Derivatives Valuation Based on Arbitrage: The Trade is Crucial

Stephen Figlewski*

Derivatives valuation has strong theoretical support because models are derived from the principle that arbitrage between the derivative and its underlying will eliminate riskless profits and drive the market price to the model value. "No-arbitrage" is invoked routinely whenever a new pricing model is developed. But real world market prices are determined by trades, not by theories. In this talk, I discuss how different the arbitrage trade is for different markets and different models and I review articles from the literature that illustrate how limits to the arbitrage trade have affected the way derivatives theory gets into prices in practice. © 2016 Wiley Periodicals, Inc. Jrl Fut Mark 37:316–327, 2017

1. INTRODUCTION

It is an honor and a pleasure to have this opportunity to discuss with you today one of the major themes that has motivated my research over the last 40 years. It also allows me to shamelessly promote a number of my earlier papers as I describe the evolution of these ideas through a series of articles that all focus on arbitrage, the trade that connects our theoretical derivatives valuation models to the real world markets for futures, options and other contingent claims.

Arbitrage, or more precisely, the lack of profitable arbitrage opportunities, is what drives our pricing models. Derivatives theory is based on the principle that when the same payoff can be produced in two or more different ways in the market, arbitrage (true arbitrage!) must force them all to be priced exactly the same. In theoretical modeling, this is often summarized in the shorthand expression: "We assume no-arbitrage," meaning that prices within the model allow no possibility of profitable riskless arbitrage. Since theoretical model values are derived from riskless positions, they do not depend on how risk averse traders are. Models that satisfy no-arbitrage exhibit the extremely useful property of "risk-neutral" pricing.

In theoretical modeling, no-arbitrage is a mathematical condition that the model specification must satisfy. If there are two ways to achieve exactly the same payoff on a given future date (e.g., futures expiration), but they have different costs today, an arbitrageur buys the cheaper position and sells (short) the higher priced one. This trade locks in the initial price difference as a riskless profit. Because everyone would like to make free money with no risk, the forces to eliminate arbitrage opportunities are very strong, even if there are only a

Stephen Figlewski is Professor of Finance at New York University Stern School of Business, New York, New York. I thank the participants at the 2016 China Derivatives Markets Conference for their feedback on the ideas presented here. Financial support from the NASDAQ Educational Foundation is gratefully acknowledged.

*Correspondence author, New York University Stern School of Business, 44 West 4th Street, Suite 9-160, New York, NY 10012-1126. Tel: 212-998-0712, Fax: 212-995-4220, e-mail: sfiglews@stern.nyu.edu

Received June 2016; Accepted July 2016

The Journal of Futures Markets, Vol. 37, No. 4, 316–327 (2017)

Published online 7 September 2016 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/fut.21806

small number of arbitrageurs. Their trading eliminates the profit and pushes the market price for the derivative instrument into line with the no-arbitrage based theoretical value.

In the real world, arbitrage is a trade. Given real world transactions costs, risks of various kinds, finite capital, and other limits to arbitrage, the trade is more costly and risky than arbitrage in the theoretical framework, and may not be possible at all in some cases. As a mathematical property of a theoretical model, no-arbitrage pricing must hold exactly. The underlying assumption is that if a mispricing were to arise, unlimited arbitrage trading would immediately drive the market price back to fair value. In case the actual arbitrage trade is not possible in the market (e.g., if the underlying asset is not traded), derivatives models revert to "The Law of One Price," which holds that two identical items cannot have different prices simply because buyers will choose to buy in the low price market while sellers will want to sell in the high price market, making the only equilibrium one where the same price prevails in both markets.

This brings to mind one of my favorite quotes, that sums up the difference between theory and practice:

In theory, there is little difference between theory and practice. But in practice, there is.

The origin of this quote is unclear. Some attribute it to Yogi Berra, who made so many other wise observations, but somehow it does not sound very Yogi-like. Wikipedia suggests it may have come from Jan L. van de Snepscheut, an interesting character who was a physics professor at Caltech when he attacked his wife with an ax and set his house on fire. She escaped with their children, but he perished in the blaze. So we may never know his exact words on theory and practice, although the outcome of his plan might be a good illustration of the main point in the quote.

The idea I want to explore in my comments today is that while no-arbitrage is a strict mathematical property in our theoretical models, in the real world it is a trade, and no-arbitrage pricing only works in practice if there are active arbitrageurs doing the trade. If the trade is impossible, the Law of One Price is normally much too weak to force market prices to their theoretical values. How closely we should expect market prices to match those from a theoretical pricing model depends on how easy or hard the arbitrage trade is in practice and how aggressively arbitrageurs pursue it. Limits to arbitrage can explain a lot of the theoretical "mispricing" in the real world.

Impediments to the arbitrage trade include transactions costs of various kinds. Transaction costs induce a band around the theoretical value within which there is an apparent arbitrage profit but it is not large enough to cover the cost of doing the trade. Above the upper arbitrage point, the derivative can be sold and replicated with a cheaper trade in the cash market. The price difference minus the transactions cost is the arbitrageur's profit.

When the derivative's price is too low, the arbitrage trade typically requires buying the cheap contract and replicating its payoff by selling short the underlying asset. But short sales are generally constrained in various ways and may not be allowed at all. This causes an asymmetry in the arbitrage bounds. Overpricing of the derivative relative to its theoretical value may bring out active arbitrage trading, while the trade of buying the derivative and shorting the underlying to eliminate an underpricing of the same size is constrained. The force of arbitrage in this direction is much weaker and can allow larger and more persistent underpricing of the derivative.

Real world arbitrage is not costless, nor is it riskless. Indeed, almost nothing is ever riskless in the real world. One problem is that trading is not continuous—markets close at night and on weekends—so options and products with option-like features cannot be perfectly replicated using the continuously rebalanced delta hedging arbitrage strategy envisioned by Black and Scholes. Moreover, most models require parameter inputs that are not known exactly and estimated parameters from a finite sample are subject to sampling error. More problematical is that real world parameters drift over time, so even if we knew volatility over the past year exactly, we would still not have a perfect forecast of volatility over the remaining life of an option traded today. Finally, real world arbitrageurs face margin requirements (that can lead to liquidity problems even for a "locked arbitrage" trade), counterparty risk, position limits, and so on, all of which impede unlimited arbitrage.

The key point is that limits to arbitrage limit how well market prices will satisfy theoretical derivatives valuation models.

2. EXAMPLES OF ARBITRAGE-BASED DERIVATIVES MODELS

How easy or hard the underlying arbitrage trade is in the real world varies tremendously across the range of derivatives models. In this section, I will look at some familiar examples that illustrate the point.

There are four general categories:

- Case 1: The arbitrage trade can be done with relatively low cost and low risk.
- Case 2: The arbitrage trade can be done only in theory. It is costly and risky in real world markets.
- Case 3: The arbitrage trade is impossible in theory, but the Law of One Price can be invoked.
- Case 4: The arbitrage trade is impossible and the Law of One Price does not apply.

Let us begin with Case 4, to get it out of the way.

<u>Case 4</u>: Arbitrage is impossible and the Law of One Price does not apply.

One of the easiest arbitrage strategies is the "cash and carry" trade in forwards and futures. The underlying asset is purchased in the spot market and stored ("carried") until contract maturity. Selling the forward or futures contract (i.e., taking a short position, which will require delivering the underlying asset on the expiration day) locks in a riskless profit on the trade (or loss, if the futures price is too low).

Think about a forward or futures contract based on an underlying that cannot be bought today and carried to expiration. If no "cash and carry" arbitrage trade is possible, there is no trade that connects the forward/futures price directly to today's spot price. Models based on arbitrage do not apply.

Examples:

- the VIX index and other financial and economic variables, like an overnight interest rate, inflation, weather, and so on (Obviously, an intangible statistic like the Consumer Price Index or heating degree days cannot be bought in the spot market and carried over time);
- capital structure, as in "structural" models of default (Arbitrage in such models would require trading the whole firm against a portfolio of its stock and bonds, maybe with continuous rebalancing of the proportions);
- notional underlyings for "real" options, such as embedded future investment opportunities (in this case, the underlying may not even exist in the present).

Key point: If there is no riskless arbitrage trade, we can no longer assume risk neutral valuation.

Case 1: Arbitrage is Easy in Theory and in Practice

The classic example of an arbitrage-based derivatives model that works in practice is the cost of carry model for forwards and futures just mentioned, in which the underlying security or commodity is purchased in the spot market, the future is sold, and the resulting hedged position is carried to futures expiration.
Let,

S = current underlying spot price;
F = forward price;
T = maturity of forward;
r = risk free interest rate.

$$\text{Fair value}: \; F(T) = \mathrm{S}e^{rT} \tag{1}$$

If transactions costs are low, the underlying can be easily bought and also sold short, there are no cash payouts on the underlying, and there are no other impediments to trade, this pricing model holds quite closely in the real world.

Example: Gold futures arbitrage. Gold can be bought easily and stored for very low cost. The only substantial expense in the cash and carry trade for gold is interest on the funding, which is represented by the exponential term in equation (1). There is also an active "gold lease" market where physical gold can be borrowed to facilitate short selling (generally for just a few basis points annually). The bid-ask spread plus typical commissions to trade spot gold are well under 1%. In practice, the gold futures price does not deviate far from the theoretical value shown above.

And yet even though this trade locks in the price at which today's gold purchase will be liquidated on expiration day, it is not free of risk. If gold prices rise and the future or forward is marked to market, the arbitrageur will need a source of cash to meet margin calls. Inability to finance the position all the way to maturity can force early termination and large losses despite the guaranteed return if the position could be carried through expiration.

Key point: Even with a locked arbitrage, collateral requirements can still introduce risk from uncertain financing.

Arbitrage and the Introduction of Stock Index Futures: the introduction and development of the stock index futures contract in the 1980s illustrates very well how the arbitrage trade is what connects theory to practice in the derivatives world. This is a clear case where "in theory there is little difference between theory and practice, but in practice there is."

The pricing model for a financial future like the one on the S&P 500 introduced in the spring of 1982 is a simple modification of equation (1) to account for the payout of cash dividends during the lifetime of the contract.

$$\text{Fair value}: \; F(T) = \mathrm{S}e^{(r-d)T} \tag{2}$$

where d is the annual dividend yield. This model applies to stock index and foreign exchange futures and forwards.

For FX contracts, the foreign interest rate takes the place of the dividend yield. In that case, the underlying is just money, homogeneous, and with no physical storage costs. The arbitrage bounds around the theoretical forward exchange rate given in equation (2) are very tight and violations are small.

The stock index futures market is tightly arbitraged today, but not when the contracts were first introduced. It was my first look at how limits to arbitrage can affect derivatives pricing in the real world.

S&P 500 index futures began trading in April 1982. The contract is based on the S&P index, which references the value of a specific portfolio containing 500 large capitalization stocks in precise proportions. The contract size was originally set at $500 times the index, which was trading at an average level of around 120 during the 6-month period I examined in a paper at that time.

In the first few months of trading, index futures prices deviated far from the theoretical model values, and were too low most of the time. The price bias even brought out academic journal articles explaining why, in theory, the equilibrium index futures price should always be below the equation (2) value.[1] I felt the main problem was simply that in 1982 the arbitrage trade was very hard to do, and argued as much.[2] Here is an excerpt from a table in that paper showing how S&P 500 index futures were mispriced in the early months of trading in 1982.

Table I shows that the standard deviation of the mispricing was about 1 index point around a negative mean. These figures translated to negative average excess returns to the cash and carry trade of over 6 percent annually during the first 3 months. The return standard deviation of more than 15% shows how large the deviations from theoretical fair value were, corresponding to theoretically riskless arbitrage returns. Notice that as the market developed, after a few months of trading the average underpricing went away in the second 3-month period and the standard deviation also dropped substantially.

Enforcement of theoretical valuation in the real world marketplace depends on arbitrageurs actively pursuing trading profits. But at this time in this new market, the arbitrage trade was hard to do. There was no electronic order transmission to the specialists on the stock exchange, so to execute a "program trade," as the transaction of trading an entire portfolio at the same time came to be called, it was necessary for runners to hand deliver the necessary orders to the stock exchange trading posts, in principle for 500 stock trades. Not surprisingly, the infrastructure to do this quickly, efficiently, and in size was not in place at the time, so mispricings in the market grew large at times without drawing out large offsetting trades from arbitrageurs.

Execution problems were especially difficult on the short side, when arbitrageurs had to buy underpriced futures, sell short the stocks in the index, invest the proceeds at the riskless interest rate, and carry the position to futures expiration. Arbitrage to offset underpricing in the futures market was harder, riskier, and more expensive than the cash and carry trade, which meant that average deviations below fair value could be larger and longer lasting than those above fair value.

In October 1984, stock index futures went the other way and became seriously overpriced. This persisted for many weeks. As just described, overpricing is harder to explain than underpricing, yet in this period the cash and carry arbitrage trade offered "riskless" alpha that was more than 10% annually at times. What was going on?

This was shortly after Fischer Black had quit M.I.T. and taken a job at Goldman Sachs, starting what has since become a significant two-way migration between academia and Wall

[1] In one, the (somewhat farfetched) reasoning turned on the difference between tax treatment of futures versus stock short sales for positions held open for more than 6 months. See Cornell and French (1983).
[2] Figlewski (1984).

TABLE I
Excerpt from Figlewski, "Explaining the Early Discounts on Stock Index Futures: The Case for Disequilibrium" Financial Analysts Journal, July–August 1984

			1st Half	*2nd Half*
		Full Sample 6/1/82 to12/20/82	*6/1/82 to 9/14/82*	*9/15/82 to 12/20/82*
% Discount from futures fair	Mean	−0.18	−0.26	−0.11
value	Std dev	1.02	1.11	0.92
Excess return to cash and carry	Mean	−3.14	−6.14	−0.19
arbitrage (annualized %)	Std dev	12.89	15.16	9.37

Street. One evening in November 1984, I was leaving my NYU office to go home and I ran into Fischer on the street. We started talking and at one point he asked, "So, Steve, what do you think about the way index futures are priced these days?" My answer was: "I don't get it. Why aren't you guys making the overpricing go away?"

Fischer's response was that they were trying to. But the trading desk had a maximum position size allowed for any individual trade. They had put on the cash and carry trade up to the desk's limit. Then, they raised the limit and traded up to the new maximum. But no prudent firm will take an arbitrarily large position on anything, no matter how attractive it looks, so they could not do any more. This was my first insight into the fact that even in a market where arbitrage is not too hard and well-established arbitrageurs are doing the trade, they still have limited capital. When a strong imbalance between supply and demand in the market pushes a derivative's price well away from fair value, arbitrageurs may simply not have enough capital to fully offset the mispricing. I never understood why demand was so much stronger than supply in the fall of 1984, but I did come to understand why arbitrage did not make the resulting futures mispricing go away.

The year 1987 brought the rapid rise and fall of the portfolio insurance strategy, which revealed in another way the crucial differences between arbitrage in theory and arbitrage in the real world. In a portfolio insurance strategy, an investor holding a portfolio of stocks would like to buy a protective put option in order to place a floor under the portfolio value no matter how far stock prices might fall. But such puts were not available in the market. The portfolio insurance strategy amounts to creating the put's payoff synthetically by dynamically trading between the stock portfolio and bonds. If the market falls, the strategy reduces stock market exposure, eventually to zero when the chosen floor level is hit.

This idea works perfectly in theory, but on October 19, 1987, when a large number of investors following the strategy tried to reduce their stock market exposure simultaneously, it produced a cascade of selling that quickly exhausted all sources of liquidity from buyers. Falling prices induced accelerating attempts from the portfolio insurers to liquidate stocks as portfolio values dropped through their intended floors. We learned at that point that, in practice, market liquidity is limited and there can be a huge difference between owning an option and trying to replicate that option by dynamic trading of the underlying asset.

By 1989, the necessary infrastructure was in place and stock index futures arbitrage had become so well-established that it was known on the Street as a "commodity trade," meaning that since any large firm could do it with no difficulty, arbitrage returns were no more than was justified by the effort and risk exposure the trade entailed. George Sofianos showed this in a paper that I liked so much I put it in as the lead article in the very first issue of the Journal of Derivatives, in 1993.[3] Using real-time intraday data to analyze all of the index futures arbitrage trades over a 6-month period in 1989, he found that by that time arbitrage opportunities were small and they disappeared within 3 minutes on average; trades were not carried through futures expiration, but rather, the average trade was unwound within 24 hours; and the excess returns to index futures arbitrage were only about 5% annually.

The key points to focus on in the historical development of arbitrage in this market are

- **even for linear contracts (forwards and futures), theoretical pricing models can perform badly or fail entirely if the arbitrage trade they are based on is hard to do in the actual market;**
- **limits to arbitrage can arise from impediments to trading like transactions costs, but also because arbitrageurs simply do not have infinite capital;**
- **asymmetric costs or risks lead to asymmetric mispricing;**
- **mispricing should diminish as the technology to execute arbitrage trades improves.**

The cash and carry trade is easy. It only requires setting up the arbitrage position at the beginning and carrying it (no longer than) to futures maturity. Options arbitrage is a lot harder.

Case 2: Arbitrage is Easy Only in Theory

Black and Scholes' option pricing model was an intellectual breakthrough because they figured out how to do a riskless arbitrage dynamically in continuous time.
This required

- asset returns follow a lognormal diffusion with known and nonstochastic volatility;
- continuous rebalancing of the hedge proportions is possible;
- there are no transactions costs.

The fact that none of these conditions hold in the real world led me to consider how much risk remained (in theory) in a delta hedge with only daily rebalancing, and how wide a market maker's bid-ask spread would have to be to cover trading costs on average. In a simulation study of these questions, I found that[4]

For 1-month at the money calls with lognormal returns, volatility of 15%, and daily rebalancing:

- rebalancing required trading about five shares for every share in the initial hedge;
- residual standard deviation that remained unhedged was about 6.5%;
- marketmaker transactions costs for the hedge (at 1987 levels) were about 30 b.p. on average;
- the bid-ask spread had to be more than 30% of the option's price to give an even chance of covering the hedging costs.

Key point: Delta hedging with real world hedging strategies and trading costs is expensive and far from risk free even for a market maker.

[3] Sofianos (1993).
[4] Figlewski (1989).

Index Arbitrage and Implied Volatility: In the 1980s, it was widely accepted that implied volatility was the market's volatility forecast, and since the market is very efficient, this was the most informed forecast available. But option prices must match model values if (and only if) arbitrageurs are strong enough to offset supply and demand imbalances in the market.

The most actively traded option in the 1980s was on the S&P 100 index, often known by its ticker symbol OEX. The OEX is very like the S&P 500 index, in that it contains only large cap stocks with weighting by market capitalization. And as with the S&P 500, dynamically trading the 100 stocks continuously to arbitrage a mispriced option was not feasible in practice for any reasonable cost.

This raised the question: was implied volatility still an efficient forecast without active arbitrageurs?

Canina and Figlewski (1993) tested this proposition in the following familiar regression:

$$\text{Realized}\,\sigma_t = \alpha + \beta\ \text{IV}_t + u_t. \tag{3}$$

The realized volatility from date t through option expiration was regressed on the implied volatility forecast. If IV is the true expected value of future realized volatility, the regression constant α should be 0, the β coefficient should be 1.0, and the R^2 should be high.

What we found was the following (standard errors in parentheses):

$$\underset{(0.012)}{\alpha = 0.136} \qquad \underset{(0.050)}{\beta = 0.022} \quad R^2 = 0.002.$$

In the 1980s, implied volatility in this active options market contained virtually no information about future realized volatility on the underlying index!

In view of the results we have seen earlier, this strong, and at the time controversial, result is actually not surprising. If no one does the arbitrage trade that the theoretical model calls for, there is no trading to offset the natural variations in supply and demand from non-arbitrageurs and to drive the market price toward the theoretical fair value.

Our result has often been misinterpreted, so let me emphasize here that we always considered the OEX to be an extreme polar case, because model-based arbitrage was exceptionally difficult. In other markets, where the arbitrage trade is easier to execute, and in later time periods after the infrastructure to do sophisticated program trading was in place, one should expect implied volatility to contain valid information about future volatility, and quite probably more than can be extracted from historical returns data alone. This is what has been found in many subsequent studies, although it virtually never turns out that α and β are 0 and 1.0, respectively.

Key point: How closely the market obeys the model depends on how actively traders do the arbitrage trade. Option replication for a broad stock index is especially hard to do efficiently and in size.

Option pricing models require the returns volatility of the underlying stock, but in practice volatility must be forecasted and this might not be easy. Green and Figlewski (1999) explored the performance of delta hedged options positions in the context of real world market making, with volatility estimated from past returns and daily rebalancing of hedge proportions (but no transactions costs).

TABLE II
Standard Deviation of Annualized Excess Return on Daily Rebalanced Delta Hedges, 1971–1996

Option	*Volatility*	*S&P 500*	*Deutschemark Exchange Rate*
ATM call	Min RMSE estimate	33.4	36.3
	Realized	12.4	12.2
ATM put	Min RMSE estimate	41.6	47.8
	Realized	15.4	14.6
OTM call	Min RMSE estimate	55.3	60.9
	Realized	22.0	22.9
OTM put	Min RMSE estimate	69.5	84.6
	Realized	27.2	30.3

From Green and Figlewski (1999).

The market maker was assumed to write options on each of four major asset classes every day, using historical volatility estimated from past returns data. The sample size, and in some cases a decay factor to downweight observations as they aged, were set equal to the values that would have provided the minimum root mean squared error volatility predictions in the past. The sample covered returns over the period 1971–1996 for the S&P 500, the Deutschemark (DM) exchange rate, 3-month LIBOR, and 10-year U.S. Treasury yields. The options were priced and subsequently delta hedged using the Black–Scholes model.

Table II gives a selection of results we found for writing 1 month at the money calls and puts on the S&P 500 index and the DM. We considered results with estimated volatility and for comparison, the same trades priced and hedged using the volatility that was actually realized over the option's life.

Hedge performance was even worse than this for options on 3-month LIBOR and 10-year Treasury yields.

The key points from this study are

- **even minimum RMSE estimates of volatility from past data have large forecast errors, which translate into large delta hedging errors;**
- **rebalancing only once a day leads to substantial hedging risk, even with the true volatility;**
- **out of the money options are even harder to hedge than at the money options.**

The paper I presented at this conference yesterday provides a final striking example, where in theory, it would seem that no-arbitrage, or more correctly in this case the Law of One Price, should be satisfied easily, yet it fails in practice.[5] A very well-known principle in call valuation is that early exercise of an American call on a non-dividend paying stock is irrational. In theory, exercise prior to expiration gives up the option's remaining time value, and this is always positive, so it is better to sell the call in the market. The right to exercise early is an option that has no value, so an American call and a European call should be priced the same. The extremely important corollary to this is that the Black–Scholes European call formula can also be used for American calls. I, and many of us at this conference, have been teaching this to students for years.

[5]Battalio, Figlewski, and Neal (2016).

The problem is that the theory does not hold in the real world, because of illiquidity in the options market. We have found that in the U.S. stock options markets, for an in the money call (or put) nearing expiration, there is almost never a bid in the market even as high as intrinsic value. It is impossible to liquidate an option early without losing its remaining time value, and early exercise is generally the only way to get even the option's intrinsic value. In practice, optimal liquidation of an American call position before expiration will probably require exercise. And, contrary to theory, an American call that allows early liquidation through exercise offers a higher expected return, and should be priced higher, than a European call with the same terms.

The key point is that market illiquidity is a real world impediment to arbitrage that can invalidate theoretical prescriptions for pricing and optimal investor behavior, including option exercise.

Case 3: The arbitrage is impossible in both theory and practice, but...

"Plain vanilla" (Black–Scholes) theoretical options arbitrage is risky and costly in practice (yet everyone uses the model anyway, with ad hoc adjustments). Research has shown that the assumption of a constant volatility lognormal diffusion process is too simple to capture the behavior of stock returns. We seem to need stochastic volatility, and jumps at least in returns but maybe also in volatilities.

Here is the Heston model extended to allow jumps in returns, one of the simpler modern specifications for the returns process.

$$dS/S = (r - d)dt + \sqrt{V_t}\, dW_t^1 \; + \; \left(e^J - 1\right)dN_t, J \sim \text{Normal}(m, v) \tag{4}$$

$$dV_t = K(\theta - V_t)dt + \sigma_v \sqrt{V_t}\, dW_t^2, \;\; E\left[dW_t^1 dW_t^2\right] = \rho dt.$$

The returns equation specifies that the percentage stock return follows a jump-diffusion with time-varying diffusive volatility and lognormal jumps with mean m and variance v. dN_t is a Poisson process with arrival intensity λdt. The variance equation specifies that variance follows a mean-reverting square root process with its own diffusion dW_t^2 that has correlation ρdt with the diffusive return shocks.

The model has seven parameters that all must be "calibrated" from market returns: λ, m, v, K, θ, σ_v, and ρ.

This puts tremendous pressure on the accuracy of the assumed model specification. How accurate are parameter estimates in a sample of reasonable size? How much information can the sample contain about low probability tail events? For example, when jumps are only expected to occur a couple of times a year at most, is it reasonable to think that the data will allow a clear choice of the lognormal for jumps versus, say, the exponential distribution, or a clear test of the constraint that up and down jumps have the same variance? How much might some or all of these parameters drift over time?

These concerns prompt the natural question: what is the arbitrage trade if an option based on this underlying asset is mispriced according to the model? The answer is easy: there is none. The problem is that there is only one underlying asset to use in the hedged portfolio, but three different sources of risk. There is no riskless hedge between the underlying and its options even in theory. And if there is no risk free arbitrage trade, we cannot expect arbitrageurs to push market prices aggressively toward model values. Moreover, without arbitrage, risk neutral valuation disappears and the equations in (4) need to be modified to include risk premia.

As our theoretical models become more realistic, but inevitably more complicated, the arbitrage trade that is required to connect theory to practice has become harder and harder, to the point that it is not possible at all with the current generation of derivatives models. It raises the question of what support for such models we can expect from real world data, and also how much faith we can have in apparent empirical support for models derived from arbitrage trades that cannot be executed in the real world.

3. WHAT SHOULD A DERIVATIVES RESEARCHER DO?

This discussion has focused on the real world problems in executing the arbitrage trades that are assumed to connect pricing theory to market practice and to drive price determination in derivatives markets. We have seen that they can be severe even for the most basic trades. How should these lessons be incorporated in our research designs?

First, it is important to try to understand the limits to arbitrage in each market. "Absence of arbitrage" is not a (0,1) condition in the real world. Risk in "arbitrage" trades varies across instruments, markets, and time periods, and we should expect to see effects of this variation reflected in market prices.

We sometimes imagine that one should be, and can be, pure in theorizing, starting from first principles of rational choice and optimal behavior, building a pricing model in a stylized market setting, and then "taking the model to the data" in empirical testing. I would argue almost the opposite position: Learning what investors actually do and then modeling that is likely to lead to better empirical results and more understanding of the markets than modeling what von Neumann-Morgenstern expected utility maximizers ought to be doing.

"Mispricing" relative to arbitrage-based model prices is not just noise. It is to be expected when there are limits to arbitrage. The deviations can provide valuable information about the factors driving supply and demand from non-arbitrageurs. In trying to understand what is happening in a financial market, it can be very valuable, especially for theorists, to talk with traders and market professionals. They understand their markets quite well and one often learns that much, if not all, of an apparent mispricing uncovered by empirical research can be understood not as a trading opportunity that has been overlooked by investors, but rather as the result of rational behavior in the face of limits to arbitrage.

Recognizing the problems in connecting theory to real world practice, we should not expect better hedging and pricing performance than such a theoretical model can reasonably deliver. One saving feature of many option models is that they tend to differ less in their estimates of delta than in model prices, meaning adequate hedging performance may be obtained with a fairly broad range of models.

In trying to adapt models to the real world, one should note that complexity tends to make a model look good in sample, but performance degrades sharply when an overly complex model is taken out of sample. Along the same lines, it does not make sense to obsess over mathematical fine points of densities and returns processes: they are all wrong. That is, market prices arise out of the interplay of supply and demand from an ever-changing mass of individual and idiosyncratic traders. There is no reason at all that the resulting returns should consistently follow any particular stochastic process, particularly not one whose main appeal is that it is familiar and mathematically convenient for us.

Finally, recognizing that when the arbitrage trade is hard, limited, expensive, risky, uncertain, or impossible, it makes sense to give up on risk neutral pricing that arises solely from arbitrage, and instead build in risk premia formally. In some cases, it may OK simply to assume a required cost of capital that includes a suitable risk premium (and to explore alternative values for the premium in robustness tests). Or one might assume investors have utility functions of a plausible kind (constant relative risk aversion with a risk aversion coefficient consistent with holding 60% of wealth in stocks and 40% in bonds perhaps?).

One direction in which it will certainly be appropriate to extend our theoretical models is to incorporate behavioral factors beyond the von Neumann-Morgenstern axioms. Deviations from the benchmark arbitrage-based model price can be used to explore the effects of investor expectations and risk preferences that are excluded from Black–Scholes.

REFERENCES

Battalio, R., Figlewski, S., & Neal, R. (2016). Early exercise boundary violations in American-style options: The rule, not the exception. Working paper, February 2016. Available at http://ssrn.com/abstract=2544915.

Canina, L., & Figlewski, S. (1993). The informational content of implied volatility. Review of Financial Studies, 6, 659–681.

Cornell, B., & French, K. (1983). The pricing of stock index futures. Journal of Futures Markets, 3, 1–14.

Figlewski, S. (1984). Explaining the early discounts on stock index futures: The case for disequilibrium. Financial Analysts Journal, 40, 43–47+67.

Figlewski, S. (1989). Options arbitrage in imperfect markets. Journal of Finance, 44, 1289–1311.

Green, T. C., & Figlewski, S. (1999). Market risk and model risk for a financial institution writing options. Journal of Finance, 54, 1465–1499.

Sofianos, G. (1993). Index arbitrage profitability. Journal of Derivatives, 1, 6–20.