Regulatory costs and risk neutrality
© Malcolm Kemp, Nematrian Limited, 18 October 2014

Abstract

Some commentators have argued that regulatory compliant pricing cannot be risk neutral and that this has implications for exit prices and marking-to-market. We argue in this paper that the demise of risk neutrality as a central component of instrument valuation is not a necessary consequence of regulatory change, at least not if subtleties in the concept of risk neutrality and how regulations might influence it are fully explored.

Introduction

In their paper "Regulatory costs break risk neutrality", Kenynon and Green (2014) argue that if different market participants receive different dividend rates from the same stock then there is no market-wide risk neutral measure that is common to all market participants. They note that different (bank) trading businesses have different holding costs because of different regulations such as the Capital Requirements Directive IV (in the EU), the Dodd-Frank Act (in the USA) and BCBS liquidity and bilateral margin requirements. They argue that this "*means that all valuations are private, in the sense that they can be derived from idiosyncratic risk neutral measures (that is, the valuations are local to the individual pricing institution). Executable screen prices are components of value, not valuations by themselves, because of these idiosyncratic and asymmetric costs*".

If this assertion was the last word on the topic then there must have been some myopic behaviour across the entire financial sector well before the regulations mentioned in their paper were conceived. Risk neutral concepts have played an important part in derivative pricing theory for at least the last 20 years. We argue that there are some additional subtleties involved in risk neutrality and on the impact of regulation that are not fully addressed in their paper. The purpose of this paper is to elucidate some of these subtleties. The subtleties are relevant not just to banking but to other parts of the financial services industry, including insurance.

A market-wide risk neutral pricing framework

Kenynon and Green (2014) refer to the definition of risk neutrality given in Shreve (2004), which in turn refers to the martingale concept. The same result can be derived more simply and in a manner that is possibly less likely to obscure some of the mathematical features implicit in a risk neutral measure by using the following simpler definition of risk neutrality as described in e.g. Kemp (2009).

Suppose we have a set of contingencies, $X_1, \dots, X_n$, all mutually exclusive which together span all possible outcomes at some future point in time, T. Suppose instrument Q_i pays out 1 (in a given numeraire, which would usually but not always be a currency such as the Euro) at time T (and nothing at any other time) if X_i occurs and 0 otherwise. A portfolio of such instruments that has k_i of instrument Q_i can then be described by an n dimensional vector $q = (k_1, \dots, k_n)^T$. In this notation, a portfolio consisting of just 1 unit of instrument Q_i can be expressed as $q_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ where the 1 is the i'th element of the vector.

Suppose we have a real-valued function $V(q)$, which we will call the 'value' of q (although we won't for the moment identify how 'value' might be interpreted). Suppose $V(q)$ satisfies axioms of additivity, scalability (sometimes instead called linearity) and uniqueness, i.e. $V(q)$ is single valued for any given q and satisfies $V\big(k(q_A + q_B)\big) = kV(q_A) + kV(q_B)$ for all portfolios $q_A, q_B \in \mathbb{R}^n$ and for all values of $k \in \mathbb{R}$. Here $q_A + q_B$ and kq are to be interpreted as portfolios whose vector description is derived using the standard axioms of vector calculus. These latter axioms also include additivity and scalability. Adherence to these axioms implies that $V(q)$ satisfies the Principle of No Arbitrage. For a 'pure' arbitrage to exist we must have $V(q) + V(-q) > 0$ for some q. However, the above axioms imply $V(q) + V(-q) = V(q - q) = 0.V(q) = 0$.

Suppose also that I is a portfolio of instruments that pays out 1 in all circumstances, i.e. $I = (1, \dots, 1)^T$, that $V(I) \neq 0$ and that $V(q_i) \geq 0$ as might seem reasonable for most plausible definitions of value that might have a sensible economic meaning. Then algebraic manipulation allows us to define a 'risk neutral' probability p_i of event X_i occurring and a (continuously compounded) 'risk-free' or 'risk neutral' discount rate, r (for the given numeraire and for term T), given the valuation function V, as follows:

$$p_i \equiv p_{i|V} = \frac{V(q_i)}{V(I)}$$

$$r_T \equiv r_{T|V} = -\frac{\log V(I)}{T} \quad \text{so that} \quad e^{-r_T T} = V(I)$$

For the p_i to correspond to a probability measure we need $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$. The latter is presumed and the former is trivially satisfied since:

$$\sum_{i=1}^{n} p_i = \frac{1}{V(I)} \sum_{i=1}^{n} V(q_i) = \frac{1}{V(I)} V\left(\sum_{i=1}^{n} q_i\right) = \frac{1}{V(I)} V(I) = 1$$

The most obvious conclusion we can draw from this analysis is that risk neutral probabilities are not inherently unique. Instead they depend on how we define 'value'. The essential point of Kenyon and Green (2014) is that different institutions have different ways of interpreting 'value' which depend on their internal position including the impact that regulatory regimes have on them. They come from a (bank) front office perspective and in effect are seeking to identify the 'right' way to derive the price to offer to do a given trade with a given counterparty. Most banks are strongly constrained by regulatory capital requirements. We can therefore expect the 'right' value for trade origination purpose for any given firm, after allowing for its own idiosyncratic funding and other costs, to depend on the exact way in which regulatory regimes impact these costs. In this sense, their analysis cannot be faulted.

And yet there seems to be something missing from this picture. There has always been scope for 'value' to be viewed idiosyncratically in this manner. The comparatively recent changes in regulation referred to by Kenyon and Green (2014) should not have intrinsically altered this picture. Nor is the issue dependent on specific regulatory regimes per se. We may expect funding costs to vary according to the credit standing of the firm needing the funding, so we should also expect firm-specific risk neutral measures to be sensitive to an institution's own credit risk. And yet market-wide (not firm-specific) risk neutral probability distributions have been a very important element in instrument pricing for many years.

The missing piece of the jigsaw is that there is a specific (and widely applicable) way of defining 'value' (and hence of defining a corresponding risk neutral measure) that *does* have market-wide relevance. The resulting risk neutral probabilities are not in principle firm-specific. This valuation methodology is the 'pure' market consistent valuation of the cash flows.

Kemp (2009) defines the market consistent value of an asset or liability to mean its market value, if it is readily traded on a market at the point in time that the valuation is struck, or in all other situations a reasoned best estimate of what such a market value would have been had such a market then existed. Implicit is the assumption that such market values respect the Principle of No Arbitrage. Formal accounting definitions of 'fair value' have similar features. Points to note include:

(a) The definition is independent of the institution in question. We do not in principle care what impact regulatory regimes have on any single institution. Our aim is to identify the price at which the given asset or liability would trade in an open market, were such a market to exist.

(b) 'Readily traded' is here interpreted as being in a market that is so deep and liquid that the instrument's bid-offer spread is de minimis. Some reasoned adjustment becomes necessary when market frictions become material. These might involve using mid-market prices in some circumstances. There is, however, no theoretical justification for necessarily choosing the mid-point between bid and offer in all circumstances.

(c) The second part of the definition (i.e. the case that applies when the asset or liability is not 'readily traded') requires exercise of judgement, particularly for less liquid instruments. Although the same conceptual definition of the valuation function applies to everyone (and is institution agnostic), individuals might in practice come up with different answers because they exercise judgement differently.

(d) How 'market consistency' is defined by regulation, e.g. Solvency II in the EU insurance world, does not necessarily accord with the 'pure' definition given above. It can also include pragmatic political compromises.

We can differentiate between two main types of valuation:

(i) *Valuations that aim to apportion value equitably between different owners*. An example might be a unit valuation of an open ended unitised fund owned by several different investors. The unitisation process in effect defines the price at which different investors can trade amongst themselves and with others. For such a valuation purpose the fund provider is typically required to treat all of its clients equitably. As Kemp (2009) notes, 'fairness' in effect requires use of some form of market consistent value. Suppose instead the provider uses values that are demonstrably off market (i.e. cannot reasonably be justified from market prices that are observable). Then one or other party to such a transaction may complain that the fund provider has disadvantaged them relative to the other side of the transaction.

(ii) *Valuations where the valuer is trying to maximise its own profit when buying (or selling) the instrument from (to) another.* For example, we might work as traders for a firm acting as principal in a potential derivatives transaction. The firm would then in effect be employing us to work out whether it should enter into the transaction, at what price and using what negotiating tactics to maximise the profit accruing to the firm. There is no specific need to refer to market consistent prices for such purposes. Indeed, we would be aiming to make money for the firm, so presumably are being paid to identify when the current market price appears to be unattractive. Of course, in such circumstances we are still likely to be interested in the market consistent price. Our employer is unlikely to be happy to enter into a transaction at a price obviously *worse* than could be achieved in the open market.

Some firms may need to carry out both sorts of valuation at the same time. For example, individuals within a fund management house in charge of portfolio strategy for an actively managed UCITS fund will try to work out whether the market value of a particular instrument appears to be 'cheap' or 'dear' relative to some perceived intrinsic value ascribed to its cash flows. Simultaneously, back office staff at the same fund management house will be seeking to work out the market value of the instrument to include in portfolio valuations used for unit pricing purposes.

What if different participants receive different dividend rates from the same stock?

To illuminate further the difference between market consistent valuations and firm-specific valuations of the sort referred to by Kenyon and Green (2014) we consider a situation motivated by their paper in which different investors receive different dividend rates on the same stock. Suppose that there are two different classes of investors, class A and class B. Suppose an investor in class A is entitled to 100% of any dividend paid on any shares it holds in company C whilst an investor in class B is only entitled to 90% of such dividends. Suppose an investor XA from class A enters into a two legged transaction with investor XB from class B. The first leg involves investor XA buying from XB some shares in C (that XB already owns) shortly before the dividend on C is paid. The second leg involves XA selling the shares back to XB shortly after the dividend on C is paid (at the same price as used in the first leg). For the time being we will ignore the counterparty risk XA and XB have to each other.

If a sufficient number of such investors are prepared to enter both sides of such a transaction (for a given dividend payment) then we might expect there to be a market clearing price for such trades. The market consistent value of the corresponding derivative instrument would then be set by this market clearing price.

If there aren't enough such investor pairs to result in a liquid market then there still is conceptually a market consistent price, but it becomes more subjective exactly what it might be. For example, it is not clear exactly how the balance between supply and demand might be set. Not all investors in either class A or class B might be expected to enter into such trades, since the trades are not in general riskless. For example, the 10% difference may be due to withholding tax and some potential participants in the transactions may be worried that the relevant tax authority might penalise one or both participants for entering into such transactions. C might not pay as high a dividend as expected. An investor might be worried about the possible mark-to-market loss he or she might suffer from the trade if it needed unwinding prior to maturity.

Either way, there conceptually *is* a common market-wide market consistent valuation and associated risk neutral measure, *even though* different firms will measure the intrinsic value of such trades differently and (in the absence of such trades) receive different dividends on the same stock.

As an aside, it is worth noting that XA and XB may both be able to 'outperform' the market-wide risk neutral measure, by entering the 'correct' side of such a transaction from their own perspective. In the specific situation described above the gain may perhaps be coming at the expense of some tax authority. It may therefore have doubtful overall value to society as a whole. In other situations the 'outperformance' may correspond to some activity with a greater societal worth. For example, it may correspond to a reward for more efficient pooling of risks, better allocation of capital or provision/consumption of socially useful financial services etc.

Incorporation of own credit risk

We can use similar reasoning to identify the theoretically ideal treatment of own credit risk in minimum regulatory capital requirements. The 'pure' market consistent value described above includes the market-implied creditworthiness of each side of a transaction. For any given firm that is a counterparty to a transaction it should include an own credit risk component where relevant. This reflects the uncontroversial observation that the price charged to novate a contract in general depends on the credit standings of the parties from whom and to whom the contract is novated.

The existence of idiosyncratic credit risk does not, however, render invalid a market-wide risk neutral measure for minimum regulatory capital requirements. No practical regulatory capital framework can guarantee zero risk of failure. Customers face an expected loss because the financial contracts they have entered into are exposed to product provider credit risk. Ideally, regulatory capital frameworks should aim to apportion this expected loss in a fair manner across all customers (and other relevant stakeholders) and across all firms. We assert that valuations for minimum regulatory capital purposes are therefore a specific type of valuation for which fairness between parties ought to be important. As explained above, this should in theory favour use of market consistent approaches for such valuations.

Suppose also that we have two similar firms both running at their regulatory minimum capital levels. 'Fair' allocation of such expected losses also means that these levels should ideally be chosen so that customers with otherwise identical products from the two different firms have similar expected losses from product provider credit risk irrespective of the firm they contract with.

At least in theory, minimum regulatory capital requirements should therefore derive from a *target* credit spread on customer liabilities (versus a suitable risk-free rate), rather than depend on the credit spread actually present on these liabilities, see Kemp (2009). So the existence of idiosyncratic own credit risk does not imply abandoning market-wide risk neutral measures. Instead, it implies that we should be seeking, at least in theory, to standardise the degree of own credit risk included in valuations for regulatory minimum capital purposes. Only by doing so will we apportion fairly the credit risk exposures that customers of different firms have to these firms and, by extension, apportion fairly any associated implicit or explicit government or industry protection scheme underpins. We might want this target credit spread to change through time for counter-cyclical and other macro-prudential purposes. However, ideally such changes should be consistently applied across firms. Otherwise the regulatory capital framework could apportion the expected value of government or industry protection scheme underpins in an inequitable fashion between different customer bases.

Conclusions

Kenynon and Green (2014) compellingly argue that recent regulatory changes have created idiosyncratic holding costs for different institutions and that this results in different firms pricing otherwise identical instruments differently. However, differential views on the attractiveness of a specific instrument to a specific counterparty do not by themselves eliminate the existence or relevance of a shared market-wide risk neutral measure for many pricing purposes. We argue in this paper that one such purpose is for regulatory capital minimum requirements. Neither do different levels of own credit risk. Instead, they imply that regulatory capital minimum requirements should in theory be set in a manner that includes a standardised but potentially time-varying allowance for such credit risks.

References

Kemp, M.H.D. (2009). *Market consistency: Model calibration in imperfect markets*. John Wiley & Sons

Kenyon, C. and Green, A. (2014). Regulatory costs break risk neutrality. *Risk*, September 2014

Shreve, W. (2004). *Stochastic Calculus for Finance, Volume II: Continuous Time Models*. Springer