Does the Tail Wag the Dog?

How Options Affect Stock Price Dynamics*

David C. Yang

January 2020

Abstract

I demonstrate empirically that the existence of options affects the autocorrelation of stock returns in the cross section. When the underlying price rises, option writers must buy more stock to re-hedge. Trading costs cause slow re-hedging, leading to increased return autocorrelation. "Hedging demand" quantifies the sensitivity of the option writers' hedge to underlying price changes. Moving from the lowest to highest quintile of hedging demand increases daily return autocorrelation from -5.0% to -1.6%. The associated trading strategy's gross alpha is 18% annualized. For causality, I use an instrumental variable, which uses the institutional idiosyncrasy of round number strike prices.

JEL Codes: G12, G13

Keywords: options, dynamic hedging, stock return autocorrelation

1   Introduction

Does the existence and trading of financial options influence the price movements of their underlying assets? In classical asset pricing, options do not affect their underlying assets because they simply "derive" their value from the underlying. In this paper, I document how options on individual stocks affect their underlying assets through the hedging behavior of option writers. In my model, the end users are option buyers on net. Liquidity providers are option writers on net and sell options to accommodate end user demand. They hedge their exposure by dynamically trading the underlying stock (Black and Scholes, 1973; Merton, 1973).

Dynamic hedging by option writers creates an upward sloping demand curve: when the underlying stock price rises, option writers must buy more of the stock to remain hedged. This upward sloping demand curve applies for an option writer who has sold either call options or put options. For hedging a call option, the intuition of buying more of the underlying as the stock price rises applies naturally. For hedging a put option, it is more natural to think about the upward sloping demand curve as short selling more of the underlying as the stock price falls.

Because it is costly to trade, option writers hedge at fixed intervals.[1] Combined with demand pressure when trading the underlying stock, the model predicts that hedging by option writers increases the autocorrelation of stock returns.[2] To illustrate, suppose there is a positive news shock at period 1, which causes prices to rise. At period 2, option writers re-adjust their hedges, which forces them to purchase more stock. This re-hedging causes prices to rise even further and creates autocorrelation in the stock returns.

---

[1]Though not optimal, fixed interval hedging is often used by market participants, see Leland (1985), Boyle and Vorst (1992), and Cetin, Jarrow, Protter, and Warachka (2006).

[2]Many papers document that demand pressure affects asset prices. For the stock market, see Shleifer (1986); Wurgler and Zhuravskaya (2002); Greenwood (2005). For the effect of demand pressure of options on their own prices, see Green and Figlewski (1999), Bollen and Whaley (2004), and Garleanu, Pedersen, and Poteshman (2009). In contrast, I study the effect of options on stock prices.

If hedging is instead instantaneous, then hedging by option writers increases stock price volatility. In this scenario, when a positive news shock arrives, the option writers are instantly adjusting their hedge as well as fully internalizing the price feedback created by their own demand. In equilibrium, this amplifies the effect of news shocks and increases volatility. Frey and Stremme (1997), Platen and Schweizer (1998), Schonbucher and Wilmott (2000) model this effect theoretically. Ni et al (forthcoming) empirically document the effect of options hedging on volatility. My paper is similar to theirs in that I study the same mechanism of hedging by market makers.[3] My paper differs from their study in three ways: First, I focus on the effect of options hedging on returns and autocorrelation of returns, as opposed to volatility. Second, I study the effect on the cross section of returns, whereas they study the effect on the time series of volatility. Third, I use an instrumental variable approach (distance-to-nearest-round-number) to establish causality.

I test the "hedging increases return autocorrelation" prediction using data on options on individual stocks. I develop a measure of "hedging demand" that quantifies "if the underlying stock price increases by 1%, what fraction of the shares outstanding must option writers additionally purchase to remain hedged?". Hedging demand reflects two components: The first component is the sensitivity of the hedge to changes in the underlying price. More specifically, this sensitivity depends on the convexity of option payoff or "gamma," in option terminology. In contrast, a purely linear financial derivative, such as a futures contract, has zero gamma and the hedge is constant. The second component is the number of options outstanding. Intuitively, hedging ten option contracts generates more price impact in the underlying stock than hedging one option contract.

In the empirical tests, I study the cross sectional variation in hedging demand. I show that stocks with higher hedging demand have higher return autocorrelations. Moving from the lowest to highest quintile of hedging demand increases daily return autocorrelation from

---

[3]In terms of notation, I write gamma in terms of net options written and they write gamma in terms of net options purchased, so the "sign" of my effect is opposite theirs.

-5.0% to -1.6%. The *marginal* effect of hedging demand on return autocorrelation is positive, but the *total* return autocorrelation is still negative because the effect of hedging demand is not enough to overcome the background negative autocorrelation in stock returns. Past research attributes this background negative return autocorrelation to bid-ask bounce or compensation for liquidity provision.[4] The effect of hedging demand on return autocorrelation is robust in different subsamples of the data (first half only, latter half only, large firms only, and excluding observations from the week of option expiration).

I also examine hedging demand formed using only put options and hedging demand formed using only call options. I find that the effect on return autocorrelation is stronger within put options in both economic magnitude and statistical significance. This finding is consistent with the following hypothesis: First, there is a class of trading strategies known as "covered call" and "covered put" strategies, where one does not dynamically hedge with the underlying stock.[5] Second, anecdotal conversation with market participants suggests covered calls are more popular than covered puts.[6] Therefore, since price pressure from dynamic hedging is the mechanism I study, this imbalance would imply a stronger hedging demand effect on return autocorrelation from put options, compared to call options. My results are consistent with that hypothesis.

Because I study return autocorrelation, I can also measure the economic magnitude using a portfolio sorting strategy. This portfolio sorting analysis complements the regressions, since portfolio sorting analyses are less sensitive than regressions to outliers and parametric misspecification. Since the theory suggests that stocks with higher hedging demand have

---

[4]See Roll (1984); Lehmann (1990); Campbell, Grossman, and Wang (1993); Avramov, Chordia, Goyal (2006); Nagel (2012).

[5]In a covered call, the investor is long the underlying stock and sells a call option against it. This strategy allows the stock holder to sell some of the upside exposure of the stock in exchange for the option premium. In a covered put, the investor shorts the underlying stock and sells a put option against it. In both cases, the investor does not dynamically hedge the position.

[6]Comprehensive data measuring this imbalance is difficult to obtain because it requires linked data on investors' long/short positions in both the stock and option market. Lakonishok et al (2007) have partial linked data for a large discount brokerage. However, because they do not observe the short positions in the underlying stock, they observe covered calls, but not covered puts.

higher autocorrelation, I sort stocks into quintiles based on the interaction of past returns and hedging demand. After controlling for the return factors RMRF, SMB, HML, UMD, and Short-Term Reversal, I find a gross alpha of 0.33% weekly (18% annualized) with a t-statistic of 4.2. While not the main focus of this paper, a back-of-the-envelope calculation suggests the portfolio sorting strategy survives transaction costs with a net alpha of 0.10% weekly (5.4% annualized).

So far, I have interpreted the results as hedging demand creating more return autocorrelation. However, a potential confound is that hedging demand may instead measure news entering the option market before the stock market. For example, informed traders might purchase options to make leveraged bets. At the same time, information could slowly integrate into the stock market, causing return autocorrelation (Hong and Stein, 1999). This potential confound is related to, but distinct from, research showing that various ratios related to option volume forecast future stock returns because news enters the option market first. Examples include the put-call volume ratio (Easley, O'Hara, and Srinivas, 1998; Pan and Poteshman, 2006), the option-stock volume ratio (Johnson and So, 2012), and the delta-weighted option order imbalance (Hu, 2014). One difference between those papers and my paper is that I focus on return autocorrelation, rather than just returns. Another difference is that those papers focus on option volume (i.e. flows) and the hedging theory presented here instead focuses on the total level of options that option writers must hedge.

I address this potential confound with an instrumental variable, the absolute difference between the underlying stock price and the nearest round number. This instrument for hedging demand is based on the institutional idiosyncrasy that exchange-traded options are struck at round numbers (e.g. $600.00, rather than $613.12). Furthermore, option convexity (gamma) rises as the stock price approaches the option strike price. Therefore, the absolute difference between the underlying stock price and the nearest round number helps isolate variation in hedging demand, unrelated to news. In particular, my instrumental variable only moves the gamma of outstanding options, which is separate from variation in any

option quantities, whether level or flow. The instrumental variable regression corroborates my other results and implies that hedging demand causally increases return autocorrelation.

The effect of hedging by option writers is related to the mechanical buying/selling by portfolio insurers, which the Presidential Task Force on Market Mechanisms (the "Brady Report") argued exacerbated the October 1987 stock market crash (Brady, 1988). Portfolio insurers attempt to limit portfolio losses by selling as the stock price falls, which also generates an upward sloping demand curve for the stock. Typically, portfolio insurers implement their strategy using index futures, so as the price falls, portfolio insurers sell index futures and index arbitrageurs then transmit the effect to the individual stocks in the index.

My analysis assumes that the end users are net option buyers and hence the liquidity providers are net option writers. In the model, I also assume that the liquidity providers only dynamically hedge delta, as opposed other option exposures such as gamma. In practice, liquidity providers do hedge some of their gamma exposure, so the model should be interpreted as the dynamic hedging arising from unhedged gamma exposure. As discussed in finance textbooks such as Hull (2012, Business Snapshot 18.1), in the industry, it appears true that end users are net option buyers and that liquidity providers only partially hedge gamma exposure due to trading costs.

These assumptions are also closely related to the literature on demand effects in option pricing, which finds that hedged option writing on individual stocks earns positive returns (Bollen and Whaley, 2004; Frazzini and Pedersen, 2012; Cao and Han, 2013).[7] For example, Frazzini and Pedersen (2012) argue that hedged option writing earns positive returns because end users value embedded leverage and hence are option buyers. Bakshi and Kapadia (2003) also find positive returns to hedged option writing on individual stocks, but instead attributes

---

[7]Bollen and Whaley (2004) find that hedged option writing on individual stocks earns positive returns, but the standard errors are large, given their sample of 20 stocks. Using larger cross sections, Frazzini and Pedersen (2012) and Cao and Han (2013) find statistically significant profits. Frazzini and Pedersen (2012) and Cao and Han (2013) state their result as hedged option *buying* earns negative returns, so I state the equivalent result that hedged option *writing* earns positive returns.

it to a fundamental negative volatility risk premium. Other papers find that index options have a greater imbalance of end user demand than individual stock options, which helps explain pricing differences between the two types of options (Bollen and Whaley, 2004; Garleanu, Pedersen, and Poteshman, 2009).

The literature's finding that hedged option writing earns positive returns may appear somewhat in tension with Lakonishok et al (2007), which finds that nonmarket makers are net sellers of individual stock options (i.e. market makers are net buyers). A potential explanation is the distinction between a formal market maker and opportunistic liquidity providers in the options market. Market makers have specific obligations to stand ready to buy and sell options at all times, hence they are not the sole liquidity providers in the option market.[8] For example, nonmarket makers such as hedge funds may opportunistically provide liquidity in the option market by selling options and hedging their exposure when option premia are attractive.[9] The mechanism I study, along with the papers on demand effects in option pricing, focuses on the hedging behavior of all liquidity providers, not just the formally designated market makers.

My empirical work connects most directly to other empirical research on the effect of a derivative on its underlying asset. As discussed earlier, Ni et al (forthcoming) study the same mechanism, but my paper differs in (1) my focus on the effect of options hedging on returns and autocorrelation of returns, as opposed to volatility; (2) my focus on the cross section of returns, instead of the time series of volatility; (3) my use of an instrumental variable approach (distance-to-nearest-round-number) to establish causality. Other research has studied the effect of options on the underlying asset at specific times, including the introduction of a new derivative (Conrad, 1989; Detemple and Jorion, 1990; Sorescu, 2000) and re-hedging effects at option expiration (Ni, Pearson, and Poteshman, 2005; Golez and

---

[8]See CBOE (2009) for an example of the specific obligations of option market makers.

[9]In Lakonishok et al (2007), hedge funds are nonmarket makers. Specifically, hedge funds are clients of full-service brokerages, which they document have a significant net short position in options.

Jackwerth, 2012).

Section 2 describes the theoretical framework to motivate my measure of hedging demand. Section 3 describes my dataset and basic patterns in the time series and cross section. Section 4 describes the main empirical results (baseline results, robustness tests, long horizon effects, and decomposing returns into their systematic and idiosyncratic components). Section 5 measures the economic magnitude using a portfolio sorting strategy. Section 6 describes the instrumental variable regression. Section 7 concludes with a discussion of natural extensions of this paper.

2   Theoretical Framework and Hedging Demand

In the theoretical framework, I aggregate all the option writers into one representative agent option writer. In this section, I model how the option writer hedges, how the option writer re-hedges after changes in the underlying price, and how this re-hedging affects stock prices.

2.1   Notation and Option Terminology

Each stock has many different options on it, e.g. put/call, different maturities, and different strike prices. Let i denote the stock, t denote the time, and k denote the option. Then, let V_{itk} denote the value of option of type k (e.g. a call option with strike price \$600 that expires in 30 days). Let $written_{itk}$ denote the number of options of type k sold by the representative agent option writer. The option writer's aggregate exposure across the portfolio of options he has sold is then:

$$V_{i,t,agg} := \sum_{k} (written_{itk} \cdot V_{itk}) \tag{1}$$

For example, if the option writer has sold one unit of the option $k = 1$ and one unit of the option $k = 2$, then $V_{i,t,agg} = V_{i,t,1} + V_{i,t,2}$.

I extensively use two concepts from the options literature: "delta" (Δ) and "gamma" (Γ).

Delta and gamma characterize the sensitivity of an option or portfolio of options to the underlying price.

$$\Delta \ := \ \frac{\partial V}{\partial P} \tag{2}$$

$$\Gamma \ := \ \frac{\partial^2 V}{\partial P^2} = \frac{\partial \Delta}{\partial P} \tag{3}$$

The sensitivity of the portfolio is a linear combination of the sensitivity of the individual options:

$$\Delta_{i,t,agg} \ = \ \sum_k (written_{itk} \cdot \Delta_{itk}) \tag{4}$$

$$\Gamma_{i,t,agg} \ = \ \sum_k (written_{itk} \cdot \Gamma_{itk}) \tag{5}$$

2.2 Hedging and Re-Hedging

I assume that the end users are net option buyers and the liquidity providers are net option writers. I also assume that the liquidity providers only dynamically hedge delta and not other option exposures, e.g. gamma. In practice, liquidity providers do partially hedge gamma, so one should interpret the model as dynamic hedging due to unhedged gamma exposure. In actual trading, it appears to be true that end users are net option buyers and that option liquidity providers only partially hedge gamma due to trading costs, see finance textbooks such as Hull (2012, Business Snapshot 18.1).

Consider the representative agent option writer who has sold various options on a given stock i and wants to hedge his position. The option writer hedges his position because the end users in aggregate are option buyers. For simplicity, assume all options are European options. In the model, the option writer hedges his exposure on stock i with stock i itself. Hence, I drop the i subscript in the model for notational simplicity. In the empirical work, I focus on cross sectional variation across stocks and hence I use the i subscript in the

regressions.

The representative agent option writer W aims to collect the premium for writing options, as opposed to making any directional bet on the underlying stock. Let ξ_t^W denote the number of shares of the underlying stock that the option writer holds at time t. The option writer's objective function is to minimize the variance of his net worth N_{t+1}^W:

$$\min_{\xi_t^W} Var(N_{t+1}^W) \tag{6}$$

where

$$N_{t+1}^W = \underbrace{-V_{t+1,agg}}_{\text{Sold Options}} + \underbrace{\xi_t^W \cdot P_{t+1}}_{\text{Stock}} + \underbrace{(N_t^W + V_{t,agg} - \xi_t^W \cdot P_t)}_{\text{Cash}} \tag{7}$$

The option writer can only choose his exposure to the underlying stock ξ_t^W. The option writer cannot choose to dispose of the options and cannot choose the trivial solution of holding zero stock and zero options. Intuitively, while one option writer can sell his exposure to another option writer, the representative agent option writer hedges his exposure.

Proposition 1. *The solution to the option writer's objective function is to buy $\xi_t^W = \Delta_{t,agg}$ units of the underlying.*

Such a position is often called a "delta-neutral position," because hedge offsets the delta of the options sold. Intuitively, the option writer's wealth is now unaffected by small changes in the price of the underlying (Figure 1a).

$$\begin{aligned} \frac{\partial N^W}{\partial P} &= \text{Loss from Sold Options} + \text{Gain from Hedge} \\ &= -\frac{\partial V_{agg}}{\partial P} + \Delta_{agg} = 0 \end{aligned} \tag{8}$$

Proposition 2. *If the underlying price rises by 1%, the option writer must buy an additional $\Gamma_{agg} \cdot 1\%P$ shares of the underlying to remain hedged.*

10

If the underlying price P rises by 1%, then the option writer now demands approximately $\Delta_{agg} + \frac{\partial \Delta_{agg}}{\partial P} \cdot 1\% P$ shares of the underlying stock (Figure 1a). Since $\Gamma_{agg} = \frac{\partial \Delta_{agg}}{\partial P}$, the *change in the hedge* is $\Gamma_{agg} \cdot 1\% P$. If $\Gamma_{agg} > 0$, then this is an upward sloping demand curve. As the price rises, hedging requires buying more of the underlying.

An option writer hedging a put option has the same upward sloping demand curve. For put options, $\Delta_{put} < 0$, so the option writer hedges his position by "buying Δ_{put}" (i.e. "selling $|\Delta_{put}|$") units of the underlying stock. Since put options are also convex ($\Gamma_{put} > 0$), as the underlying price rises, the option writer buys more of the underlying stock (i.e. reduces his short position).

Above, I assumed that the option writer hedges by dynamically trading the underlying stock. One could alternately model the option writer selling many options on the different firms in an index and then attempting to hedge his exposure with the index itself. However, this hedge is imperfect because options have non-linear payoffs: an option on a portfolio is distinct from a portfolio of options. As the underlying stock prices change, the required amount of stock in each company (to remain hedged) can deviate from the weights in the index. For example, suppose the index is just an equal-weighted average of two stocks: X and Y. Assume X and Y have the same number of shares outstanding, so we only need to track the share prices. Suppose the option writer sells one call option on each stock. If X rises by $10 and Y falls by $10, then the index is unchanged. However, the option writer now needs more of stock X and less of stock Y. Hence, the weights in his hedge now deviate from the weights in the index. For this reason, I model hedging with the stock itself.

2.3   Price Impact of Re-Hedging

Next, I describe how re-hedging by the option writer increases autocorrelation in the stock returns. There are two assets in this model: the risky asset ("stock") and the risk-free asset. At some final date T, the stock pays a single terminal dividend of $F_T = F_0 + \sum_{j=0}^{T} \epsilon_j$, where

11

$\epsilon_j \sim N(0, 1)$. I normalize the stock to have a total quantity of 1. I also normalize the net risk-free rate to 0, by assuming the risk-free asset is elastically supplied at that rate.

There are three groups of representative agents: the option buyer, the option writer, and the "fundamental investor." In the option market, the option buyer purchases options from the option writer. For simplicity, assume that the option buyer does not interact with the stock market. The option writer sells the options and hedges his risk in the stock market. Alternately, one can allow the option buyer to trade in the stock market, as long as the option buyer's actions do not fully offset hedging from the option writer.

In the stock market, the representative agent option writer interacts with the representative agent fundamental investor (FI). The fundamental investor's role is simply to provide a downward sloping demand curve, against which the option writer trades. The following setup is one way to generate this downward-sloping demand curve: The fundamental investor only trades based on one-period ahead fundamentals F_{t+1}. This assumption prevents the fundamental investor from front running the option writer's hedging. The fundamental investor has constant absolute risk aversion (CARA) utility, where τ denotes the risk tolerance, i.e. reciprocal of risk aversion. Let N_{t+1}^{FI} denote her net worth in the next period and ξ_t^{FI} denote the fundamental investor's demand for the underlying stock. Given CARA utility and normally distributed risk, her objective function is equivalent to mean-variance optimization:

$$\max_{\xi_t^{FI}} E[N_{t+1}^{FI}] - \frac{1}{2\tau} Var[N_{t+1}^{FI}] \tag{9}$$

where $N_{t+1}^{FI} = F_{t+1}\xi_t^{FI} + (N_t^{FI} - P_t\xi_t^{FI})$. This setup implies that the fundamental investor has a downward sloping demand curve for the underlying stock $\xi_t^{FI} = \tau \cdot (E[F_{t+1}] - P_t)$.

As a benchmark, consider the equilibrium with only the fundamental investor, i.e. no hedging by the option writer. In this benchmark, the stock price is simply the fundamental value less a risk premium, $P_t^{bmark} = F_0 + \sum_{j=1}^{t} \epsilon_j - \frac{1}{\tau}$. Therefore, when there is only the

12

fundamental investor, stock prices are a random walk:

$$Cov(P_{t+1}^{bmark} - P_t^{bmark}, P_t^{bmark} - P_{t-1}^{bmark}) \;=\; Cov(\epsilon_{t+1}, \epsilon_t)$$

$$= 0 \tag{10}$$

Now, I add the hedging by the option writer. I assume that the option writer can only hedge every other period $t \in \{0, 2, 4, ...\}$, e.g. because it is too costly to hedge every period. Therefore, when the option writer delta hedges, his demand for the underlying stock is $\xi_t^W = \Delta_{t,agg}$ if $t \in \{0, 2, 4, ...\}$ or $\xi_t^W = \Delta_{t-1,agg}$ if $t \in \{1, 3, 5, ...\}$. Setting supply equal to demand, we get the following prices:

$$P_t = \begin{cases} P_t^{bmark} + \frac{\Delta_{t,agg}}{\tau} & \text{if } t \in \{0, 2, 4, ...\} \\ P_t^{bmark} + \frac{\Delta_{t-1,agg}}{\tau} & \text{if } t \in \{1, 3, 5, ...\} \end{cases} \tag{11}$$

When the option writer hedges (in the even periods $t \in \{0, 2, 4, ...\}$), we must solve for a fixed point because $\Delta_{t,agg}$ is a function of P_t. This fixed point exists as long as $\Gamma_{agg}/\tau < 1$. Intuitively, as the option writer buys more of the underlying stock, the price rises, which induces more desire for shares, and so on, until we reach the fixed point solution. The condition $\Gamma_{agg}/\tau < 1$ ensures this process converges. If the fundamental investor has infinite risk tolerance τ, then prices are equal to the random-walk benchmark. Intuitively, this is because higher risk tolerance lowers the price impact of the hedging by the option writer.

To illustrate this equilibrium, consider the impulse response where there is a single positive news shock at time 1 (i.e. $\epsilon_j = 0$ for $j \neq 1$). Figure 1b depicts this impulse response. Using the fact that $\Delta_2 \approx \Delta_0 + \Gamma_1(P_2 - P_0)$, I can solve:

13

Returns	With Option Writers	Benchmark
$P_1 - P_0$	ϵ_1	ϵ_1
$P_2 - P_1$	$\epsilon_1 \cdot \frac{\Gamma_{1,agg}/\tau}{1-\Gamma_{1,agg}/\tau}$	0
$P_{t+1} - P_t$ for $t \geq 2$	0	0

Proposition 3. *Hedging by the option writer creates return autocorrelation, $Cov(P_2 - P_1, P_1 - P_0) > 0$. As $\Gamma_{1,agg}$ rises, return autocorrelation rises. Also, as the fundamental investor's risk tolerance τ falls, return autocorrelation rises.*

I normalized $Var(\epsilon_1) = 1$, so $Cov(P_2 - P_1, P_1 - P_0) = \frac{\Gamma_{1,agg}/\tau}{1-\Gamma_{1,agg}/\tau} > 0$. The key ratio is $\Gamma_{1,agg}/\tau$, which is the ratio of the option writer's hedging demand to the fundamental investor's risk tolerance. Intuitively, as $\Gamma_{1,agg}$ rises, autocorrelation rises because the option writer must buy more shares to re-hedge. And, as the fundamental investor's risk tolerance τ falls, autocorrelation rises because there is more price impact from the hedging.

The model also predicts that the delta hedging affects the stock price permanently, as opposed to causing a temporary "overshooting." For example, in the impulse response, the returns after period 2 (i.e. $P_{t+1} - P_t$ for $t \geq 2$) are zero, not negative. Intuitively, at period 2, the option writer increases his holdings of the underlying stock because his option exposure has increased. However, after period 2, because there are no more shocks, the option writer remains hedged and does not need to adjust his stock holdings.

Instead of assuming that the option writer can only hedge every other period as I have done here, one could alternately model the option writer as hedging every period in a backward-looking manner. That is, one could assume that the option writer holds $\Delta_{t-1,agg}$ shares of the underlying at period t. This alternate setup avoids the fixed point equilibrium above and predicts more sluggish price adjustment. However, this alternate setup has the drawback that the option writer is no longer forward-looking. If hedging is instead instantaneous, then the full price adjustment occurs in period 1, leading to increased volatility and no effect on subsequent returns.

2.4   Hedging Demand and its Empirical Counterpart

I now define the key variable "hedging demand." My model showed that higher Γ_{agg} is associated with higher serial correlation, holding all other variables fixed. Intuitively, when Γ_{agg} is higher, for the same shock, the option writer must buy more of the underlying stock. Hedging demand is Γ_{agg} re-scaled, so I can compare across different stocks with different share prices and shares outstanding.

Re-scaling is also necessary because the price impact model used a setup with CARA utility and normally distributed risk. This setup is convenient for models with heterogeneous agents because the demand for risky assets is independent of wealth. Hence, the equilibrium does not depend on the distribution of wealth amongst the agents. However, one drawback of this setup is that the product of two normal random variables is not normally distributed. Therefore, in these types of models, "returns" are typically just the change in the price level $P_{t+1} - P_t$. However, in empirical work, returns are $\frac{P_{t+1}-P_t}{P_t}$, so I re-scale to account for this difference.

Definition 1. *Theoretical Version of Hedging Demand: If the price of the underlying stock increases by 1%, what fraction of the shares outstanding must the option writer additionally purchase to remain hedged? Answer:*

$$
\begin{aligned}
H^* &:= \frac{1\%P}{\text{SharesOut}} \cdot \Gamma_{agg} \\
&= \frac{1\%P}{\text{SharesOut}} \cdot \sum_k (written_k \cdot \Gamma_k)
\end{aligned}
\tag{12}
$$

Definition 2. *Empirical Counterpart of Hedging Demand:*

$$
H_{it} := \frac{1\%P_{it}}{\text{SharesOut}_{it}} \cdot \sum_k (100 \cdot OI_{itk} \cdot \Gamma_{itk})
\tag{13}
$$

$$
LogH_{it} := Log(H_{it})
\tag{14}
$$

where i denotes the stock, t denotes the time, and k denotes the option.

I denote the theoretical version H^* with an asterisk because the empirical counterpart of hedging demand contains two adjustments. First, in actual trading, each option contract corresponds to 100 shares. Hence, I include 100 as a normalizing constant. Second, I approximate $written_{itk}$ with open interest OI_{itk}. In the model, the two are equivalent (i.e. $written_{itk} = OI_{itk}$) because the representative agent option writer hedges all the sold options.

However, in a more nuanced model, there may be an additional group of option writers who are end users and do not hedge their exposure. Let $written'_{itk}$ denote the options sold by these option writers who do not hedge. (Let $written_{itk}$ continue to denote the options sold by the option writers who do hedge.) In this more nuanced model, $OI_{itk} = written_{itk} + written'_{itk}$. If a constant fraction of the open interest is always actively hedged, then the empirical definition $LogH_{it}$ equals the "true" log hedging demand plus a constant. For this reason, I focus on $LogH_{it}$ in the empirical work.

The other reason I focus on $LogH_{it}$ is because the log form allows us to more easily control for the model of price pressure. Price impact may depend on some combination of fraction of shares outstanding, fraction of volume, etc. By using the log form, log shares outstanding enters linearly into the estimation. I then add controls of log market cap, log dollar volume, and log share price. It also turns out that using just hedging demand H_{it} gives similar empirical results.

3 Dataset and Summary Statistics

I use panel data on two levels: the option level and the equity level. The option-level data aggregates to equity-level data. At both levels, data frequency is daily. The dataset spans from Jan 1996 to Aug 2013 for a total of roughly 4400 days. Table 1 contains a list of the main variables used in this paper.

At the option level, an observation is an equity-date-option triplet. For example, an observation might be the Apple, Inc. call option with strike price $600.00 that expires on Feb 16, 2013 as observed on Jan 02, 2013. Option-level data come from OptionMetrics. I include all observations from January 1996 (start of the database) to August 2013. In terms of cross sectional span, the OptionMetrics database covers all U.S. exchange-listed options. While OptionMetrics also has data for index options, in this paper, I focus on options on individual stocks. Option-level data is 118 GB in size and contains 977 million observations.

For option risk sensitivities gamma Γ, I use the estimates from OptionMetrics, which are based on the Cox, Ross, and Rubinstein (1979) binomial tree method. Their model accommodates discrete and continuous dividends. The other key option-level datum is open interest, which is the total number of contracts that are not settled. My main explanatory variable Hedging Demand (H_{it}) collapses these option-level data into an equity-level statistic at the daily frequency.

At the equity level, an observation is an equity-date pair. For example, an observation might be Apple, Inc. on Jan 02, 2013. Across the 17 years, the dataset of 977 million option-level observations rolls up into 4.3 million equity-level observations, when merged with equity-level data. The mean number of equities per daily cross section is roughly 1000. Using data from the Center for Research in Security Prices (CRSP), I obtain equity-level data on holding period returns, shares outstanding, trading volume, close prices, and bid-ask spreads. From CRSP, I also obtain lower-frequency data on CRSP size decile cutoffs. From Compustat, I obtain data on accounting book value. I lag lower-frequency data to avoid look-forward bias.

In Section 5, I analyze the results using a long-short portfolio sorting strategy. In that section, I further collapse the data to the day-level. For example, an observation might be the return spread for the high vs low quintile of stocks on Jan 02, 2013. For that analysis, I also use data from Ken French's Data Library for the factors of RMRF, HML, SMB, UMD, and Short-Term Reversal.

17

3.1 Time Series and Cross Sectional Patterns in the Dataset

Figure 2 plots the time series and cross sectional patterns of hedging demand. Figure 2a plots the average of log hedging demand *LogH* over time. The large decline in hedging demand in 2008-2009 is due to both a decline in the open interest of options and a large decline in stock prices. A decline in open interest lowers hedging demand because it reduces the number of options the option writers must hedge. A large decline in stock prices lowers hedging demand because it reduces the gamma, the sensitivity of the hedge to changes in the underlying stock price. In particular, options are generally issued with strike prices close to the current stock price. Hence, when there is a large price rise or fall, the stock price is now far away from the strike price of previously issued options (Figure 4). As gamma is highest near the strike price (i.e. near the "kink" in the option payoff diagram), large price movements reduce the gamma of previously issued options and reduce hedging demand.

Hedging demand has a strong monthly cyclic variation due to options expiration. Figure 2a also shows average log hedging demand for Jan 2012 to Aug 2013 (end of the dataset). Exchange traded equity options expire the Saturday after the third Friday of the expiration month. For example, all exchange-traded equity options that expired in the month of January 2012, expired on Saturday Jan 21, 2012. I observe that open interest falls dramatically right before each month's expiration. This monthly periodicity offers an alternate interpretation to the results of Ni, Pearson, and Poteshman (2005), which finds that stock prices cluster near the strike price on option expiration dates. In theory, delta hedging by option buyers should cause clustering and delta hedging by option writers should cause de-clustering. However, when Ni, Pearson, and Poteshman (2005) roughly classify traders into delta hedgers and end users, they find clustering for both groups. As a result, they conclude that effects other than re-hedging are at work. The alternate explanation suggested by Figure 2a is that the option writers are the delta hedgers and they have the fewest number of option contracts to hedge right before option expiration (due to the monthly periodicity). Therefore, in the

18

time series, there would be the less de-clustering (i.e. more clustering) at expiration. This prediction is a time series prediction, which is separate from the cross sectional variation I use in the main results of this paper.

Figure 2b plots the cross sectional dispersion of $LogH$. Specifically, it displays the cross sectional standard deviation over time and displays a histogram of de-meaned $LogH$. In the histogram, I de-mean $LogH$ because average $LogH$ changes over time. I observe that the standard deviation of $LogH$ is higher in more recent years.

Table 2 displays the summary statistics. Panel (a) displays the summary statistics for the entire dataset. Panel (b) shows how the averages vary across the quintiles of hedging demand. Roughly speaking, the dataset is 1000 firms per day across 4400 days. The firms in the sample are relatively large, with a mean market capitalization of $7.9 billion and a median market capitalization of $2.2 billion. The variable $H \cdot \frac{ShareOut}{Volume}$ re-scales hedging demand by daily volume instead of shares outstanding. Hence, it measures the fraction of *daily volume* the option writer would need purchase to remain hedged after a 1% increase in the underlying stock price. Because I use open interest as a proxy for the number of options being hedged, both H and $H \cdot \frac{ShareOut}{Volume}$ are an upper bound on the hedging demand–specifically, the hedging demand if all open option contracts were actively hedged by option writers.

Average hedging demand H is 0.028% and average $H \cdot \frac{ShareOut}{Volume}$ is 4.24%, which means a 1% increase in the underlying stock price would require purchasing an additional 0.028% of the shares outstanding (4.24% of the daily volume). For the highest quintile, $H = 0.09\%$ and $H \cdot \frac{ShareOut}{Volume} = 10.08\%$. Stocks with higher hedging demand are larger and more liquid, i.e. higher turnover and lower spread. CAPM beta rises with hedging demand. The mean of volatility does not rise with hedging demand, but the median (not shown) rises slightly.

4 Main Empirical Results

The theory in Section 2 predicts that stocks with higher hedging demand have higher return autocorrelation, due to hedging by option writers. In this section, I test this prediction using a regression framework. In addition to the baseline regressions, I also analyze the robustness to different subsamples and different estimation methods; the effect on long horizon returns; and the effect of decomposing returns into their systematic and idiosyncratic components.

4.1 Baseline Regression

Let r_{it}^e denote the log returns of equity i at time t in excess of the log risk-free rate. For the risk-free rate, I use the one-month Treasury bill. The subscript ma denotes a lagged moving average over the last five trading days (i.e. one trading week). For example, $r_{ma,i,t}^e$ denotes the lagged moving average of returns. In the baseline regression, the right hand side regressors are returns $r_{ma,i,t}^e$, log hedging demand $LogH_{ma,i,t}$, controls $X_{ma,i,t}$, and their interactions.

$$
\begin{aligned}
r_{i,t+1}^e \;=\;& b_{0,t} + b_1 \cdot r_{ma,i,t}^e + \lambda \cdot r_{ma,i,t}^e \times LogH_{ma,i,t} \\
& + b_2 \cdot LogH_{ma,i,t} + \underbrace{b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t}}_{\text{Controls}} + \epsilon_{i,t+1}
\end{aligned} \tag{15}
$$

To focus on cross sectional variation, I allow for a different intercept $b_{0,t}$ for each time period. My main regressions use the Fama-MacBeth methodology, which implicitly includes the time varying intercept; the panel regressions explicitly include the time fixed effect.

Since I study return autocorrelation in the cross section, I test whether cross sectional variation in past returns forecasts cross sectional variation in future returns. In Equation 15, the regressors involving past returns are $r_{ma,i,t}^e$, $r_{ma,i,t}^e \times LogH_{ma,i,t}$, and $r_{ma,i,t}^e \times X_{ma,i,t}$. Therefore, the *total* return autocorrelation is $b_1 + \lambda \cdot LogH_{ma,i,t} + b_3 \cdot X_{ma,i,t}$. I cross-sectionally

de-mean log hedging demand $LogH_{ma,i,t}$ and the controls $X_{ma,i,t}$, so the coefficient b_1 has the natural interpretation as the return autocorrelation for a firm with average characteristics (i.e. mean log hedging demand, mean log market capitalization, etc.).

The coefficient of interest is the λ coefficient on the interaction term $r^e_{ma,i,t} \times LogH_{ma,i,t}$. This coefficient measures the *marginal* effect of hedging demand on return autocorrelation. The theory in Section 2 predicts that $\lambda > 0$, i.e. higher $LogH_{ma,i,t}$ is associated with more autocorrelation.

Since I estimate the impact of hedging demand on autocorrelation, the key controls $X_{ma,i,t}$ of interest are other variables that might also affect autocorrelation or price impact (e.g. size of the firm). I control for these covariates by adding them as interactions with $r^e_{ma,i,t}$. Strictly speaking, since I control for multiple covariates, I should express $X_{ma,i,t}$ as a vector. However, I express it this way for notational simplicity.

The lagged moving average allows us to estimate the average effect over multiple days because there is no fundamental reason to believe that hedging must happen within one day. Market practitioners, with whom I spoke, confirmed that they often rebalance over a few days to reduce transaction costs. In a more generalized regression, one could alternately estimate a separate coefficient for each lag. However, I use the lagged moving averages because it is easier to display the results concisely. Note also that I use the average of the log, i.e. $LogH_{ma,i,t}$, as opposed to the log of the average, i.e. $Log(H_{ma,i,t})$. I use the former because it more easily generalizes to explicitly estimating a separate coefficient for each jth lag of $Log(H_{i,t-j})$. Both the average of the log and the log of the average give very similar estimates.

Table 3 displays the baseline results, estimated using the Fama-MacBeth methodology. The controls are log market capitalization, log dollar volume, and log share price. Since $LogH = Log(\text{Num Shares to Remain Hedged}) - Log(\text{SharesOut})$, these controls are also equivalent to normalizing hedging demand by some linear combination of log market capitalization, log dollar volume, and log share price. Since log turnover equals log dollar

volume minus log market capitalization, if I alternately use log turnover in the controls, I find that autocorrelation falls as turnover increases, which is consistent with the literature (e.g. Campbell, Grossman, and Wang, 1993). Later, in Table 6, I use alternate estimation methods, including Fama-MacBeth with Newey-West standard errors and panel regression with clustered standard errors. For legibility, displayed coefficients are regression coefficients multiplied by 100.

Table 3 Column (1) shows the regression of returns on lagged returns. Columns (2), (3), (4), and (5) then estimate the effect of hedging demand (the λ coefficient on the interaction term $r^e_{ma,i,t} \times LogH_{ma,i,t}$) with different controls. The b_1 coefficient on r^e_{ma} is negative throughout. This negative coefficient matches the general finding that equity returns have negative autocorrelation in the cross section at short horizons, which past research has linked to bid-ask bounce or compensation for liquidity provision (e.g. Roll, 1984; Nagel, 2012).

Next, I examine the λ coefficient on the interaction term $r^e_{ma,i,t} \times LogH_{ma,i,t}$ across Columns (2), (3), (4), and (5). In terms of statistical significance, the estimates are all significant with t-statistics exceeding 3.0. The estimated λ coefficient is positive, which implies that higher hedging demand is associated with higher return autocorrelation. This finding is matches the prediction from the theory in Section 2. The estimated economic magnitude depends on the controls, but is roughly similar across the columns.

In Table 3 Column (5), which includes all the controls of log market capitalization, log dollar volume, and log share price, the marginal impact of log hedging demand is $\hat{\lambda} = 0.79\%$. The standard deviation of the de-meaned $LogH_{ma}$ is 1.5, so a one standard deviation increase in log hedging demand increases the return autocorrelation by 1.1%. This increase is meaningful as the average firm has an estimated return autocorrelation of $\hat{b}_1 = -3.15\%$. Put differently, as we move from the lowest to highest quintile of hedging demand, the total return autocorrelation $(b_1 + \lambda \cdot LogH_{ma,i,t})$ increases from -5.0% to -1.6%. The total autocorrelation is still negative for the highest quintile because the effect is not enough to offset the fact that stocks have reversion in general at this time scale $(b_1 < 0)$. However, in other asset classes

with less background reversion, hedging option writing could push the total autocorrelation positive.

While not the focus of this paper, the tables show that $LogH_{ma}$ robustly predicts lower returns in the cross section, i.e. $b_2 < 0$. This effect is driven by the fact that option open interest is a strong negative predictor of cross sectional returns (Yang, 2020). This finding is related to the research showing that high option *volume* forecasts low returns (Johnson and So, 2012).

4.2   Robustness: Different Subsamples and Estimation Methods

In this subsection, I explore the robustness of the baseline results to different subsamples and different estimation methods. To fix a point of reference, I compare relative to the main baseline regression in Table 3 Column (5). Naturally, one can choose a different point of reference as well. In the main baseline regression, I found $\hat{\lambda}_{baseline} = 0.79\%$ with t-statistic of 4.1. For ease of reference, I repeat the results of that regression estimation in each robustness table. As before, we are interested in the λ coefficient on the interaction term $r^e_{ma,i,t} \times LogH_{ma,i,t}$. This coefficient λ estimates the additional return autocorrelation associated with an increase in log hedging demand. Also, as before, for legibility, displayed coefficients are regressions estimates multiplied by 100.

Table 4 illustrates that the main baseline result is robust in various subsamples of the main dataset. Column (1) repeats the main baseline regression from Table 2. The rest of the columns display the different subsample regressions. In Column (2) and Column (3), I verify that the effect exists in both the first half and second half of the sample. One could potentially be concerned that option behavior has shifted over time. For example, dynamic hedging techniques have improved over time, as participants lower the price impact of hedging. Furthermore, as discussed by Campbell, Lettau, Malkiel, and Xu (2001), idiosyncratic volatility has risen over time, which affects the pricing of options on individual stocks. Col-

umn (2) uses the first half of the sample, i.e. before 2005. The effect is slightly weaker in the first half. However, it is still similar in magnitude $\hat{\lambda}^{pre2005} = 0.72\%$ and is still significant with a t-statistic of 2.3. Column (3) uses the second half of the data, i.e. 2005 and later. The economic magnitude increases to $\hat{\lambda}^{post2005} = 0.86\%$ and t-statistic of 3.9. Column (4) uses the subsample of firms with market capitalization in the top 50% of the CRSP universe, i.e. larger firms. In the sample, that corresponds to \$11.8 billion (average market capitalization) and \$1.4 billion (average minimum required market capitalization). The economic magnitude is larger among these firms with $\hat{\lambda}^{bigfirms} = 0.85\%$ (t-statistic = 3.6).

Table 5 examines different subsamples of options. Column (1) repeats the main baseline regression from Table 2 for comparison. Column (2) omits observations from the week of option expiration, which is the Saturday after the third Friday of the expiration month. For example, all exchange-traded equity options that expired in the month of January 2012, expired on Saturday Jan 21, 2012. Since Ni, Pearson, and Poteshman (2005) and Golez and Jackwerth (2012) find unusual stock price behavior near option strike prices at option expiration, we want to verify that my effect is separate from theirs. The estimated coefficient is very similar to the baseline regression (0.77% vs 0.79%), suggesting that my effect is not driven by option expiration week.

Column (3) uses hedging demand computed only from put options and Column (4) uses hedging demand computed only from call options. I find that hedging demand of put options has a stronger effect on return autocorrelation of the underlying stock than hedging demand of call options in both economic magnitude and statistical significance ($\hat{\lambda}^{put} = 1.39\%$ with $t = 8.5$; $\hat{\lambda}^{call} = 0.36\%$ with $t = 1.9$). I attribute this difference to the fact that dynamic hedging of written options is much less prevalent in call options than put options, due to "covered call" and "covered put" strategies. In both strategies, the investor does not dynamically hedge the position by trading the underlying stock.[10] Anecdotal conversations with market participants suggest that covered call strategies account for a meaningful share

---

[10] See footnote 5 for description of "covered call" and "covered put" trading strategies.

of call options written, but the analogous is not true for put options written. This imbalance predicts that the hedging demand from put options will have a stronger effect on the return autocorrelation of the underlying stock than the hedging demand from call options, which is what I find in the data.

Table 6 illustrates that the main baseline result is robust to different estimation methods. In this table, the estimation method varies across columns. Columns (1), (2), and (3) compare variants of the Fama-MacBeth estimation method. Column (1) repeats the main baseline result from Table 2. Column (2) shows a Fama-MacBeth regression with additional controls of beta, log book-to-market ratio, and log spread. These additional controls have only a small effect on economic magnitude and statistical significance of the main baseline result.

Column (3) shows the main baseline regression using the Fama-MacBeth methodology, but with Newey-West standard errors with lag length of 20 trading days (i.e. one trading month). Newey-West standard errors adjust for time series correlation in the error terms, using a triangle (Bartlett) kernel for the correlation structure. The key parameter in the Newey-West procedure is the lag length. In this dataset, lag lengths of 1 trading day to 120 trading days give similar t-statistics, so I simply choose the round number of 20 trading days (i.e. one trading month). In Column (3), the t-statistic actually increases from 4.1 to 4.7 after applying Newey-West standard errors. This is because the errors are slightly negatively correlated up to two trading months, so accounting for the negative autocorrelation improves the t-statistic. In any case, the effect is modest since the dependent variable is (non-overlapping) returns and the usual concern with returns is cross sectional correlation, not time series correlation.

Column (4) shows a panel regression with fixed effects by time and standard errors clustered by time. This panel regression is similar to the Fama-MacBeth estimation procedure. The main difference is that the panel regression more heavily weights cross sections with more observations and cross sections where there is more heterogeneity in the explanatory

variables. In my dataset, this translates to more heavily weighting the more recent observations, as there are more observations per cross section in recent years and more dispersion in *LogH* in recent years (Figure 2). Since hedging demand has a stronger effect in the second half of the sample, this econometric logic predicts the panel estimates should exceed the Fama-MacBeth estimates, which is what I find.

The panel regression results are significantly stronger in both economic magnitude and statistical significance. The economic magnitude of the panel estimate is $\hat{\lambda}^{panel} = 1.35\%$, roughly 1.7x the Fama-MacBeth estimates. The t-statistic of the panel estimate is 5.5, which is significantly larger than the t-statistic for the Fama-MacBeth estimates.

Column (5) shows a panel regression with fixed effects by time and standard errors clustered by *time and firm*. Clustering by firm is similar to Newey-West standard errors. The main distinction is that the Newey-West procedure assumes that the serial correlation decays over time, whereas clustering by firm does not. As discussed in Petersen (2009), if we set the Newey-West lag length to $T - 1$, the formula for the standard error is the same as clustering by firm, modulo a weighting function. Additionally clustering by firm only affects the t-statistics modestly. As before, this is because the general concern for (non-overlapping) returns is usually cross sectional correlation, not time series correlation.

4.3 Effect on Long Horizon Returns

To analyze how hedging affects return autocorrelation at long horizons, I replace the dependent variable in Equation 15 with $\sum_{j=1}^{N} r_{i,t+j}^{e}$, the N period cumulative return.

$$
\begin{aligned}
\sum_{j=1}^{N} r_{i,t+j}^{e} = \quad & b_{0,t} + b_1 \cdot r_{ma,i,t}^{e} + \lambda^N \cdot r_{ma,i,t}^{e} \times LogH_{ma,i,t} \\
& + b_2 \cdot LogH_{ma,i,t} + b_3 \cdot r_{ma,i,t}^{e} \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+N} \quad (16)
\end{aligned}
$$

The coefficient of interest is λ^N, the effect of the interaction of past returns with hedging demand on the future N period cumulative return. Since the dependent variable is now overlapping returns, we must account for serial correlation in the error term. I use Newey-West standard errors with a lag length of 400 trading days, since that is the maximum horizon I consider. For reference, my total dataset has roughly 4400 trading days.

Figure 3 displays two graphs. Figure 3a plots the λ^N estimate at different horizons N. The effect of hedging on long horizon returns is statistically significant up to 100 trading days and the point estimate is positive up to 325 trading days. Figure 3b expresses the same information in the form of an impulse response. It plots the total effect of a positive 1% return shock on the highest and lowest quintile of hedging demand. Because of stocks in general have reversion at this time scale ($b_1 < 0$), the total autocorrelation remains negative for the highest quintile. However, we clearly see that the highest quintile has less reversion than the lowest quintile.

This persistence is in line with the theory. In Section 2, I showed that after a positive shock to the stock price, option writers choose to hold more stock to re-hedge and continue to hold that amount of stock as long as there are no additional price shocks. The effect after option expiration depends on the option buyers' actions. If the option buyer holds the stock after expiration, then the effect will persist. (In my sample, average days to maturity, weighted by open interest, is around 100 trading days.) My data are for exchange-traded equity options, which are "physically delivered" not "cash settled" (CBOE, 2014). Hence, at expiration, the option writer delivers the underlying shares to the option buyer and the persistence I observe suggests that the option buyers continue to hold these delivered shares for some time. The effect eventually decays at long horizons, which suggests that (1) the option buyers eventually sell the stock or (2) arbitrage capital slowly enters the stock market, which increases the risk-tolerance τ of the fundamental investors (Duffie, 2010). The time horizon for this slow moving capital can be roughly comparable to my results. For example, Duffie (2010) discusses the price impact when stocks are deleted from an index. He notes

that other studies have generally found the statistical significance fades around 60 days and in Duffie (2010) Figure 1 the point estimate is still negative at 90 days, which is when his graph ends.

4.4 Systematic vs Idiosyncratic Components of Returns

I examine whether the effect of hedging demand comes from the systematic component of returns or the residual idiosyncratic component of returns. I decompose returns using the CAPM and compute betas using the Scholes and Williams (1977) method. The Scholes and Williams (1977) method accounts for potentially non-synchronous trading, as that can create biased estimates due to measurement error. I compute the betas quarterly and use the estimates from the previous quarter to avoid look-forward bias. I can then decompose the returns into the systematic component and the residual idiosyncratic component:

$$r_i^{e,sys} = \beta \cdot r_{mkt}^e \tag{17}$$

$$r_i^{e,idio} = r_i^e - r_i^{e,sys} \tag{18}$$

Hence, adding in the lagged moving averages, the regression is:

$$
\begin{aligned}
r_{i,t+1}^e = \ & b_{0,t} + b_1^{sys} \cdot r_{ma,i,t}^{e,sys} + b_1^{idio} \cdot r_{ma,i,t}^e \\
& + \lambda^{sys} \cdot r_{ma,i,t}^{e,sys} \times LogH_{ma,i,t} + \lambda^{idio} \cdot r_{ma,i,t}^{e,idio} \times LogH_{ma,i,t} \\
& + b_2 \cdot LogH_{ma,i,t} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}
\end{aligned}
\tag{19}
$$

We are interested in the estimates for the coefficients λ^{sys} and λ^{idio} on the interaction terms $r_{ma,i,t}^{e,sys} \times LogH_{ma,i,t}$ and $r_{ma,i,t}^{e,idio} \times LogH_{ma,i,t}$. These coefficients estimate the additional return autocorrelation associated with an increase in log hedging demand, decomposed into the systematic and idiosyncratic component of returns.

Table 7 displays the results of the decomposition regression. For reference, Column (1) repeats the main baseline result from Table 3, where I found that $\hat{\lambda} = 0.79\%$. Column (2) shows the decomposed results. $\hat{\lambda}^{sys}$ is statistically insignificant, with a t-statistic of 1.1. On the other hand, $\hat{\lambda}^{idio} = 0.90\%$ with a t-statistic of 4.7. The standard errors on the systematic component are large enough that we cannot reject the hypothesis $\hat{\lambda}^{sys} = \hat{\lambda}^{idio}$.

From this decomposition, we can conclude that the idiosyncratic component of returns drives at least part of the effect of hedging demand. As for the systematic component of returns, there are two possibilities: The first is that the systematic component has no effect. The second is that the systematic component has a partial effect, but there simply is not enough statistical power to detect it, as $StdError(\hat{\lambda}^{sys})$ is large in the regression. This standard error is large because the regressions rely on cross sectional variation. Most of the cross sectional variation comes from the idiosyncratic component and so the effect of the systematic component is poorly estimated.

5   Portfolio Sorting Strategy

As an alternate way to quantify the economic magnitude of hedging demand by option writers, I analyze the returns to a portfolio sorting strategy. This methodology expresses the economic magnitude in terms of an annualized alpha, which may be easier to grasp intuitively given the large literature using the portfolio sorts. I sort stocks into quintiles based on $r^e_{ma,i,t} \times LogH_{ma,i,t}$ and analyze the portfolio returns of each quintile. Controlling for the factors RMRF, SMB, HML, UMD, and Short-Term Reversal, I find that stocks in the highest quintile outperform stocks in the lowest quintile by 0.33% alpha per week (t-statistic of 4.1). On an annualized basis, that is equivalent to gross alpha of 18%.

The portfolio sorting methodology also helps address potential concerns with the Fama-MacBeth regressions of Section 4. First, the Fama-MacBeth methodology is sensitive to outliers. In contrast, the portfolio sorting methodology dampens the effect of outliers by sorting

stocks into quintiles. Second, the Fama-MacBeth methodology is sensitive to regression mis-specification. The portfolio sorting strategy (in particular, the per quintile regressions) helps check for potential non-monotonic relationships in the alphas. Another difference is that in the portfolio sorting methodology, one controls for "covariances," instead of "characteristics" (Daniel and Titman, 1997).

The Fama-MacBeth regressions used the interaction term $r^e_{ma,i,t} \times LogH_{ma,i,t}$ to measure the additional return autocorrelation associated with log hedging demand. Hence, I use that interaction term to sort stocks into quintiles. I skip one day between portfolio formation and the portfolio holding period, to avoid potential concerns about different closing times across markets. The (non-overlapping) portfolio holding period is weekly. That is, each week, I sort stocks into quintiles using the value of $r^e_{ma,i,t} \times LogH_{ma,i,t}$ at the close on Thursday. When there are market holidays, I adjust correspondingly. I then measure the returns over the following week.

For quintile i, let $R^e_{Qi,t+1} = R_{Qi,t+1} - R_{f,t+1}$ denote the excess returns for that portfolio. I use simple returns, as that is the standard in this literature. Portfolio returns are value-weighted. I test if the return spread between the highest and lowest quintiles persists after controlling for the standard return factors f_{t+1}, using the regression:

$$R_{Q5,t+1} - R_{Q1,t+1} = \alpha + \beta' \cdot f_{t+1} + \epsilon_{t+1} \tag{20}$$

Table 8, Columns (1) to (3) shows the results of this regression under various factor controls. The constant term estimates the alpha of the portfolio sort strategy. All the alphas are statistically significant with t-statistics exceed 4.0. Harvey, Liu, and Zhu (2016) argue that data mining is a concern in empirical asset pricing and recommend that 3.0 should be the new minimum required t-statistics for tests of expected returns in the cross section. The portfolios here pass that test. Column (1) controls for the Fama-French 3-factor model (RMRF, SMB, HML). Column (2) controls for the Carhart 4-factor model (RMRF,

SMB, HML, UMD). Across both specifications, $\alpha \approx 0.49\%$ weekly (24.5% annualized, before transaction costs) with t-statistics exceeding 6 in magnitude.

In Column (3), I further add the Short-Term Reversal Factor. The sorting variable is $r_{ma,i,t}^e \times LogH_{ma,i,t}$, which is an interaction term involving lagged returns. Hence, it is important to verify that the alpha is robust to controlling for the Short-Term Reversal Factor. By comparison, in the Fama-MacBeth baseline regressions in Section 4, I controlled for r_{ma}^e directly in the regression. After adding the Short-Term Reversal Factor, I find that the weekly alpha is 0.33% with t-statistic of 4.1. In economic magnitude, that is equivalent to 18% annualized alpha, before transactions costs.

While not the main focus of this paper, a back-of-the-envelope calculation suggests the portfolio sorting strategy survives transactions costs. On average, about 75% of the stocks change quintiles each week. The median bid-ask spread is 0.15%. (The mean bid-ask spread is higher, but in my sample that is driven by times such as the fall of 2008.) Therefore, a rough approximation of round-trip transactions costs is $0.225\% = 0.15\% \times 2 \times 75\%$. The portfolio sorting strategy therefore has a net weekly alpha of $0.105\% = 0.33\% - 0.225\%$, which is a net alpha of 5.4% annualized. While this calculation does not include the cost of price impact, more careful trading methods could further reduce transaction costs. For example, here I trade a stock as soon as it changes quintiles. However, that may non-optimal, depending on how often a stock subsequently drifts back into its original quintile in the next couple weeks.

In cross sectional asset pricing, it is also typical to verify that the alphas are monotonic across the quintiles. Unless one has strong theoretical reasons for non-monotonicity, finding non-monotonic alphas suggests a potentially spurious relationship. Table 8, Columns (4) to (8) break down the returns by quintile. Specifically, I run the regression:

$$R_{Qi,t+1}^e = \alpha + \beta' \cdot f_{t+1} + \epsilon_{t+1} \tag{21}$$

We see that the alphas are indeed monotonic across the quintiles, as desired.

6 Identification via Instrumental Variable

The results from Section 4 and Section 5 show that higher log hedging demand is associated with higher return autocorrelation. However, to establish causality, we require exogenous variation. One potential confound is that the earlier results measure news entering option markets first, instead of the effect of hedging. When open interest increases, hedging demand also increases because there are more options to hedge. However, open interest also increases when informed traders purchase options to make leveraged bets. If the information slowly integrates into the stock market, then we will observe return autocorrelation (Hong and Stein, 1999). As discussed in Section 1, this potential confound is related to, but distinct from, research showing that various ratios related to option volume forecast stock returns because news enters the option market first (Easley, O'Hara, and Srinivas, 1998; Pan and Poteshman, 2006; Johnson and So, 2012; Hu, 2014).

To address this potential confound, I use the absolute distance of the underlying stock price to nearest round number as my instrumental variable for hedging demand. My instrument focuses on variation in the convexity/gamma of outstanding options. In contrast, the news-entering-option-markets-first confound focuses on changes in the open interest. My instrument relies on two facts about options: (1) For a put or call option, gamma is highest when the underlying stock price equals the strike price on the option contract. For example, Figure 4a depicts this relationship for an option with strike price $600 that expires in one month. Intuitively, the strike price is the "kink" of the option payoff and so that is where the gamma is the highest. In the limit, on expiration day, gamma is infinite when the underlying price is at the strike price. (2) As a matter of institutional detail, option exchanges only issue options with strike prices that are "round" (e.g. $600, not $613.12). Further, there is more coordination around "rounder" numbers. That is, even though both

$600 and $590 are possible strike prices, the strike at $600 will be a stronger coordination point and have more options issued. For example, Figure 4b shows how Sum(Open Interest) and Sum(Gamma*Open Interest) vary with respect to the strike price for options on Apple Inc. (AAPL) on Dec 31, 2012. We can clearly see spikes in the open interest around the "rounder" numbers of $550, $600, $650, $700, etc.

To use this instrument properly, we need to discard observations where there have been big price movements in the recent past. Specifically, for the two stage least-square (2SLS) estimation, I discard observations where $r^e_{ma,i,t}$ is in the upper and lower 10%. Otherwise, the first stage of the 2SLS estimation is weak for two reasons. First, as the stock price moves away from the strike price, gamma falls to zero (Figure 4a). Second, options are generally issued with strike prices near the current stock price (Figure 4b). Therefore, after a large price increase/decrease, hedging demand is near zero and the instrument is weak. As a result, I discard observations where $r^e_{ma,i,t}$ is in the upper and lower 10%.

Based on a firm's fundamentals, there is no reason to believe that stock prices behave different around round numbers. However, non-fundamental factors could also be a concern. Hence, to verify the exclusion restriction of my instrument, I consider three strands of prior research on stock price behavior near round numbers: (1) Ni, Pearson, and Poteshman (2005) document unusual stock price behavior near option strike prices at option expiration. They attribute it either re-hedging or stock price manipulation by option traders. Re-hedging is precisely my mechanism and so that satisfies the desired exclusion restriction. Stock price manipulation likely only happens on the expiration day, since it is extremely costly to manipulate stock prices for more than brief periods of time. In contrast, my instrumental variable regressions focus the effect of a stock price being close to a round number price, in general. (2) Prior research has shown that stock prices cluster at round decimals, e.g. Harris (1991) which argues that traders focus on discrete price sets to simplify negotiations. These findings do not appear to pose an issue to the exclusion restriction. First, it is unclear why such coordination points would create more return autocorrelation. Second, since the average

stock price is around \$30, my instrumental variable focuses on much coarser round numbers than those studies, which focus on round decimals. In my instrument, \$30 is a "round" number, whereas \$32.25 is not. (3) Some papers have argued that there are behavioral effects near round number prices of stock indices. For example, Donaldson and Kim (1993) argues that multiples of 100 of the Dow Jones Industrial Average are salient reference points and hence create "price barriers." However, other papers argue that these effects are not robust (Ley and Varian, 1994; De Ceuster, Dhaene, and Schatteman, 1998). At the very least, my instrument appears to satisfy the exclusion restriction for the primary confound of news entering option markets first.

To map these institutional details into an instrument, I construct a metric that measures the distance from the stock price to the nearest "round" price. I also need to account for the fact that the distance between successive strike prices rises as with the stock price. For example, for stocks with prices around \$600, open interest clusters around every \$50 (e.g. \$550, \$600, etc.). For stocks with prices around \$60, open interest clusters around every \$5 (e.g. \$55, \$60, etc.). I use the following procedure: First, I normalize prices to have two digits before the decimal point. So, for example \$5.32, \$53.20 and \$532.00 all map to a "normalized price" P^{norm} of \$53.20. Then, my instrumental variable is the distance of the normalized price to the nearest integer multiple of 5:

$$IV_{it} = Abs(P_{it}^{norm} - \text{NearestMultipleOf5}) \qquad (22)$$

To illustrate, on December 31, 2012, the close price of AAPL stock is \$532.17. Therefore, the normalized price is \$53.21. Since the nearest multiple of 5 is \$55, so IV_{it} is $1.79 = 55.00 - 53.21$. One can alternately create multiple instrumental variables using the distance of the normalized price to the nearest multiple of \$10 and \$2.5. However, using multiple instruments puts more onus on the first stage not being weak. If the first stage is weak, multiple instruments actually worsens the weak instruments bias (Chapter 4 of Angrist and

Pischke, 2009).

Because $LogH$ appears as the main effect ($LogH_{ma,i,t}$) and the interaction effect ($r^e_{ma,i,t} \times LogH_{ma,i,t}$), there are two endogenous variables. Hence, my two instruments are $IV_{ma,i,t}$ and $r^e_{ma,i,t} \times IV_{ma,i,t}$. This implementation avoids the econometrically incorrect "forbidden regression" (see textbooks such as Wooldridge, 2001). It is incorrect to only have *one* first-stage fitted value $\widehat{LogH}_{ma,i,t}$ and use it to construct $r^e_{ma,i,t} \times \widehat{LogH}_{ma,it}$ in the second stage. This incorrect procedure produces estimates that are inconsistent, i.e. do not converge to the true parameter. Instead, we must estimate a separate first-stage fitted value for the *entire* interaction term $\widehat{r^e_{ma,i,t} \times LogH}_{ma,i,t}$. Intuitively, the essence of the forbidden regression problem is that if f is a linear projection and g is a non-linear transformation, then $f(g(x)) \neq g(f(x))$. In this situation, the 2SLS estimation is a linear projection and the interaction term is a non-linear transformation.

The 2SLS estimator is

$$
\begin{aligned}
r^e_{i,t+1} \quad = \quad & b_{0,t} + b_1 \cdot r^e_{ma,i,t} + \lambda^{IV} \cdot \widehat{r^e_{ma,i,t} \times LogH}_{ma,i,t} \\
& + b_2 \cdot \widehat{LogH}_{ma,i,t} + b_3 \cdot r^e_{ma,i,t} \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}
\end{aligned} \tag{23}
$$

Table 9 displays the results. As before, we are interested in the coefficient on the interaction term $r^e_{ma,i,t} \times LogH_{ma,i,t}$, which is $\hat{\lambda}^{IV}$. Because the 2SLS estimator is more comparable to a panel regression than a Fama-MacBeth regression, I compare my 2SLS estimates to a panel regression. Column (1) is a panel regression with fixed effects by time and clustering by time. It is the same as Table 6 Column (5). Column (2) is the same panel regression, but with the highest 10% and lowest 10% of $r^e_{ma,i,t}$ discarded. As discussed earlier, this discarding is necessary for instrument strength.

Column (3) shows the results of the 2SLS regression. It is statistically significant and has the same sign the theory in Section 2. Combined with the exclusion restriction, this result corroborates the assertion that hedging demand causally increases return autocorrelation.

The IV estimate $\hat{\lambda}^{IV}$ is also significantly larger than the panel ordinary least squares estimate $\hat{\lambda}^{panel}$. However, the two estimates are not directly comparable. Variation in hedging demand comes from two sources: the gamma of each option and the number of options being hedged. This instrument focuses on variation due to the former whereas the panel regression focuses on both sources of variation combined.

I also formally test the strength of the instrument in the first stage. As a general econometric matter, while the 2SLS estimator is biased, it is consistent so the estimate converges to the true parameter as the sample size increases. However, when the first stage is weak, asymptotic consistency can be surprisingly slow (Bound, Jaeger, and Baker, 1995). A commonly used rule of thumb is that the F-statistic on the excluded instruments should exceed 10 (Angrist and Pischke, 2009). My instrument comfortably passes that test with F-statistics of 266.0 (Kleibergen-Paap test). The F-statistics show that the potential bias due to weak instruments is not a major concern for this setting.

7   Conclusion

In the textbook model of derivatives, two end users who want opposite financial exposures come together and make a side bet. For example, a baker and a farmer who use a derivative contract to take opposite positions about wheat prices. Such end-user-to-end-user trade is a side bet and does not affect the underlying asset. However, when the counterparty is instead a liquidity provider and when the derivative has convex payoffs (as options do), derivatives are not mere side bets. In such situations, the derivative can affect the price dynamics of the underlying asset due to dynamic hedging by the liquidity provider. Previous research has shown that net demand by end users can affect option prices (e.g. Green and Figlewski, 1999; Bollen and Whaley, 2004; Garleanu, Pedersen, and Poteshman, 2009). I show how it can affect the returns and return autocorrelation of the underlying asset as well.

I show that hedging by option writers creates an upward sloping demand curve in the

underlying stock: to remain hedged, option writers must buy stock after the price rises and sell stock after the price falls. If trading the stock has price impact and dynamic hedging is not instantaneous, then this upward sloping demand curve increases the autocorrelation of stock returns. This mechanism is similar to the forced buying/selling by portfolio insurers, which the Brady Report argues was a significant factor in the October 1987 stock market crash (Brady, 1988).

"Hedging demand" quantifies the sensitivity of the option writers' hedge to changes in the underlying price. I document empirically that stocks with higher hedging demand have higher return autocorrelations in the cross section. Using this effect on return autocorrelation, one can formulate a trading strategy that generates positive alpha. I further provide causal identification using an instrumental variable based on the institutional idiosyncrasy that exchange-traded options are struck at round number prices. The robustness of the results supports the view that hedging demand causally increases the return autocorrelation of the underlying stock.

One direct area of future research is to test if options hedging affects other asset classes as well. For example, according to the Bank for International Settlements, the notional amount of interest rate options is roughly ten times the combined notional amount of equity index and individual stock options (BIS, 2014). On the other hand, the interest rates market is more liquid, so dynamic hedging with the underlying asset has less price impact. While many options trade over-the-counter, future research that uses the subset of data on exchange-traded options in other asset classes could likely also apply my instrumental variable (the absolute difference between the underlying price and the nearest round number) for exogenous variation and establish causality.

Another area of future research is to study derivatives other than options. As the upward sloping demand curve comes from the convexity of options, a similar mechanism may apply to hedging other derivatives that are also convex. For example, the theory suggests that credit default swaps could have similar effects. Sellers of credit default swaps often hedge

using the stock of the underlying firm and the value of a credit default swap is convex with respect to the underlying stock. Estimates from the Bank for International Settlements (2014) suggest that the market for credit default swaps is also many times the size of the market for equity-related options. I plan to explore these directions in future research.

References

[1] Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton University Press.

[2] Avramov, D., Chordia, T., & Goyal, A. (2006). Liquidity and autocorrelations in individual stock returns. The Journal of Finance, 61(5), 2365-2394.

[3] Bakshi, G., & Kapadia, N. (2003). Volatility risk premiums embedded in individual equity options: Some new insights. The Journal of Derivatives, 11(1), 45-54.

[4] Bank for International Settlements. (2014). Statistical annex of September 2014 report. BIS Quarterly Review.

[5] Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. The Journal of Political Economy, 637-654.

[6] Bollen, N. P., & Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions?. The Journal of Finance, 59(2), 711-753.

[7] Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association, 90(430), 443-450.

[8] Boyle, P. P., & Vorst, T. (1992). Option replication in discrete time with transaction costs. The Journal of Finance, 47(1), 271-293.

[9] Brady, N. F., & United States Presidential Task Force on Market Mechanisms. (1988). Report of the presidential task force on market mechanisms. U.S. Government Printing Office.

[10] Campbell, J. Y., Grossman, S. J., & Wang, J. (1992). Trading volume and serial correlation in stock returns. Quarterly Journal of Economics, 108 (4), 905-939.

[11] Campbell, J. Y., Lettau, M., Malkiel, B. G., & Xu, Y. (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. The Journal of Finance, 56(1), 1-43.

[12] Cao, J., & Han, B. (2013). Cross section of option returns and idiosyncratic stock volatility. Journal of Financial Economics, 108(1), 231-249.

[13] Carhart, M. M. (1997). On persistence in mutual fund performance. The Journal of Finance, 52(1), 57-82.

[14] Cetin, U., Jarrow, R., Protter, P., & Warachka, M. (2006). Pricing options in an extended Black Scholes economy with illiquidity: Theory and empirical evidence. Review of Financial Studies, 19(2), 493-529.

[15] Chicago Board Options Exchange. (2009). Market maker program. Accessed from website at https://www.cboe.org/publish/mmfaq/mmfaq.pdf during November 2014.

[16] Chicago Board Options Exchange, 2014, Product specifications. Accessed from website at https://www.cboe.com/Products/EquityOptionSpecs.aspx during August 2014.

[17] Conrad, J. (1989). The price effect of option introduction. The Journal of Finance, 44(2), 487-498.

[18] Cox, J. C., Ross, S. A., & Rubinstein, M. (1979). Option pricing: A simplified approach. Journal of Financial Economics, 7(3), 229-263.

[19] Daniel, K., & Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. The Journal of Finance, 52(1), 1-33.

[20] De Ceuster, M. J., Dhaene, G., & Schatteman, T. (1998). On the hypothesis of psychological barriers in stock markets and Benford's Law. Journal of Empirical Finance, 5(3), 263-279.

[21] Detemple, J., & Jorion, P. (1990). Option listing and stock returns: An empirical analysis. Journal of Banking & Finance, 14(4), 781-801.

[22] Donaldson, R. G., & Kim, H. Y. (1993). Price barriers in the Dow Jones industrial average. Journal of Financial and Quantitative Analysis, 28(03), 313-330.

[23] Duffie, D. (2010). Presidential address: Asset price dynamics with slow-moving capital. The Journal of Finance, 65(4), 1237-1267.

[24] Easley, D., O'Hara, M., & Srinivas, P. S. (1998). Option volume and stock prices: Evidence on where informed traders trade. The Journal of Finance, 53(2), 431-465.

[25] Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3-56.

[26] Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. The Journal of Political Economy, 607-636.

[27] Frey, R., & Stremme, A. (1997). Market volatility and feedback effects from dynamic hedging. Mathematical Finance, 7(4), 351-374.

[28] Garleanu, N., Pedersen, L. H., & Poteshman, A. M. (2009). Demand-based option pricing. Review of Financial Studies, 22(10), 4259-4299.

[29] Golez, B., & Jackwerth, J. C. (2012). Pinning in the S&P 500 futures. Journal of Financial Economics, 106(3), 566-585.

[30] Green, T. C., & Figlewski, S. (1999). Market risk and model risk for a financial institution writing options. The Journal of Finance, 54(4), 1465-1499.

[31] Greenwood, R. (2005). Short-and long-term demand curves for stocks: Theory and evidence on the dynamics of arbitrage. Journal of Financial Economics, 75(3), 607-649.

[32] Harris, L. (1991). Stock price clustering and discreteness. Review of Financial Studies, 4(3), 389-415.

[33] Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... And the cross section of expected returns. Review of Financial Studies, 29(1), 5-68.

[34] Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. The Journal of Finance, 54(6), 2143-2184.

[35] Hu, J. (2014). Does option trading convey stock price information?. Journal of Financial Economics, 111(3), 625-645.

[36] Hull, J. (2012). Options, Futures, and Other Derivatives. Prentice Hall.

[37] Johnson, T. L., & So, E. C. (2012). The option to stock volume ratio and future returns. Journal of Financial Economics, 106(2), 262-286.

[38] Lakonishok, J., Lee, I., Pearson, N. D., & Poteshman, A. M. (2007). Option market activity. Review of Financial Studies, 20(3), 813-857.

[39] Lehmann, B. N. (1990). Fads, martingales, and market efficiency. Quarterly Journal of Economics, 105 (1): 1-28.

[40] Leland, H. E. (1985). Option pricing and replication with transactions costs. The Journal of Finance, 40(5), 1283-1301.

[41] Ley, E., & Varian, H. R. (1994). Are there psychological barriers in the Dow-Jones index?. Applied Financial Economics, 4(3), 217-224.

[42] Merton, R. C. (1973). The theory of rational option pricing. Bell Journal of Economics and Management Science, 4: 141-83.

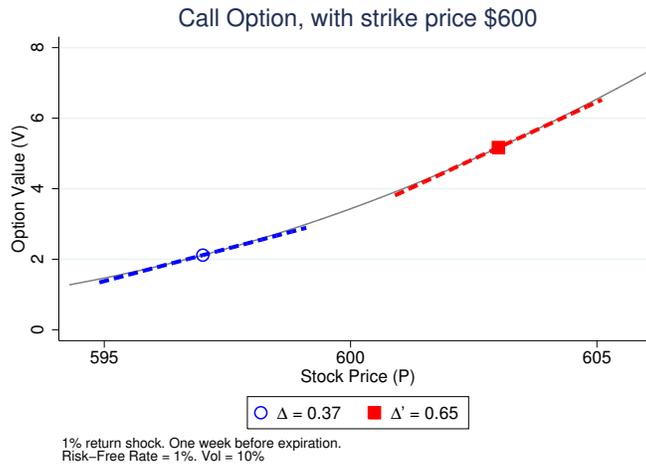[43] Nagel, S. (2012). Evaporating liquidity. Review of Financial Studies, 25(7), 2005-2039.

[44] Ni, S. X., Pearson, N. D., & Poteshman, A. M. (2005). Stock price clustering on option expiration dates. Journal of Financial Economics, 78(1), 49-87.

[45] Ni, S. X., Pearson, N. D., Poteshman, A. M., & White, J. S. (forthcoming). Does Option Trading Have a Pervasive Impact on Underlying Stock Prices?, Review of Financial Studies.

[46] Pan, J., & Poteshman, A. M. (2006). The information in option volume for future stock prices. Review of Financial Studies, 19(3), 871-908.

[47] Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. Review of Financial Studies, 22(1), 435-480.

[48] Platen, E., & Schweizer, M. (1998). On feedback effects from hedging derivatives. Mathematical Finance, 8(1), 67-84.

[49] Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. The Journal of Finance, 39(4), 1127-1139.

[50] Shleifer, A. (1986). Do demand curves for stocks slope down?. The Journal of Finance, 41(3), 579-590.

[51] Scholes, M., & Williams, J. (1977). Estimating betas from nonsynchronous data. Journal of Financial Economics, 5(3), 309-327.

[52] Schönbucher, P. J., & Wilmott, P. (2000). The feedback effect of hedging in illiquid markets. SIAM Journal on Applied Mathematics, 61(1), 232-272.

[53] Sorescu, S. M. (2000). The effect of options on stock prices: 1973 to 1995. The Journal of Finance, 55(1), 487-514.

[54] Wooldridge, J. M. (2001). Econometric analysis of cross section and panel data. MIT press.

[55] Wurgler, J., & Zhuravskaya, E. (2002). Does arbitrage flatten demand curves for stocks?. The Journal of Business, 75(4), 583-608.

[56] Yang, D. C. (2020). Disagreement and the option stock volume ratio. Working Paper.

Figure 1: How Hedged Option Writing Affects Stock Prices

The figures below overview the theory in Section 2. Dynamic hedging by option writers creates an upward sloping demand curve for the underlying stock and return autocorrelation.

(a) Upward Sloping Demand Curve
As the stock price rises, purchase more of underlying stock to remain hedged.

Call Option, with strike price $600

Option Value (V)

Stock Price (P)

○ Δ = 0.37   ■ Δ' = 0.65

1% return shock. One week before expiration.
Risk–Free Rate = 1%. Vol = 10%

(b) Price Impact of Re-Hedging Causes Return Autocorrelation

Hypothetical Impulse Response

Cumulative Return

Time

—— High H   - - - - H = 0
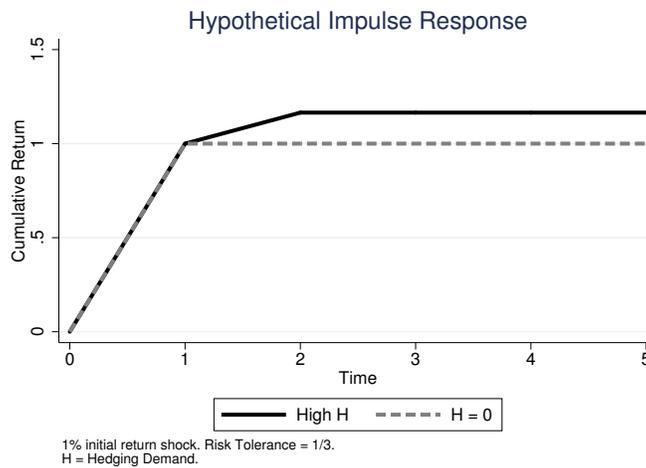
1% initial return shock. Risk Tolerance = 1/3.
H = Hedging Demand.

Figure 2: Time Series and Cross Sectional Properties of Log Hedging Demand

In my empirical work, I define hedging demand as

$$H_{it} \quad := \quad \frac{1\% P_{it}}{\text{SharesOut}_{it}} \cdot \sum_{k}(100 \cdot OI_{itk} \cdot \Gamma_{itk})$$

where P_{it} is the stock price for stock i at time t, OI_{itk} is the open interest of option k, and Γ_{itk} is the convexity/gamma of option k. Below, I plot the log of hedging demand ($LogH_{it}$), which is the key explanatory variable in my analysis. While the time series patterns in Panel (a) convey a general sense of log hedging demand, my analysis relies on the cross sectional variation in log hedging demand in Panel (b). See also Table 2 for summary statistics by quintile of hedging demand.

(a) Average Log Hedging Demand
Left Figure: full time series. Right Figure: Jan 2012 to end of dataset
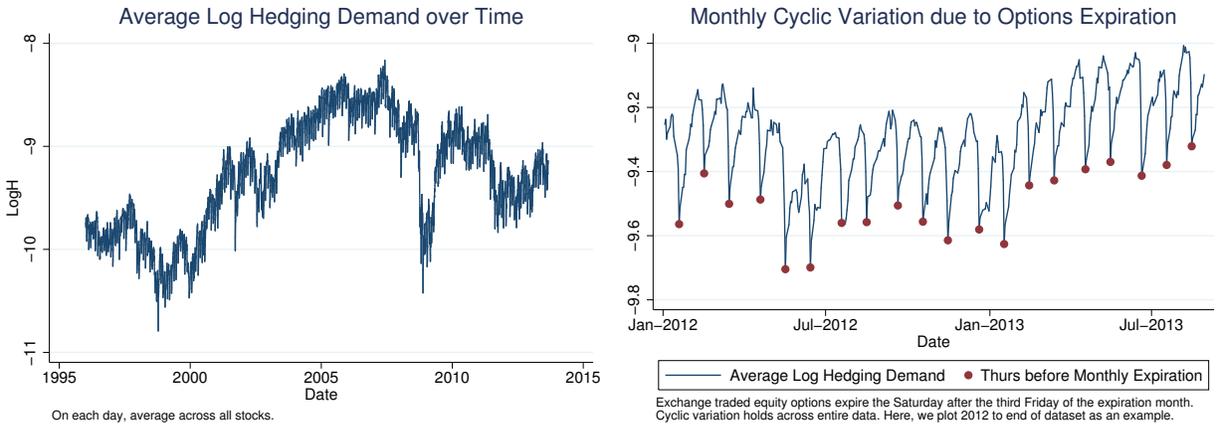
(b) Cross Sectional Dispersion

44
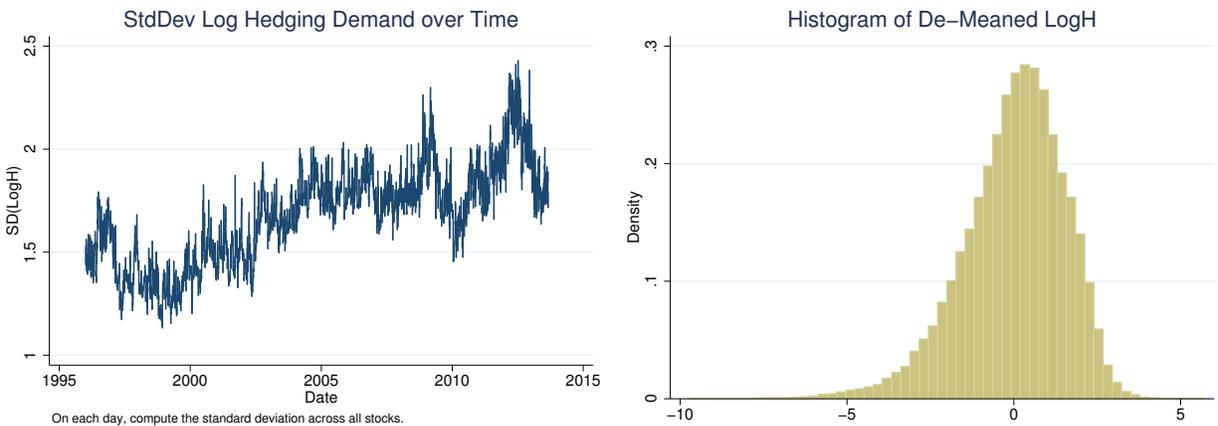
Figure 3: Impact at Long Horizons

The graph below illustrates the relationship between log hedging demand and long horizon returns. The dependent variable is the long horizon return $\sum_{j=1}^{N} r_{i,t+j}^e$. We are interested in the coefficient λ^N on the interaction term $r_{ma,i,t}^e \times LogH_{ma,i,t}$, where r^e denotes excess log returns and $LogH$ denotes log hedging demand.

$$\sum_{j=1}^{N} r_{i,t+j}^e = b_{0,t} + b_1 \cdot r_{ma,i,t}^e + \lambda^N \cdot r_{ma,i,t}^e \times LogH_{ma,i,t}$$
$$+ b_2 \cdot LogH_{ma,i,t} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+N}$$

Newey-West standard errors with a lag length of 400 trading days. A panel regression with clustering by time and firm and time fixed effects gives similar results.

(a) λ^N Estimates at Different Horizons
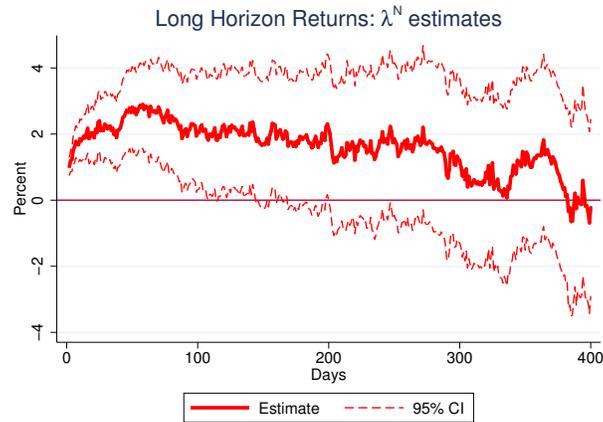
(b) Impulse Response: Highest vs Lowest Quintile of Hedging Demand
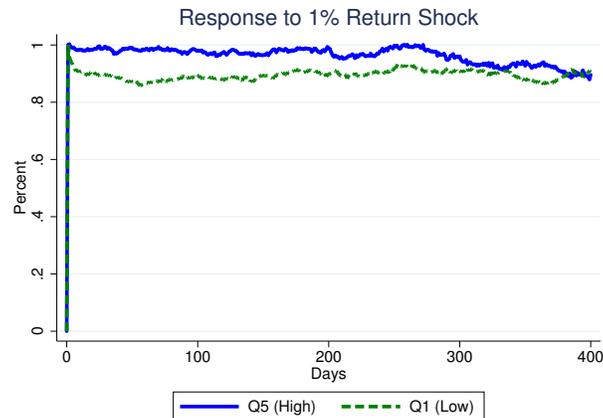
Figure 4: IV, Distance to Nearest Round Number: First Stage Graphs

My instrument for hedging demand is the absolute difference between the underlying stock price and the nearest round number. This instrument is based on two facts: (1) put and call gamma is highest near the strike price and (2) as a matter of institutional detail, exchange-traded options are struck at round number prices (e.g. $600 as opposed to $613.12).

(a) Example: Gamma is Highest Near the Strike Price
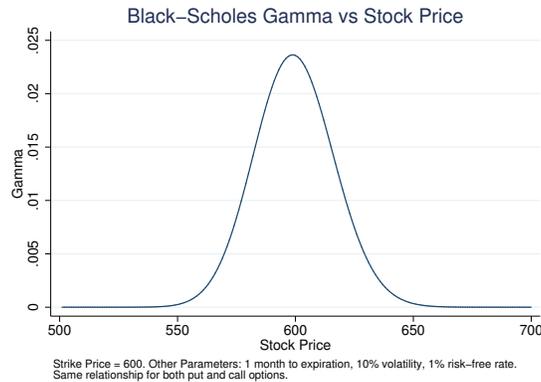
(b) Example: Options are Struck at Round Number Prices
On December 31, 2012, Apple Inc. (AAPL) closed at a stock price of $532.17.
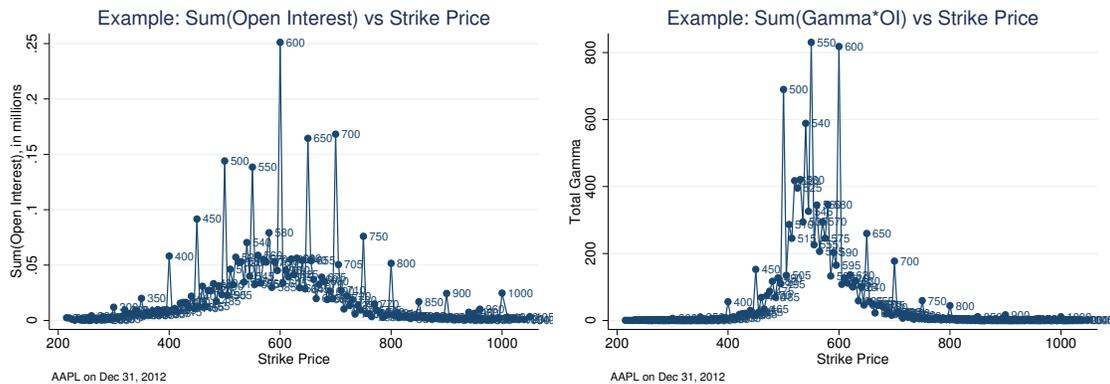
(c) Binscatter on Full Dataset

46

Table 1: List of Key Variables

This table lists the key variables used in this paper.

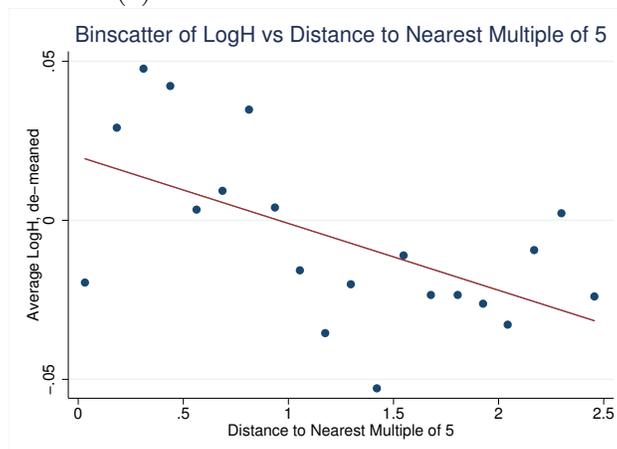| Variable	Symbol	Description
Hedging Demand	H	Proxy for *"If the underlying stock price rises by 1%, what fraction of the shares outstanding must the option writer additionally purchase to remain hedged?"* See Section 2.4 for formal definition.
Moving Average	Subscript ma	Lagged moving average over the last five trading days.
Option Delta (subscript k for individual option and agg for portfolio)	Δ	First derivative of an option (or portfolio of options) with respect to underlying price, i.e. $\frac{\partial V}{\partial S}$. For individual options, I use the estimates computed by Option Metrics, which is based on the Cox, Ross, and Rubinstein (1979) binomial tree model.
Option Gamma (subscript k for individual option and agg for portfolio)	Γ	Second derivative of an option (or portfolio of options) with respect to underlying price, i.e. $\frac{\partial^2 V}{\partial S^2}$. For individual options, I use the estimates computed by Option Metrics, which is based on the Cox, Ross, and Rubinstein (1979) binomial tree model.
Open Interest	OI	Total number of options contracts that are currently not settled ("open").
Simple Return	R	Net simple returns, including dividends.
Log Return	r	$r = Log(1 + R)$.
Excess Log Return	r^e	Log return in excess of the log risk-free rate. For the risk-free rate, I use the one-month Treasury bill.
Market Cap	$MktCap$	Total equity market capitalization, using close prices.
Dollar Volume	$DVolume$	Total dollar volume for a given stock on a given day.
Share Price	P	Unadjusted price per share on the close of each day.
CAPM Beta	β	Estimated using the Scholes and Williams (1977) method to account for potentially nonsynchronous trading. Computed quarterly and lagged by one quarter to avoid look-forward bias.
Book/Market Ratio	$Book/Mkt$	(Book Equity)/(Market Capitalization). Lagged by two quarters to ensure that the accounting data was already publicly available on each date. Following Fama and French (1993), Book Equity = Stockholder's Equity + Deferred Taxes and Investment Tax Credit (if available) - Preferred Stock.
Bid-Ask Spread	$Spread$	(Ask - Bid)/(Midpoint)
Turnover	$Turnover$	(Share Volume)/(Shares Outstanding)

Table 2: Summary Statistics

This table displays summary statistics of the main variables. I express hedging demand H in percent for legibility. I include $LogH_{ma}$ and de-meaned $LogH_{ma}$, since I use those values in the text. The average of the de-meaned $LogH_{ma}$ is not exactly 0 because I de-mean cross sectionally, as opposed to over the entire dataset. The variable $H \cdot \frac{ShareOut}{Volume}$ is hedging demand scaled by volume, instead of shares outstanding. I abbreviate market capitalization as "MktCap" and dollar volume as "DVolume." The variable "Volatility" refers to the daily return volatility for the individual stocks in the sample. See Table 1 for full list of variable definitions. $N = 4,326,618$.

(a) Full Dataset

	mean	sd	p10	p50	p90
Year	2004.8	5.24	1997	2005	2012
Num Firms per Day	987.0	124.3	828	1027	1105
H, in %	0.028	0.044	0.0012	0.011	0.073
$H \cdot \frac{ShareOut}{Volume}$, in %	4.20	5.72	0.28	2.13	10.5
$LogH_{ma}$	-9.22	1.64	-11.3	-9.11	-7.24
$LogH_{ma}$, de-meaned	-0.0063	1.56	-2.00	0.15	1.82
MktCap, in billions	7.94	21.0	0.40	2.25	17.6
DVolume, in millions	49.3	122.5	1.40	14.0	119.1
Turnover, in %	0.84	0.90	0.16	0.55	1.81
Spread, in %	0.65	1.11	0.026	0.15	1.87
β	1.08	8.84	0.23	1.00	2.07
Book/Market Ratio	0.60	0.55	0.18	0.49	1.08
Volatility, in %	2.37	1.56	1.03	1.97	4.10

(b) By Quintile of Hedging Demand

	1	2	3	4	5
H, in %	0.00	0.01	0.01	0.03	0.09
$H \cdot \frac{ShareOut}{Volume}$, in %	0.80	1.75	3.11	5.38	10.05
$LogH_{ma}$	-11.53	-9.91	-9.05	-8.31	-7.27
$LogH_{ma}$, de-meaned	-2.31	-0.69	0.16	0.91	1.94
MktCap, in billions	2.78	4.01	6.27	13.47	13.27
DVolume, in millions	9.28	18.73	33.21	72.74	113.69
Turnover, in %	0.52	0.65	0.78	0.91	1.34
Spread, in %	0.81	0.69	0.63	0.58	0.56
β	1.00	1.06	1.05	1.12	1.18
Book/Market Ratio	0.71	0.63	0.58	0.53	0.53
Volatility, in %	2.37	2.31	2.32	2.35	2.51

mean coefficients

This table displays the baseline results. We are interested in the coefficient λ on the interaction term $r^e_{ma,i,t} \times LogH_{ma,i,t}$, where r^e denotes excess log returns and $LogH$ denotes log hedging demand. This coefficient λ estimates the additional return autocorrelation associated with an increase in hedging demand. Subscript ma (e.g. $LogH_{ma}$) denotes a lagged moving average over the last five trading days. I cross-sectionally de-mean $LogH_{ma,i,t}$ and the controls $X_{ma,i,t}$, so the b_1 coefficient measures the return autocorrelation for average firm. Data are 1996 to 2013, daily. See Table 1 for full list of variable definitions.

$$
\begin{aligned}
r^e_{i,t+1} = {} & b_{0,t} + b_1 \cdot r^e_{ma,i,t} + \lambda \cdot r^e_{ma,i,t} \times LogH_{ma,i,t} \\
& + b_2 \cdot LogH_{ma,i,t} + b_3 \cdot r^e_{ma,i,t} \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}
\end{aligned}
$$

For legibility, displayed coefficients are regressions estimates multiplied by 100. Regression methodology is Fama-MacBeth. For brevity, I do not display the non-interacted controls ("main effects").

VARIABLES	(1)	(2)	(3)	(4)	(5)
r^e_{ma}	-1.39***	-1.47***	-2.32***	-2.68***	-3.15***
	(-3.1)	(-3.3)	(-5.1)	(-5.8)	(-7.0)
$r^e_{ma} \times LogH_{ma}$		0.61***	0.94***	0.72***	0.79***
		(3.5)	(5.5)	(3.8)	(4.1)
$LogH_{ma}$		-0.01***	-0.01***	-0.01***	-0.01***
		(-3.4)	(-3.5)	(-5.0)	(-5.3)
$r^e_{ma} \times LogMktCap_{ma}$			-1.10***	-1.64***	-1.90***
			(-5.9)	(-4.6)	(-5.2)
$r^e_{ma} \times LogDVolume_{ma}$				0.63*	0.61*
				(1.9)	(1.8)
$r^e_{ma} \times LogPrice_{ma}$					0.59
					(1.4)
Observations	4,326,618	4,326,618	4,326,618	4,326,618	4,326,618
R-squared	0.017	0.028	0.043	0.056	0.068
Number of groups	4,434	4,434	4,434	4,434	4,434
Main effects	Y	Y	Y	Y	Y

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Baseline Results, Using Different Subsamples of the Main Dataset

This table estimates the baseline results using different subsamples of the main dataset. We are interested in the coefficient λ on the interaction term $r^e_{ma,i,t} \times LogH_{ma,i,t}$, where r^e denotes excess log returns and $LogH$ denotes log hedging demand. Column (1) is the baseline regression from Table 3 Column (5), which uses the full dataset. Column (2) uses the first half of the dataset, i.e. before 2005. Column (3) uses the second half of the dataset, i.e. 2005 and later. Column (4) uses the subsample of firms with market capitalization in the top 50% of the CRSP universe on each day, i.e. larger firms. Data are 1996 to 2013, daily. See Table 1 for full list of variable definitions.

$$
\begin{aligned}
r^e_{i,t+1} \;=\; & b_{0,t} + b_1 \cdot r^e_{ma,i,t} + \lambda \cdot r^e_{ma,i,t} \times LogH_{ma,i,t} \\
& + b_2 \cdot LogH_{ma,i,t} + b_3 \cdot r^e_{ma,i,t} \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}
\end{aligned}
$$

Regression methodology is Fama-MacBeth. For legibility, displayed coefficients are regressions estimates multiplied by 100. For brevity, I do not display the non-interacted controls ("main effects").

VARIABLES	(1)	(2)	(3)	(4)
r^e_{ma}	-3.15***	-3.83***	-2.45***	-3.46***
	(-7.0)	(-6.4)	(-3.7)	(-6.8)
$r^e_{ma} \times LogH_{ma}$	0.79***	0.72**	0.86***	0.85***
	(4.1)	(2.3)	(3.9)	(3.6)
$LogH_{ma}$	-0.01***	-0.01***	-0.01***	-0.00
	(-5.3)	(-4.7)	(-2.8)	(-0.9)
$r^e_{ma} \times LogMktCap_{ma}$	-1.90***	-2.61***	-1.16**	-0.71*
	(-5.2)	(-4.9)	(-2.3)	(-1.8)
$r^e_{ma} \times LogDVolume_{ma}$	0.61*	0.63	0.59	0.00
	(1.8)	(1.3)	(1.3)	(0.0)
$r^e_{ma} \times LogPrice_{ma}$	0.59	2.12***	-1.00**	0.16
	(1.4)	(3.2)	(-2.0)	(0.3)
Observations	4,326,618	1,993,710	2,332,908	2,784,652
R-squared	0.068	0.068	0.069	0.076
Number of groups	4,434	2,256	2,178	4,434
Main effects	Y	Y	Y	Y
Dataset	All	Pre-2005	Post-2005	Size>p50

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Baseline Results, Using Different Subsamples of Options

This table estimates the baseline results using different subsamples of the options. We are interested in the coefficient λ on the interaction term $r^e_{ma,i,t} \times LogH_{ma,i,t}$, where r^e denotes excess log returns and $LogH$ denotes log hedging demand. Column (1) is the baseline regression from Table 3 Column (5), which uses the full dataset. Column (2) omits days during the week of option expiration. Column (3) only uses data from put options to calculate hedging demand. Column (4) only uses data from call options to calculate hedging demand. Data are 1996 to 2013, daily. See Table 1 for full list of variable definitions.

$$
\begin{aligned}
r^e_{i,t+1} \;=\;& b_{0,t} + b_1 \cdot r^e_{ma,i,t} + \lambda \cdot r^e_{ma,i,t} \times LogH_{ma,i,t} \\
& + b_2 \cdot LogH_{ma,i,t} + b_3 \cdot r^e_{ma,i,t} \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}
\end{aligned}
$$

Controls: log market capitalization, log dollar volume, log share price, and their interactions with returns. For legibility, displayed coefficients are regressions estimates multiplied by 100. Regression methodology is Fama-MacBeth.

VARIABLES	(1)	(2)	(3)	(4)
r^e_{ma}	-3.15***	-3.89***	-3.18***	-3.23***
	(-7.0)	(-7.6)	(-7.1)	(-7.2)
$r^e_{ma} \times LogH_{ma}$	0.79***	0.77***		
	(4.1)	(3.5)		
$r^e_{ma} \times LogH^{put}_{ma}$			1.39***	
			(8.5)	
$r^e_{ma} \times LogH^{call}_{ma}$				0.36*
				(1.9)
$LogH_{ma}$	-0.01***	-0.01***		
	(-5.3)	(-4.2)		
$LogH^{put}_{ma}$			-0.01***	
			(-6.2)	
$LogH^{call}_{ma}$				-0.01***
				(-4.6)
Observations	4,326,618	3,311,810	4,266,788	4,314,979
R-squared	0.068	0.068	0.068	0.068
Number of groups	4,434	3,393	4,434	4,434
Controls + xterms	Y	Y	Y	Y
Dataset	All	Omit Expir Wk	Only Puts	Only Calls

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Baseline Results, Using Different Estimation Methods

This table displays the effect of varying the estimation method. The key coefficient of interest is λ. Column (1) is the baseline regression from Table 3 Column (5), which uses the Fama-Macbeth procedure ("FM" in the table). Column (2) is a Fama-MacBeth regression with the additional controls of beta, log book-to-market ratio, and log spread. Column (3) the baseline regression using Fama-MacBeth with Newey-West standard errors, lag length of 20 trading days. Column (4) is a panel regression with fixed effects by time and standard errors clustered by time. Column (5) clusters standard errors by time and firm. The r-squared of the panel regressions does not include the time fixed effect and hence one cannot directly compare it to the r-squared of the Fama-MacBeth regressions. Data are 1996 to 2013, daily. See Table 1 for full list of variable definitions.

$$
\begin{aligned}
r^e_{i,t+1} =\ & b_{0,t} + b_1 \cdot r^e_{ma,i,t} + \lambda \cdot r^e_{ma,i,t} \times LogH_{ma,i,t} \\
& + b_2 \cdot LogH_{ma,i,t} + b_3 \cdot r^e_{ma,i,t} \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}
\end{aligned}
$$

Controls: log market capitalization, log dollar volume, log share price, and their interactions with returns. For legibility, displayed coefficients are regressions estimates multiplied by 100.

VARIABLES	(1)	(2)	(3)	(4)	(5)
r^e_{ma}	-3.15***	-3.69***	-3.15***	-3.96***	-3.96***
	(-7.0)	(-9.1)	(-7.0)	(-4.4)	(-4.1)
$r^e_{ma} \times LogH_{ma}$	0.79***	0.68***	0.79***	1.35***	1.35***
	(4.1)	(3.4)	(4.7)	(5.5)	(5.0)
$LogH_{ma}$	-0.01***	-0.01***	-0.01***	-0.02***	-0.02***
	(-5.3)	(-4.9)	(-5.2)	(-7.5)	(-7.4)
Observations	4,326,618	4,326,618	4,326,618	4,326,618	4,326,618
R-squared	0.068	0.112	0.068	0.001	0.001
Number of groups	4,434	4,434	4,434	4,434	4,434
Controls + xterms	Y	Y + Extras	Y	Y	Y
StdErr Cluster	FM	FM	FM+NW	Time	Time+Firm
Fixed Effect	FM	FM	FM	Time	Time

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Decomposing Returns into Systematic and Idiosyncratic Components

This table displays the effect of decomposing returns into a systematic and an idiosyncratic component. Column (1) shows the main baseline result from Table 3 Column (5) for reference. Column (2) shows the decomposed results. Using CAPM β, I decompose r_i^e into the systematic component $(r_i^{e,sys} = \beta \cdot r_{mkt}^e)$ and the residual idiosyncratic component $(r_i^{e,idio} = r_i^e - r_i^{e,sys})$; in the regressions, I use lagged moving averages of this decomposition. Data are 1996 to 2013, daily. See Table 1 for full list of variable definitions.

$$
\begin{aligned}
r_{i,t+1}^e = \ & b_{0,t} + b_1^{sys} \cdot r_{ma,i,t}^{e,sys} + b_1^{idio} \cdot r_{ma,i,t}^{e,idio} \\
& + \lambda^{sys} \cdot r_{ma,i,t}^{e,sys} \times LogH_{ma,i,t} + \lambda^{idio} \cdot r_{ma,i,t}^{e,idio} \times LogH_{ma,i,t} \\
& + b_2 \cdot LogH_{ma,i,t} + b_3 \cdot r_{ma,i,t}^e \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}
\end{aligned}
$$

Controls: log market capitalization, log dollar volume, log share price, and their interactions with returns. For legibility, displayed coefficients are regressions estimates multiplied by 100. Regression methodology is Fama-MacBeth.

VARIABLES	(1)	(2)
r_{ma}^e	-3.15***	
	(-7.0)	
$r_{ma}^{e,sys}$		-0.09
		(-0.0)
$r_{ma}^{e,idio}$		-3.55***
		(-8.4)
$r_{ma}^e \times LogH_{ma}$	0.79***	
	(4.1)	
$r_{ma}^{e,sys} \times LogH_{ma}$		14.45
		(1.1)
$r_{ma}^{e,idio} \times LogH_{ma}$		0.90***
		(4.7)
$LogH_{ma}$	-0.01***	-0.01***
	(-5.3)	(-2.6)
Observations	4,326,618	4,326,618
R-squared	0.068	0.087
Number of groups	4,434	4,434
Controls + xterms	Y	Y

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Alpha of Portfolio Sorting Strategy (Weekly Holding Periods)

This table displays the returns to a portfolio sorting strategy, where I sort stocks into quintiles based on the interaction term $r^e_{ma,i,t} \times LogH_{ma,i,t}$. The constant term estimates the portfolio alpha. I skip one day between portfolio formation and the portfolio holding period, to avoid potential concerns about different closing times across markets. Portfolios are held for one week and are value-weighted. In Columns (1) to (3), the dependent variable is the return of the highest quintile minus the lowest quintile, $R_{Q5,t+1} - R_{Q1,t+1}$. In Columns (4) to (8), the dependent variable is the return of quintile i, $R_{Q_i,t+1}$. Data frequency is weekly so each 0.01% weekly alpha translates into roughly 0.50% annual alpha, before transactions costs. Newey-West standard errors with lag length of four trading weeks (one trading month).

VARIABLES	(1) Q5-Q1	(2) Q5-Q1	(3) Q5-Q1	(4) Q1	(5) Q2	(6) Q3	(7) Q4	(8) Q5
RMRF, in %	0.11*	0.14**	-0.01	1.05***	0.98***	0.94***	0.96***	1.05***
	(1.8)	(2.1)	(-0.1)	(39.1)	(60.0)	(50.5)	(40.3)	(26.7)
SMB, in %	-0.00	-0.02	-0.04	-0.02	-0.12***	-0.16***	-0.12***	-0.06
	(-0.0)	(-0.2)	(-0.5)	(-0.3)	(-3.9)	(-4.4)	(-2.7)	(-1.0)
HML, in %	-0.00	0.05	0.14	0.13**	0.24***	0.21***	0.27***	0.28***
	(-0.0)	(0.4)	(1.3)	(2.5)	(7.0)	(5.7)	(5.4)	(3.7)
UMD, in %	-0.00	0.10*	0.16**	-0.16***	-0.04*	-0.02	0.07**	-0.00
	(-0.0)	(1.7)	(2.5)	(-4.6)	(-1.7)	(-0.6)	(2.4)	(-0.1)
ST Reversal, in %			0.50***	-0.21***	-0.11***	0.00	0.09**	0.28***
			(7.9)	(-5.6)	(-3.7)	(0.1)	(2.2)	(6.6)
Constant	0.50***	0.48***	0.33***	-0.02	0.04	0.08***	0.13***	0.31***
	(6.4)	(6.1)	(4.1)	(-0.4)	(1.1)	(2.6)	(3.8)	(5.3)
Observations	916	916	916	916	916	916	916	916
R-squared	0.014	0.022	0.179	0.802	0.880	0.887	0.846	0.801

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9: IV, Distance to Nearest Round Number: 2SLS

This table displays the results of the two stage least square (2SLS) estimation. My instru-
mental variable (IV) is the absolute difference between the price of the underlying stock and
the nearest round number; see Section 6 for more details. Since the two endogenous variables
are $LogH_{ma,i,t}$ and $r^e_{ma,i,t} \times LogH_{ma,i,t}$, the two instruments are $IV_{ma,i,t}$ and $r^e_{ma,i,t} \times IV_{ma,i,t}$.
This implementation avoids the econometrically incorrect "forbidden regression" (see text-
books such as Wooldridge, 2001). The coefficient of interest is λ^{IV}. Column (1) is a panel
OLS regression. It is the same as Table 6 Column (5). Column (2) is the same panel re-
gression, but with the highest 10% and lowest 10% of $r^e_{ma,i,t}$ discarded. As discussed in the
main text, this is necessary for instrument strength. Column (3) shows the results of the
2SLS regression. In general, for 2SLS estimation, r-squared can be negative. F-statistic on
excluded instruments: 697.7 (Cragg-Donald test), 266.0 (Kleibergen-Paap test).

$$
\begin{aligned}
r^e_{i,t+1} \;=\; & b_{0,t} + b_1 \cdot r^e_{ma,i,t} + \lambda^{IV} \cdot \widehat{r^e_{ma,i,t} \times LogH}_{ma,i,t} \\
& + b_2 \cdot \widehat{LogH}_{ma,i,t} + b_3 \cdot r^e_{ma,i,t} \times X_{ma,i,t} + b_4 \cdot X_{ma,i,t} + \epsilon_{i,t+1}
\end{aligned}
$$

Controls: log market capitalization, log dollar volume, log share price, and their interactions
with returns. For legibility, displayed coefficients are regressions estimates multiplied by 100.

VARIABLES	(1)	(2)	(3)
r^e_{ma}	-3.96***	-0.56	-0.41
	(-4.4)	(-1.1)	(-0.7)
$r^e_{ma} \times LogH_{ma}$	1.35***	1.27***	23.07***
	(5.5)	(5.9)	(2.7)
$LogH_{ma}$	-0.02***	-0.01***	-0.00
	(-7.5)	(-6.7)	(-0.1)
Observations	4,326,618	3,456,809	3,456,809
R-squared	0.001	0.000	-0.005
Number of groups	4,434	4,434	4,434
Controls + xterms	Y	Y	Y
StdErr cluster	Time	Time	Time
Fixed effect	Time	Time	Time
Method	Panel OLS	Panel OLS	2SLS

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1