{dD_t} = R^O_{t+1} \tag{59}
$$

So then

$$
\begin{aligned}
0 &= \rho E_t \left[\exp\left(-k\left(W_{t+1}+G_{t+1}\right)\right) R^O_{t+1}\right] && (60)\\
0 &= R_f^{-1} E^d_t R^O_{t+1} && (61)
\end{aligned}
$$

It is straightforward to get a recursion for G_t by following the derivation in GPP. ■

Proposition 6 *The effect of options demand on prices is*

$$
\frac{\partial P_t}{\partial D_t} = -\gamma\left(R_f-1\right) E^d_t\left[\left(R^O_{t+1}-\hat{\beta}_t R^F_{t+1}\right) P_{t+1}\right] \tag{62}
$$

where

$$
\begin{aligned}
\hat{\beta}_t &\equiv \beta^F_t \frac{E^d_t\left[\left(R^F_{t+1}\right)^2\right]}{E^d_t\left[\left(R^F_{t+1}\right)^2\right]+k^{-1}R_f\kappa} && (63)\\
\beta^F_t &\equiv \frac{\operatorname{cov}^d_t\left(R^F_{t+1},R^O_{t+1}\right)}{\operatorname{var}^d_t\left(R^F_{t+1}\right)} && (64)
\end{aligned}
$$

Proof. Based on the analysis from the previous proof, the pricing kernel can be written as

$$
m^d_{t+1} \equiv \frac{\exp\left(-k\left(F_tR^F_{t+1}+D_tR^O_{t+1}+G_{t+1}\right)\right)}{R_fE_t\exp\left(-k\left(F_tR^F_{t+1}+D_tR^O_{t+1}+G_{t+1}\right)\right)} \tag{65}
$$

Differentiate m^d_{t+1} with respect to D_t to get

$$
\begin{aligned}
\frac{\partial m^d_{t+1}}{\partial D_t} &= \frac{-k\left(R^O_{t+1}+R^F_{t+1}\frac{\partial F_t}{\partial D_t}\right)\exp\left(-k\left(F_tR^F_{t+1}+D_tR^O_{t+1}+G_{t+1}\right)\right)}{R_fE_t\exp\left(-k\left(F_tR^F_{t+1}+D_tR^O_{t+1}+G_{t+1}\right)\right)} && (66)\\
&\quad -\frac{\exp\left(-k\left(F_tR^F_{t+1}+D_tR^O_{t+1}+G_{t+1}\right)\right)}{\left(R_fE_t\exp\left(-k\left(F_tR^F_{t+1}+D_tR^O_{t+1}+G_{t+1}\right)\right)\right)^2} && (67)\\
&\quad \times E_t\left[-kR_fP_{t+1}\exp\left(-k\left(F_tR^F_{t+1}+D_tR^O_{t+1}+G_{t+1}\right)\right)\right] && (68)
\end{aligned}
$$

$$= -k\left(R^O_{t+1} + R^F_{t+1}\frac{\partial F_t}{\partial D_t}\right) m^d_{t+1} - \frac{\exp\left(-k\left(F_t R^F_{t+1} + D_t R^O_{t+1} + G_{t+1}\right)\right)}{R_f E_t \exp\left(-k\left(F_t R^F_{t+1} + D_t R^O_{t+1} + G_{t+1}\right)\right)} \tag{69}$$

$$\times E_t\left[-kR_f R^O_{t+1}\frac{\exp\left(-k\left(F_t R^F_{t+1} + D_t R^O_{t+1} + G_{t+1}\right)\right)}{R_f E_t \exp\left(-k\left(F_t R^F_{t+1} + D_t R^O_{t+1} + G_{t+1}\right)\right)}\right] \tag{70}$$

$$= -k\left(R^O_{t+1} + R^F_{t+1}\frac{\partial F_t}{\partial D_t}\right) m^d_{t+1} \tag{71}$$

Next, we differentiate the first-order condition for F_t with respect to D_t,

$$\kappa\frac{\partial F_t}{\partial D_t} = E_t\left[\frac{\partial}{\partial D_t} m^d_{t+1} R^F_{t+1}\right] \tag{72}$$

$$= E_t\left[-k\left(R^O_{t+1} + R^F_{t+1}\frac{\partial F_t}{\partial D_t}\right) R^F_{t+1} m^d_{t+1}\right] \tag{73}$$

Solving for the derivative yields

$$k^{-1}\kappa\frac{\partial F_t}{\partial D_t} = E_t\left[-R^O_{t+1}R^F_{t+1}m^d_{t+1} - R^F_{t+1}\frac{\partial F_t}{\partial D_t}R^F_{t+1}m^d_{t+1}\right] \tag{74}$$

$$k^{-1}\kappa\frac{\partial F_t}{\partial D_t} + \frac{\partial F_t}{\partial D_t}R_f^{-1}E^d_t\left[\left(R^F_{t+1}\right)^2\right] = R_f^{-1}E^d_t\left[-R^O_{t+1}R^F_{t+1}\right] \tag{75}$$

$$\frac{\partial F_t}{\partial D_t} = -\frac{E^d_t\left[R^O_{t+1}R^F_{t+1}\right]}{E^d_t\left[\left(R^F_{t+1}\right)^2\right] + k^{-1}R_f\kappa} \equiv -\beta^F_t\frac{E^d_t\left[\left(R^F_{t+1}\right)^2\right]}{E^d_t\left[\left(R^F_{t+1}\right)^2\right] + k^{-1}R_f\kappa} \tag{76}$$

where

$$\beta^F_t \equiv \frac{E^d_t\left[R^O_{t+1}R^F_{t+1}\right]}{E^d_t\left[\left(R^F_{t+1}\right)^2\right]} = \frac{\text{cov}^d_t\left(R^F_{t+1}, R^O_{t+1}\right)}{\text{var}^d_t\left(R^F_{t+1}\right)} \tag{77}$$

The price sensitivity comes from differentiating the pricing equation for the option

$$\begin{aligned}
\frac{\partial P_t}{\partial D_t} &= E_t\left[\frac{\partial m^d_{t+1}}{\partial D_t}P_{t+1}\right] \\
&= -kE_t\left[\left(R^O_{t+1} + R^F_{t+1}\frac{\partial F_t}{\partial D_t}\right) m^d_{t+1}P_{t+1}\right] \\
&= -kE_t\left[\left(R^O_{t+1} - \hat{\beta}_t R^F_{t+1}\right) m^d_{t+1}P_{t+1}\right] \\
&= -\gamma\left(R_f - 1\right)E^d_t\left[\left(R^O_{t+1} - \hat{\beta}_t R^F_{t+1}\right) R^O_{t+1}\right]
\end{aligned}$$

where the last line uses the fact that $E^d_t\left[R^O_{t+1}\right] = E^d_t\left[R^F_{t+1}\right] = 0$ since they are excess returns that are fairly priced under the pricing measure d. ■

E.2 Proof of proposition 7

The proof involves simply analyzing the expectation in 6 above. We have

$$\begin{aligned}
E^d_t\left[\left(R^O_{t+1} - \hat{\beta}_t R^F_{t+1}\right) R^O_{t+1}\right] &= E^d_t\left[\left(R^O_{t+1} - \beta^F_t R^F_{t+1} - \left(\hat{\beta}_t - \beta^F_t\right) R^F_{t+1}\right) R^O_{t+1}\right] && (78)\\
&= E^d_t\left[\left(\varepsilon^F_{t+1} - \left(\hat{\beta}_t - \beta^F_t\right) R^F_{t+1}\right)\left(\beta^F_t R^F_{t+1} + \varepsilon^F_{t+1}\right)\right] && (79)\\
&= \mathrm{var}^d_t\left[\varepsilon^F_{t+1}\right] - \left(\hat{\beta}_t - \beta^F_t\right)\beta^F_t E^d_t\left[\left(R^F_{t+1}\right)^2\right] && (80)\\
&= \mathrm{var}^d_t\left[\varepsilon^F_{t+1}\right] - \left(\frac{-k^{-1}R_f\kappa}{E^d_t\left[\left(R^F_{t+1}\right)^2\right] + k^{-1}R_f\kappa}\right)\left(\beta^F_t\right)^2 E^d_t\left[\left(R^F_{t+1}\right)^2\right] && (81)
\end{aligned}$$

Next, we want to further decompose $\mathrm{var}^d_t\left[\varepsilon^F_{t+1}\right]$. We have

$$\begin{aligned}
R^F_{t+1} &= R^I_{t+1} + z_{t+1} && (82)\\
\beta^F_t &= \beta^I_t \frac{\sigma^2_{I,t}}{\sigma^2_{I,t} + \sigma^2_{z,t}} && (83)
\end{aligned}$$

where $\sigma^2_{I,t} = \mathrm{var}^d_t\left(R^I_{t+1}\right)$. We can write

$$R^O_{t+1} = \beta^I_t R^I_{t+1} + \varepsilon^I_{t+1} \qquad (84)$$

where β^I_t is the (d-measure) regression coefficient. Then

$$\begin{aligned}
\varepsilon^F_{t+1} &= R^O_{t+1} - \beta^F_t R^F_{t+1} && (85)\\
&= \beta^I_t R^I_{t+1} + \varepsilon^I_{t+1} - \beta^I_t \frac{\sigma^2_{I,t}}{\sigma^2_{I,t} + \sigma^2_{z,t}}\left(R^I_{t+1} + z_{t+1}\right) && (86)\\
&= \beta^I_t \frac{\sigma^2_{z,t}}{\sigma^2_{I,t} + \sigma^2_{z,t}} R^I_{t+1} + \varepsilon^I_{t+1} - \beta^I_t \frac{\sigma^2_{I,t}}{\sigma^2_{I,t} + \sigma^2_{z,t}} z_{t+1} && (87)
\end{aligned}$$

$$\mathrm{var}_t^d\left[\varepsilon_{t+1}^F\right] = \left(\beta_t^I\right)^2 \frac{\sigma_{z,t}^2\sigma_{I,t}^2}{\sigma_{I,t}^2+\sigma_{z,t}^2} + \sigma_{\varepsilon,t}^2 \tag{88}$$

where $\sigma_{\varepsilon,t}^2 \equiv \mathrm{var}_t^d\left[\varepsilon_{t+1}^I\right]$.

Up to first order in κ and σ_z^2,

$$\frac{\partial P_t}{\partial D_t} = -\gamma\left(R_f-1\right)\left(\underbrace{\left(\beta_t^I\right)^2\sigma_{z,t}^2}_{\text{Basis risk}} + \underbrace{\sigma_\varepsilon^2}_{\text{Unhedgeable risk}}\right) + \underbrace{\kappa R_f^2\left(\beta_t^I\right)^2}_{\text{Imperfect hedging}} \tag{89}$$

E.3 GPP model

This section studies a simple extension of the model of Garleanu, Pedersen, and Poteshman (GPP; 2009) to help clarify how various frictions can affect option prices when markets are segmented. The only addition to their framework is to allow for transaction costs and index-futures basis risk. The main text describes just the key parts of the setup and predictions of the model. The details are in appendix E.

E.3.1 Setup

In GPP's model, the price of the underlying – which we take here to be the market index – is determined exogenously (presumably by a much larger mass of traders, for example retail investors optimizing between index funds and cash), as is the demand or supply for a set of derivative claims, such as options. GPP show empirically that in general retail investors appear to be long index options, so that dealers must be net short.[34]

The dealers are assumed to have time-additive CARA preferences over consumption with risk aversion γ. The key equation in the model is the dynamic budget constraint,

$$W_{t+1} = \left(W_t - C_t\right)R_f + D_t R_{t+1}^O + F_t R_{t+1}^F - \frac{\kappa}{2}F_t^2 \tag{90}$$

[34] A simple model for the source of asymmetry in retail demand is that it is easier in practice to buy than sell options – buying options does not require posting margin, it has limited downside risk, and it can be done in both regular and tax-protected accounts. So if retail investors are roughly split between those who would to buy and sell, the net demand can still be positive due to asymmetry in frictions. And as those frictions decline, so should the asymmetry.

W_t is wealth and C_t consumption. The risk-free rate, R_f is constant for simplicity, R^F_{t+1} is the gross return on index futures, and R^O_{t+1} is the return on the derivatives. The dealers endogenously choose consumption and the allocations to derivatives and futures D_t and F_t, respectively.

We add two frictions: a quadratic trading cost, $\frac{\kappa}{2}F_t^2$, and a wedge between the futures return, R^F_{t+1}, and the underlying index return, R^I_{t+1},

$$R^F_{t+1} = R^I_{t+1} + z_{t+1} \tag{91}$$

z_{t+1} represents *basis risk*. Ideally the dealers would like to hedge the options they trade with the underlying, like the S&P 500. But the S&P 500 is not itself directly tradable (except at significant cost by buying 500 stocks). Instead, dealers must buy futures (or ETFs or other instruments) whose price is not guaranteed to perfectly track the index. z_{t+1} captures the risk associated with imperfect tracking.[35]

While the dealers choose D_t, markets must clear, meaning that in equilibrium their choice of D_t must perfectly offset the (exogenous) demand from retail investors. The core idea in GPP is to understand how derivative prices, denoted by P_t, vary with quantities, D_t.

E.3.2 Predictions

In the model, intermediaries hedge their options each period with a position in the underlying – it can be thought of as a delta hedge that is adjusted each period. The optimal position, in the absence of any frictions, is denoted by β^I_t (which is simply the local sensitivity of option returns to the underlying index). The unhedgeable risk is defined as

$$\sigma^2_{\varepsilon,t} \equiv \mathrm{var}^d_t\left(R^O_{t+1} - \beta^I_t R^I_{t+1}\right) \tag{92}$$

[35] The S&P 500 index is the underlying for CBOE options, but not CME futures options. For the futures options, an interpretation of basis risk would be that intermediaries price options based on a model for the underlying, meaning that deviations of the futures price from the index create risk.

where var_t^d is a variance taken under the intermediaries' pricing measure d based on date-t information.

The model's key prediction is for the sensitivity of option prices, P_t^O, to demand:

Proposition 7 *Up to first order in the transaction cost κ and the index-futures basis risk $\text{var}_t^d(z_{t+1})$,*

$$\frac{\partial P_t^O}{\partial D_t} = -\gamma\left(R_f - 1\right)\left(\underbrace{\sigma_{\varepsilon,t}^2}_{\text{Unhedgeable risk}} + \underbrace{\left(\beta_t^I\right)^2 \text{var}_t^d\left(z_{t+1}\right)}_{\text{Basis risk}}\right) + \underbrace{\kappa R_f^2\left(\beta_t^I\right)^2}_{\text{Imperfect hedging}} \tag{93}$$

The sensitivity is proportional to risk aversion, γ, and has three terms.

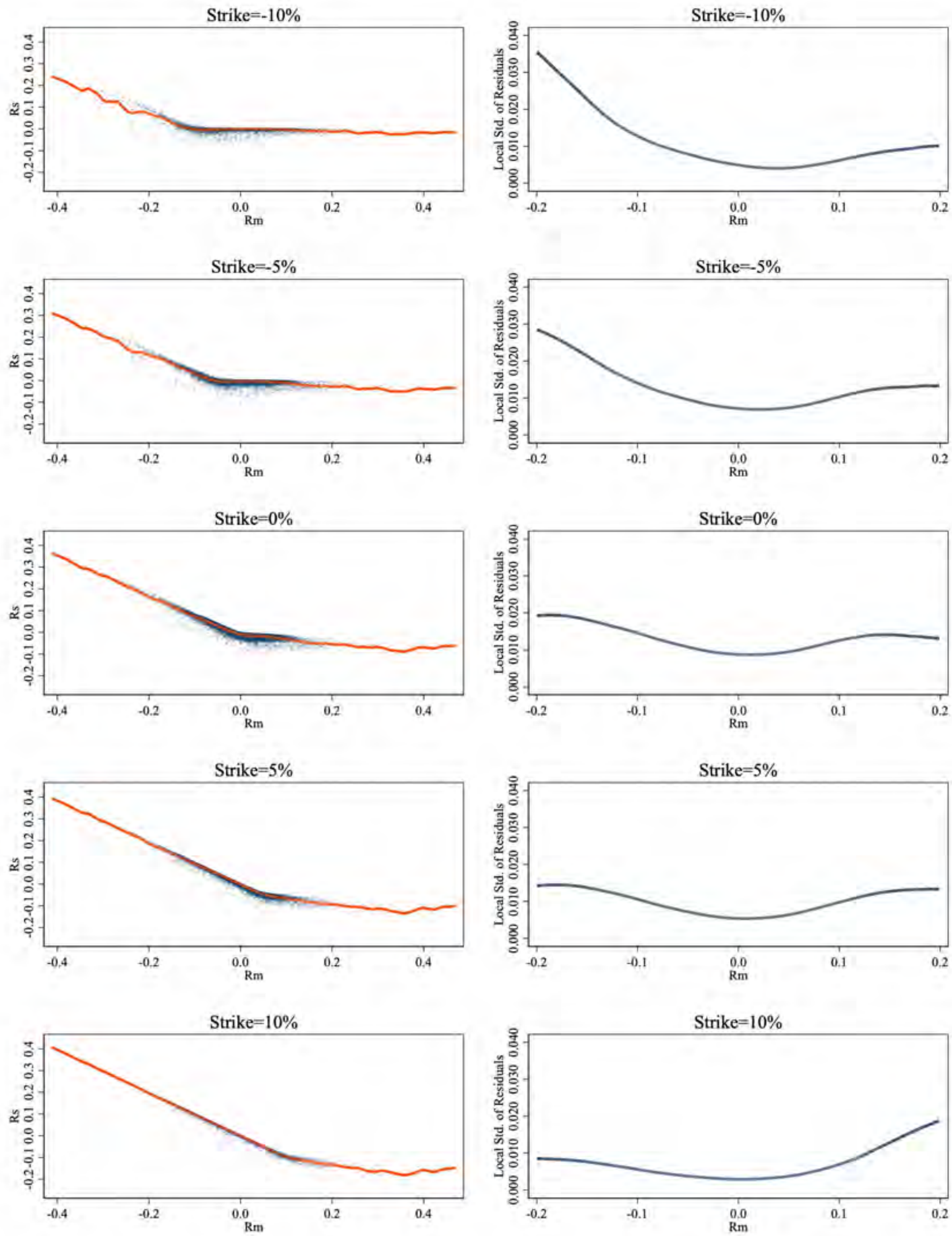
The first component, $\sigma_{\varepsilon,t}^2$, is the unhedgeable risk from (92). Dealers hedge by taking positions exposed to the underlying, but since options have nonlinear exposure, that hedge is inevitably imperfect, due to discrete hedging, jumps, and unspanned volatility. The synthetic options studied in our empirical analysis exactly map into the hedge that the dealers use here – they are updated discretely and inherit risk from deviations between the discrete replication and the traded option return.

The second term represents basis risk. When there are larger random gaps between the hedging instrument and the true underlying index, dealers face greater risk and thus demand larger premia. Finally, the third term arises due to the quadratic trading cost, κ, which causes dealers to hedge incompletely, further raising their risk from holding derivatives.

In the context of the general theoretical analysis in section 2, this is a model in which traded options are not priced by the marginal utility of retail equity investors, but instead by that of dealers. And the exogenous option demand, since it must be borne only by dealers, drives option prices up, creating negative CAPM alphas.

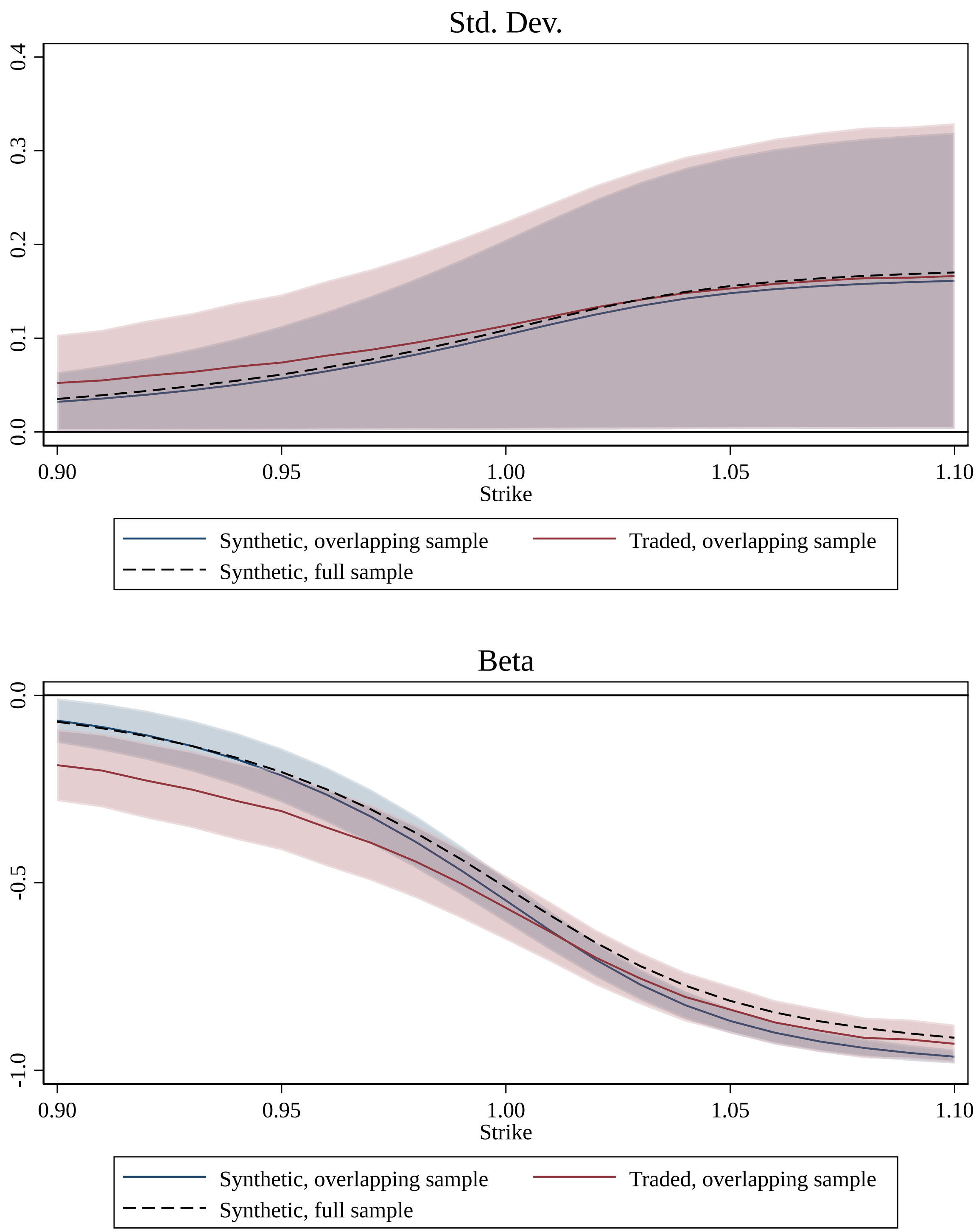
F Additional figures

Figure A.1: Synthetic put returns for various strikes



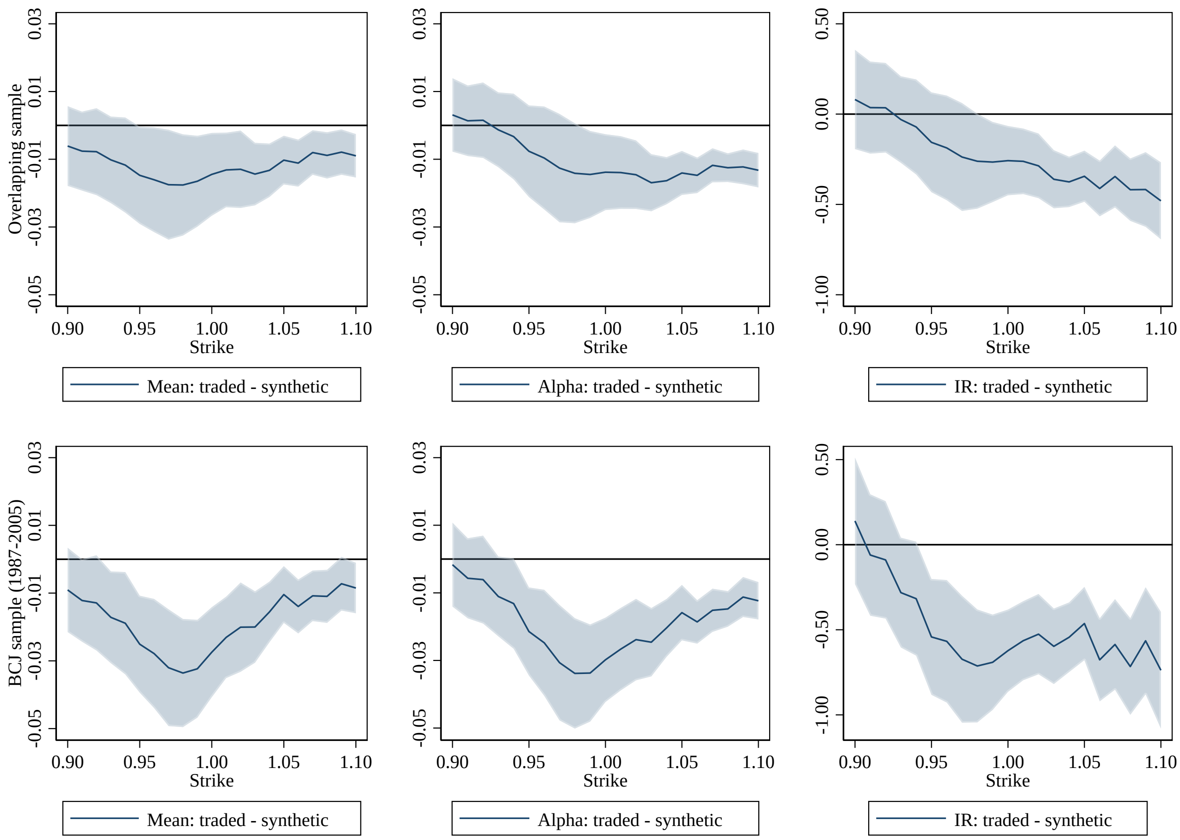
Note: these figures replicate figure 3 varying the strike used for the synthetic option.

Figure A.2: Risk measures for synthetic and traded options.



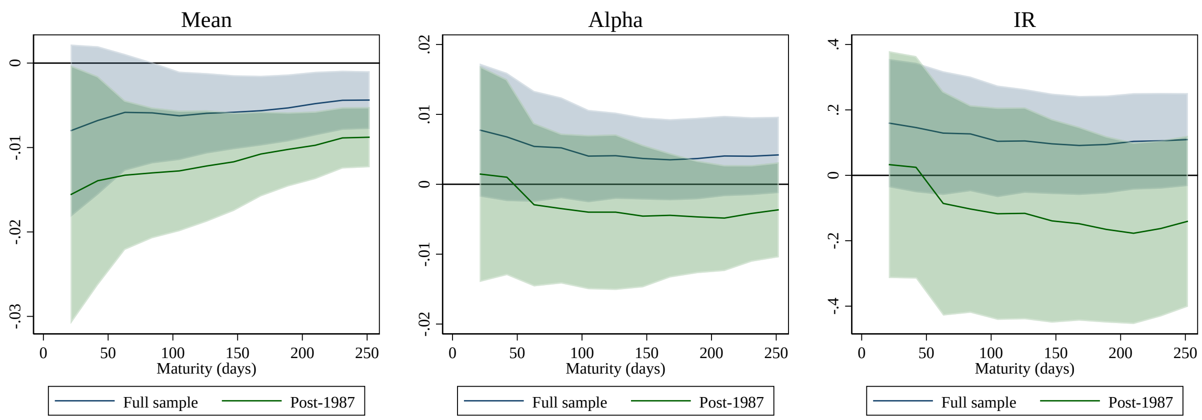
Note: Standard deviations and CAPM betas for synthetic and traded option returns. The shaded regions are 95% confidence intervals (no confidence interval is shown for the dotted line)

Figure A.3: Difference between traded and synthetic options



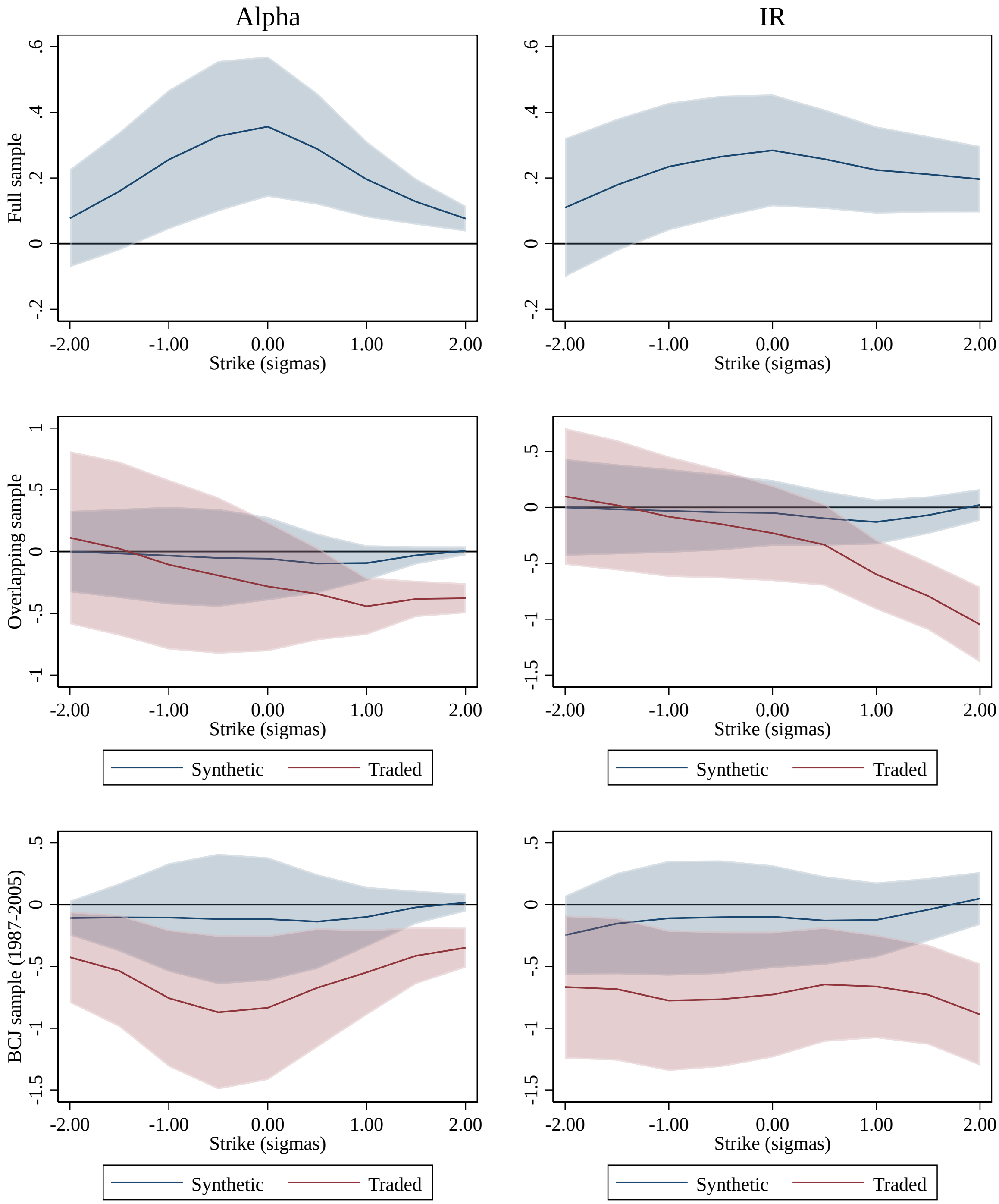
Note: Difference in means, CAPM alphas, and CAPM information ratios between traded and synthetic options, across various samples. Shaded regions represent 95-percent confidence intervals. The overlapping sample in the top row is 1987–2022, and the BCJ sample is 1987-2005.

Figure A.4: Synthetic option returns across maturities



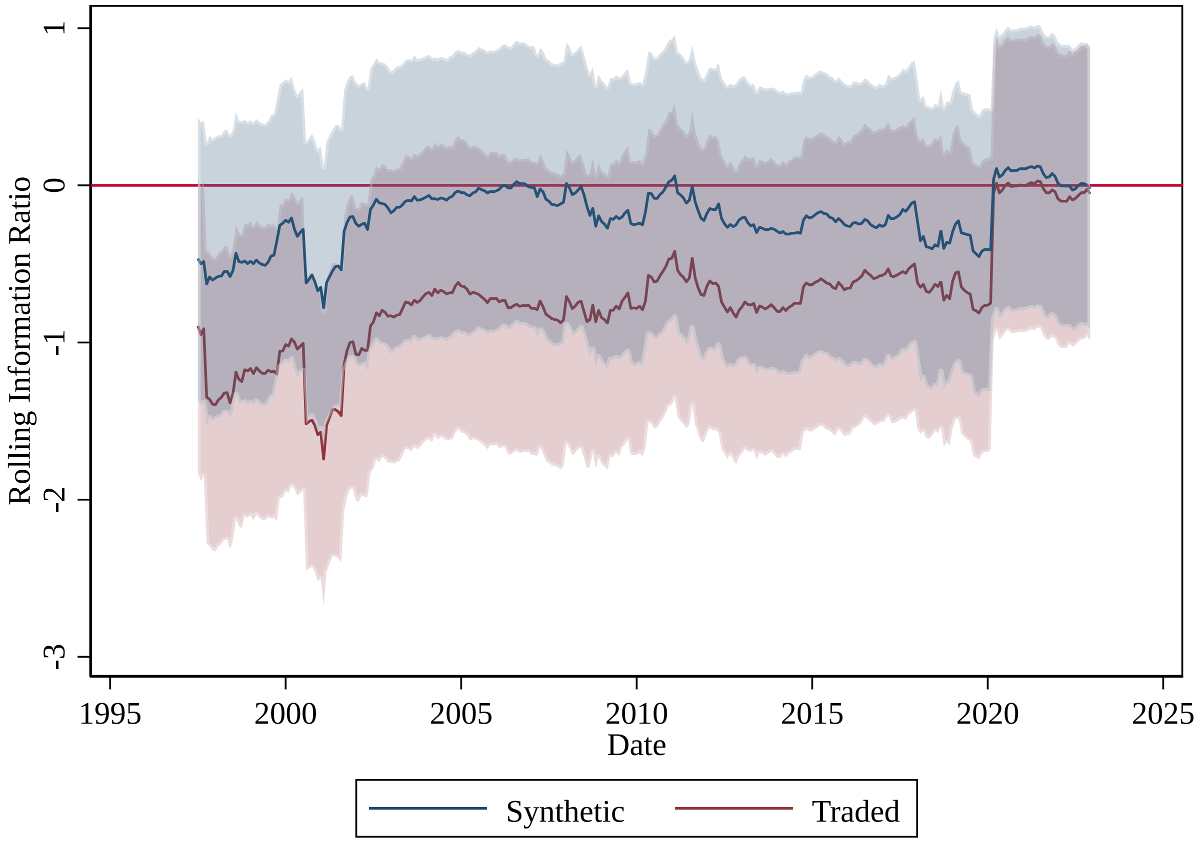
Note: Average return, CAPM alpha, and information ratio for synthetic options across maturities. In all cases, the strike is set to be equivalent to -5% at the monthly maturity (scaling with the square root of the maturity). Note that the two series plotted are not traded versus synthetic options but rather synthetic options in two different samples (the longer maturities are much less liquid for the traded options, especially early in the sample, and present a number of issues in implementation).

Figure A.5: Option returns for moneyness in volatility units



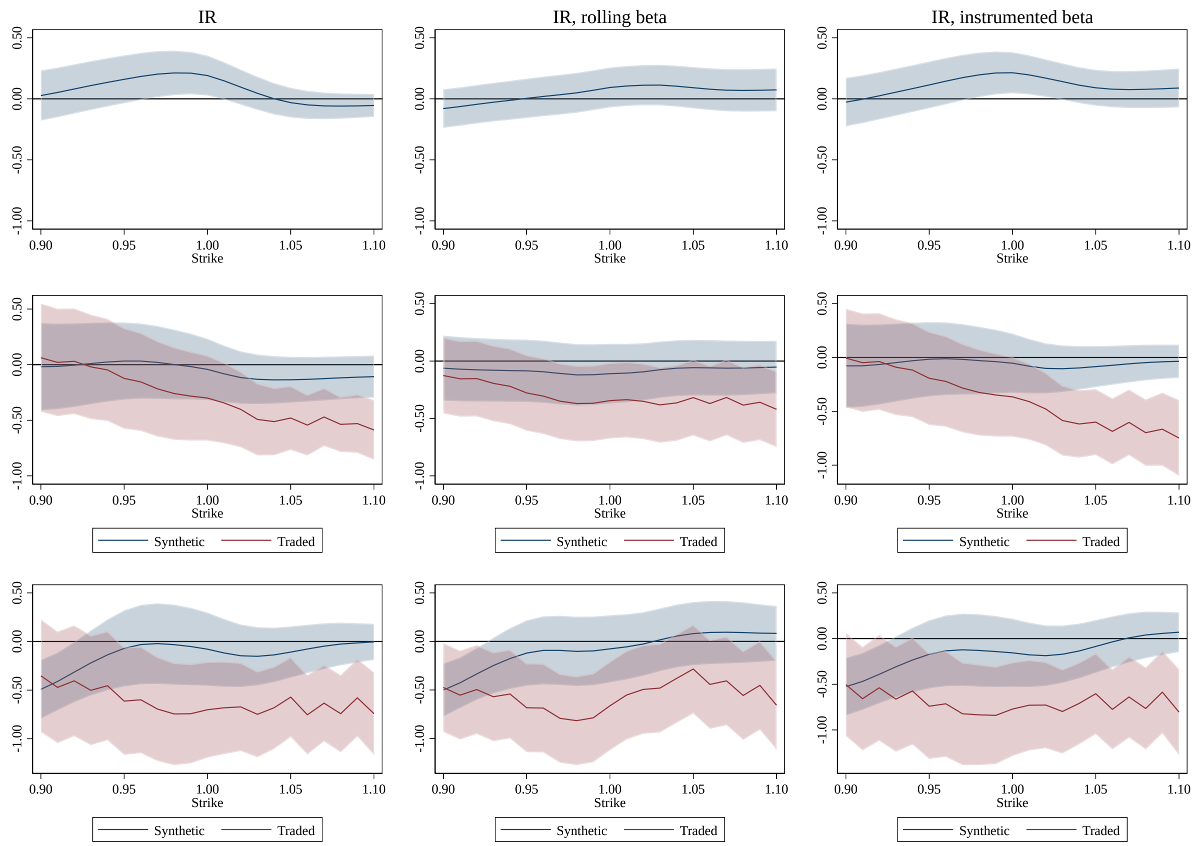
Note: Average return, CAPM alpha, and information ratio for synthetic and traded options. Strikes here are selected in units of volatility instead of as a fixed percentage of the price of the underlying.

Figure A.6: Option returns scaling by option price



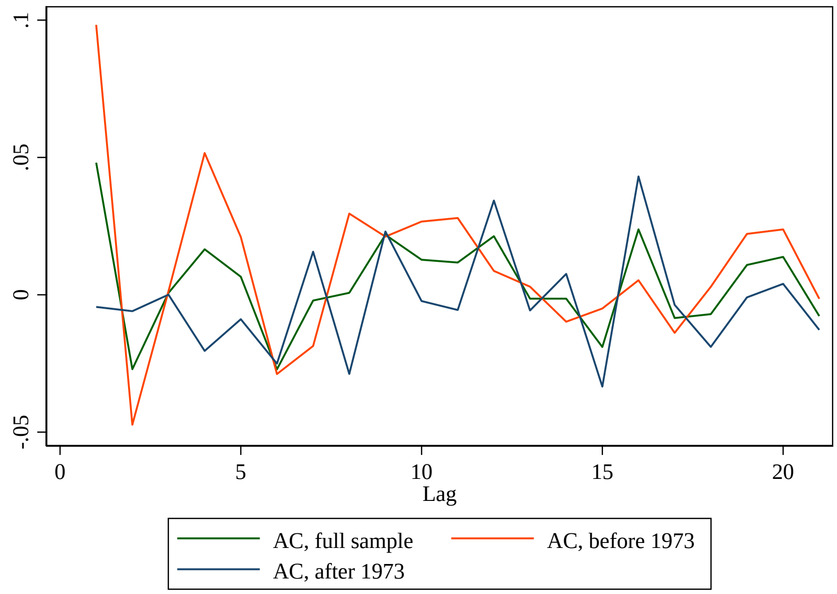
Note: Reports results where the denominator of the return is the price of the option instead of the underlying – so purching a fixed dollar amount of insurance, instead of insurance on a fixed number of units of the underlying.

Figure A.7: Information ratios under various specifications for beta

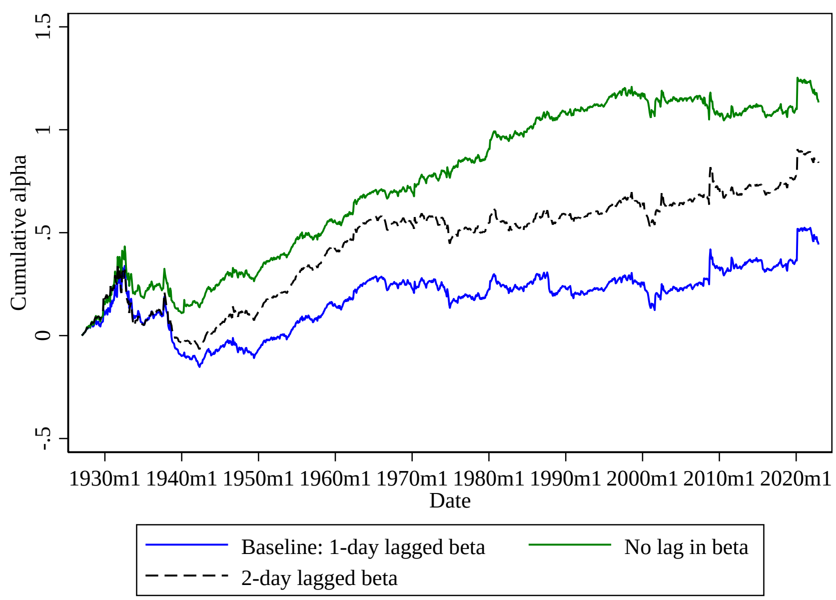


Note: Information ratios under various specifications for the betas. The left-hand column uses the full-sample beta, which is the benchmark specification. The middle column uses betas estimated from a rolling three-month window. Finally, the right-hand column instruments for the conditional beta, as described in the text.

Figure A.8: Daily autocorrelation of returns



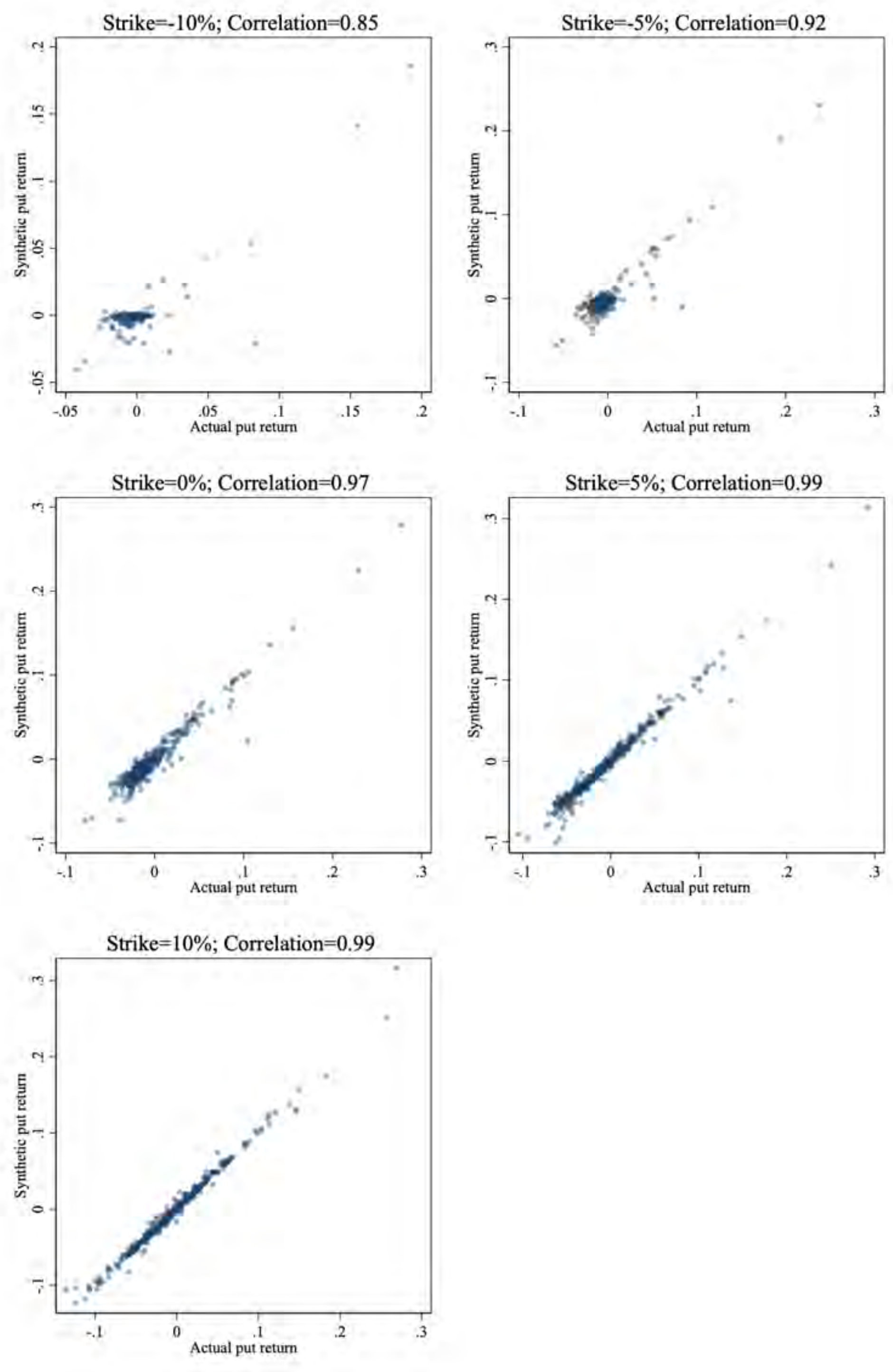
(a) Autocorrelations



(b) Cumulated returns of synthetic options

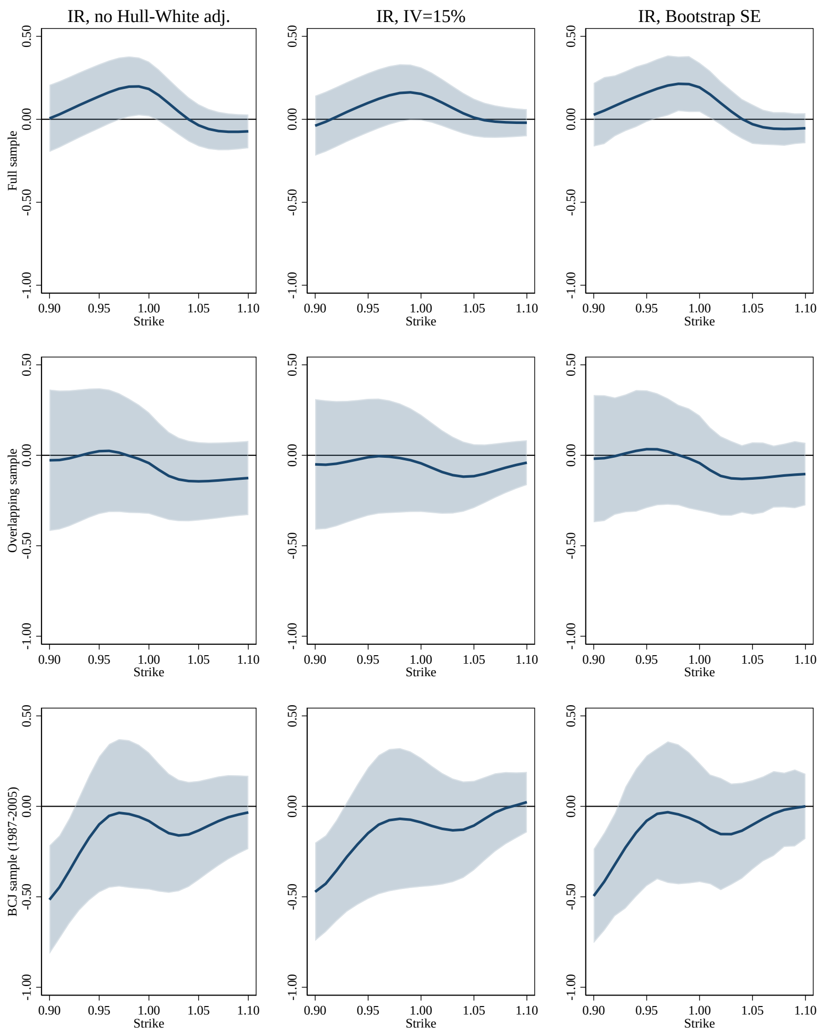
Note: Panel (a) plots the autocorrelations of daily returns up to 21 lags, using the full sample (1926-2022), and using the pre- and post-1973 data separately. Panel (b) plots cumulative CAPM alphas on the synthetic puts built using one-day lagged weights (the benchmark case), no lag, and two-day lagged weights.

Figure A.9: Synthetic versus traded option returns



Note: The scatter plots are for traded verus synthetic option returns over the period 1987–2022. The returns are monthly, rolling on the third Friday.

Figure A.10: Robustness for information ratios

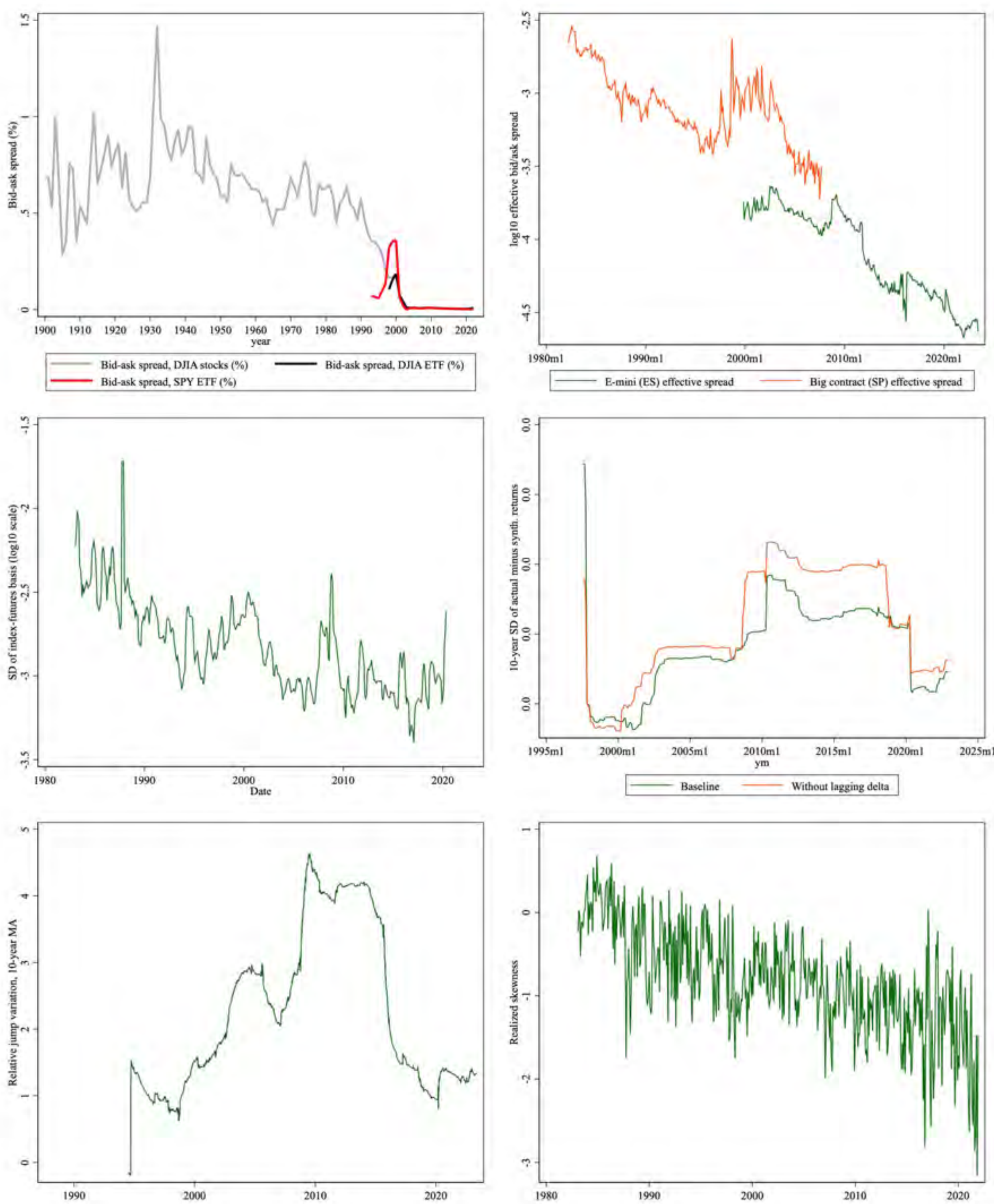


Note: The figure reports information ratios for synthetic options of different strikes, for alternative choices in the construction of the option returns. The first column uses standard Black-Scholes to compute delta (instead of Hull-White). The second column fixes implied volatility at 15% instead of estimating it using historical data. The last column computes standard errors via block bootstrap. The three rows correspond to different samples: the full 1926-2022 sample, the sample in which both traded and synthetic options are available (1987-2022) and the BCJ sample (1987-2005).

Figure A.11: Cumulative alphas for delta-hedged options

Note: This graph reports results analogous to those in figure 5, but giving cumulative alphas for delta-hedged option returns.

Figure A.12: Various measures of hedging risk **figure_frictions.jpg**



Note: The top-left panel plots bid-ask spreads: for the Dow-30 from Jones (2002) and the DIA and SPY ETFs from CRSP. The top-right panel plots effective spreads based on the Roll estimator for S&P 500 futures. The middle-left panel plots the three-month rolling average of the (log_{10}) standard deviation of the gap between the S&P 500 futures price and the level of the index based on 15-minute averages for each. The middle-right panel plots the 10-year rolling standard deviation of the gap between synthetic and traded option returns (which is simply the volatility of the delta-hedged return). The bottom-left panel plots jump variation for the S&P 500 measured as total quadratic variation minus bipower variation from 15-minute returns. The bottom-right panel plots realized skewness of S&P 500 returns from Dew-Becker (2022).