Working Paper

Credit risk models: why they failed in the credit crisis

Wilson Sy – July 2008
Credit risk models: why they failed in the credit crisis

by

Wilson Sy

Australian Prudential Regulation Authority
400 George Street
Sydney NSW 2000
Email: Wilson.Sy@apra.gov.au
Phone: 0612 9210 3507

(11 July 2008)

Presented to the 13th Melbourne Money and Finance Conference, 2-3 June 2008, Brighton, Victoria

To appear in a special issue of JASSA
Journal of the Australian Society of Securities Analysts

The views and analysis expressed in this paper are those of the author and do not necessarily reflect those of APRA or other staff. The data samples and methods used in this paper are selected for the specific research purposes of this paper and may differ from those used in other APRA publications.

APRA does not accept any responsibility for the accuracy, completeness or the currency of the material included in this Publication, and will not be liable for any loss or damage arising out of the use of, or reliance on, this Publication.

1 The author thanks Katrina Ellis and the participants of the Melbourne Money and Finance Conference for helpful comments.
Abstract

Credit risk models are shown to play a key part in the global credit crisis. We discuss how the credit market has exposed the shortcomings of the credit risk models and we identify their main shortcomings. To overcome the shortcomings, a new causal framework is proposed to build deductive credit default models which have predictive capabilities.

JEL classification: B41, C81, D14, E44, G21, G32, G33.
Keywords: Credit risk models, credit crisis, causal framework, delinquency, insolvency, mortgage default.
The global credit crisis is the failure of major parts of the credit market to function normally, leading to failures of financial institutions because they were unable to obtain funding from the market. The first part of the market to fail was the US sub-prime mortgage market. Subsequent failures of other parts of the market suggest that the problems are more fundamental than simply the mispricing of some mortgage-backed securities.

As there has been no identifiable external shock to the financial system, the crisis appears to have an endogenous origin which could have multiple systemic causes associated with the various steps of the credit process developing over time. Undoubtedly many of the “usual suspects” such as excessive credit growth, lax lending standards, conflicts of interest, complex financial instruments etc. all play some part. In this paper, we consider the key part played by the pricing of credit risk. We discuss how the credit market has exposed the shortcomings of the credit risk models. We discuss the main shortcomings and how to overcome them.

Credit market failure

In his seminal paper, Akerlof (1970) identified two ingredients for market failure: existence of low quality products (e.g. “lemons” in the used car market) and asymmetric information with buyers being less informed than sellers. Both these ingredients are present in the credit market crisis for mortgage-backed securities.

Minsky (1992) has anticipated that the flawed incentives of the mortgage securitisation process would create loans of low credit quality, which now have a variety of names: “sub-prime”, “Alt-A”, “non-conforming” etc. This insight has recently been confirmed officially by the U.S. government (Paulson, 2008). The securitisation process has also led to asymmetric information, because not all information about the loans the mortgage brokers had at loan approval was transmitted to the buyers of the mortgage-backed securities (MBS), as the crisis subsequently proved. At the time, the existence of asymmetric information was difficult to demonstrate because buyers relied on prices set by credit rating agencies who were presumably fully informed about the approved loans.

Once money was lost from the MBS, withdrawal from the market by investors gave rise to a modern version of a “run on the banks” in the financial system. The failure of the market to provide credit might not have been so evident or debilitating for the economy, if the dependence on credit from the market had not been so great. With 50% of the US$14.4 trillion (FRB, 2008) in outstanding mortgages in the United States being intermediated through securitization, at the end of the third quarter, (35% through traditional institutions and 15% was through governments and individuals), the disruption to market function was significant.

After an initial shock from unexpected news when the market temporarily becomes illiquid due to uncertainty, the market normally recovers quickly with trading resuming around new consensus price levels. However, in this crisis, market recovery did not occur quickly making obvious the existence of an asymmetric information problem. After several months of the crisis, we still seem to be a long way from a solution to the asymmetric information problem.

The fact that many parts of the credit market have remained closed to new lending suggests something still more serious has affected the market. It appears that having more information on the mortgages is insufficient to unlock the market, since we need to translate individual loan data to prices for securities with credit risk. This appears to be beyond the capabilities of the credit risk models that are currently in use and it is one of purposes of this paper to offer an explanation.
In March 2008, in his testimony before the U.S. House of Representatives, Charles Prince, then the chair and CEO of Citigroup, one of the largest banks in the world, admitted that the credit risk models used at Citigroup and in the industry were wrong. In early May 2008, Standard and Poors announced (Shenn, 2008) that it will stop rating certain types of mortgage-backed bonds citing unprecedented deterioration of the loans and that “the market segment does not allow meaningful analysis”.

The Lucas critique

What’s wrong with the credit risk models? Most credit risk models, as in much of economics and finance, are based on the assumption of market equilibrium (Blaug, 1998), which has not been adequately supported by empirical evidence in the real world (Coarse, 1998). Their basic shortcoming can be discussed firstly within a general critique of econometrics.

In his famous critique with macroeconomics in mind, Lucas (1976) pointed out the limitations of econometrics for guiding economic policy. We believe the basic argument he used applies also to economics and finance theory generally and we sketch here an outline of his argument. Econometrics is based on inductive methods and is mainly concerned with the estimation from empirical data of a general statistical model:

\[
y_{t+1} = F(y_t, x_t, \theta, \epsilon_t). \tag{1}
\]

Given a history up till time \( t \) for a vector of dependent variables \( y_t \), a vector of independent “forcing” variables \( x_t \), a vector of independent and identically distributed random shocks \( \epsilon_t \) and a set of environmental factors \( \theta \) representing public policy and decision rules, the objective of econometrics is to determine the model \( (F, \theta) \), where \( F(.) \) is a function and \( \theta \) is a set of parameters. In credit risk models, \( y_t \) would be default rates and recovery rates, \( x_t \) could be risk factors measured by accounting ratios or other characteristics of the borrower and \( \theta \) is the set of environmental factors and individual behaviours which are implicit in the model.

Given a set of historical data \( \{y_t, x_t\} \), we can, in principle (though not necessarily easily), determine a model \( (F, \theta) \) through statistical data-fitting. The presumption in econometrics is that \( (F, \theta) \) is structurally stable and does not vary with different sets of data of the “forcing” variables \( x_t \). In his critique Lucas (1976, p.25) points out: “Everything we know about dynamic economic theory indicates that this presumption is unjustified”. In other words, given another set of data \( \{y_t', x_t'\} \), it will often lead to another model \( (F', \theta') \). In empirical research, we often find significant differences between ex-post and ex-ante model results. In essence, a given \( (F, \theta) \) is a description of a state of the world, this state will not change structurally only if the world is in a state of equilibrium. An econometric model describes a historical state specified by \( (F, \theta) \) which is estimated from the empirical data. This historical state may not be an equilibrium state and may not continue into the future if environmental conditions change due to economic policy or other factors.

The Lucas critique is robust against any further developments in econometrics as defined by equation (1), as new research merely leads to more sophisticated models \( (F, \theta) \) and it “will not be of use for forecasting and policy evaluation of actual economies” (Lucas, 1976,
The world we experience is structurally unstable relative to econometric models or equilibrium states, where large deviations of several standard deviations should not be a surprise. Our world is inhabited by “black swans” (Taleb, 2007) and distributions with “fat tails”. We saw an example of this in the credit crisis where some delinquency rates were sometimes several standard deviations higher than can be expected from credit risk models.

**Credit risk models**

There are two recognised approaches to credit risk models: “reduced form” models and “structural” models. Most credit risk models in use are reduced form models, which are usually a linear subset of the general econometric model given by equation (1). Hence the Lucas critique applies to them and therefore they are not generally valid. Expansion of $F(.)$ about an equilibrium state $\theta$ in a single period model leads to

$$y_{t+1} = c_0(\theta) + c_1(\theta)x_t + x_t'c_2(\theta)x_t + ... + \epsilon_t.$$  

The third term on the right hand side is a nonlinear quadratic term where $c_2(\theta)$ is a square coefficient matrix dependent on the state $\theta$ and $x_t'$ is the transpose of $x_t$. This second order and higher order terms are ignored in linear models of default or recovery risks, because equilibrium fluctuations are assumed small. It has become a highly simplified form of the general econometric problem (1), where standard linear regression methods can be used.

One may argue that there are equilibrium environments where $F(.)$ is approximately linear and $\theta$ are stable and this approach can be used to exploit available market information for pricing and hedging (Jarrow and Protter, 2004). This appears to be true in the halcyon days of the credit market when the economy was growing steadily and credit default rates were low: the market was in a temporary quasi-equilibrium state. The models were useful in predicting credit defaults. Once the market started to move from this state, the estimated models ceased to be valid because the regression coefficients $c_0(\theta)$ and $c_1(\theta)$ are changing. New models can be estimated only when $c_0(\theta)$ and $c_1(\theta)$ stop changing and sufficient data have been collected in a new quasi-equilibrium state to make new estimates.

The main shortcoming of the econometric approach in the reduced form models is the typical reliance on having large amounts of statistical data coming from a quasi-equilibrium state. The approach cannot be used to make even short-term forecasts in rapidly changing environments such as in a credit crisis. Such inductive models have failed to predict what would happen just when they were most needed to work. The shortcomings of using inductive methods alone have been well recognised historically. In a letter to Keynes, Marshall wrote: “You talk of the inductive & the deductive methods: whereas I contend that each involves the other...” There is a need to integrate inductive and deductive methods in economics, as inductive methods alone cannot provide non-equilibrium predictions.

Inductive and deductive methods have been used in an integrated epistemology (see Figure 1) in the natural sciences leading to solid advances in human knowledge. The ultimate objective of this epistemological process is the creation of deductive models which could predict accurately empirical data. Predictability (Blaug, 1998, p.29) is “the ultimate test of whether our theories are true and really capture the workings of the economic system independent of our wishes and intellectual preferences.” Only deductive models can make predictions based on limited amount of data or assumptions.
Relatively little\textsuperscript{14} of economics and finance have gone through the whole epistemological loop in Figure 1. Most have been trapped either to the lower-left part of the loop in purely econometric studies or to the upper-right part of the loop in equilibrium theory formalism as discussed by Blaug (1998). Reduced form models as a subset of the general econometric model have been trapped to the lower-left part of the loop. Structural models have perhaps covered more parts of the loop, as they are deductive and have been applied to empirical data.

Structural models originated from Merton (1976) who applied equilibrium theory of option pricing to corporate debt. Strictly speaking, his model is not a credit default theory, but rather it is a theory of risk premium determination based on the assumption that traders will eliminate arbitrage opportunities in a bond market at equilibrium. There is a subtle but important distinction: it is an equilibrium theory about bond prices if the bonds can default. The theory identifies the expectation of loss from insolvency with the expectation of default. The Merton model has been interpreted (Kealhofer, 2003) as being causal and deductive. But that as it may, we will show the Merton model is not applicable generally to many areas of credit risk, including markets for unsecured loans and markets without securities trading, as we will discuss below.

Essentially, the Merton model estimates the probability that the stochastic insolvency variable

\[ x_v = \frac{\text{Assets}}{\text{Liabilities}} \]

falls below unity, indicating the borrower has negative equity. The variable is stochastic because the value of the firm fluctuates according to some stochastic process whilst the amount of outstanding bonds (liabilities) is assumed fixed. If we assume the stochastic process to be a standard random walk of a Gaussian process, we obtain enclosed-form solutions of Black and Scholes option pricing formulae. Over time the model has had various extensions (e.g. the KMV model), but recently Kealhofer (2003, p.42) concluded: “implementation of the model in practice has proven to be more difficult than originally anticipated”. We believe the difficulties are due to inappropriate or incomplete default causality in the Merton model.
The basic idea of insolvency as the cause of default in the Merton model is not sufficiently general, because of the existence of other possible causes. We take the view that insolvency may be a necessary, but not sufficient, condition for default of secured loans. There are probably many individuals and corporations which may be technically insolvent but may still be able to stave off default because they continue to make the necessary debt repayments. In other words, liquidity or the ability to service loan commitments is also a decisive factor in the actual occurrence of default. We provide two examples where liquidity or the ability to service loans is critical in understanding credit default.

In the aftermath of the Asian financial crisis in 1997, over a period of six years from 1998 to 2003 average property prices in Hong Kong SAR (Fan and Peng, 2005), dropped 60-70%, with as many as 30% of residential mortgage loans having negative equity. While property prices were dropping continuously throughout the period, the loan delinquency ratio rose at first from slightly above 2% to a peak of more than 7% in 1999. But thereafter delinquency rates dropped continuously to less than 3% by the end of 2003, as property prices continued to fall. Such an event would contradict the underlying assumptions and the subsequent predictions of the Merton model.

The second example is the phenomenon of Grameen micro-credit (Yunus, 2006), where loans are made to poor villagers who are without any assets as collateral. Actual experience over several years showed that the default rates averaged at less than 2% per annum, due to the ability of villagers to service the loans by making small business profits. The Merton model is inapplicable to Grameen micro-credit and to the whole class of unsecured loans, such as credit cards and other consumer credit.

Clearly, the Merton approach suffers from incomplete causality for all types of loans, as we need to also include the causal effects arising from an entity’s cash flows in a more complete theory (even for the traded bond market). In the assessment of corporate health from financial account data, for example, this is equivalent to paying attention to the profit and loss statements, as well as to the balance sheet statements, which are the only statements a Merton model would consider.

**A new causal framework**

In a new approach to credit risk (Sy, 2007), we assume the primary cause of credit default is loan delinquency due to insufficient liquidity or cash flow to service debt obligations. In the case of unsecured loans, we assume delinquency is a necessary and sufficient condition. In the case of collateralised loans, delinquency is a necessary, but not sufficient condition, because the borrower may be able to refinance the loan from positive equity or net assets to prevent default. In general, for secured loans, both delinquency and insolvency are assumed necessary and sufficient for credit default.

We introduce a liquidity model which estimates the probability that the stochastic delinquency variable

\[
x_s = \frac{\text{Cash flow to service loan}}{\text{Loan payment}}
\]  

falls below unity. The variable is stochastic because circumstances can change in random ways to affect the ability of the borrower to service loan obligations. We call our new approach a framework rather than a theory because there are arbitrary numbers of ways to model cash flow depending on the circumstances of the borrower and the information
available to quantify liquidity. Hence the framework permits many possible theories and models.

Once the stochastic process controlling the evolution of the delinquency variable is specified, we can determine its probability distribution at a later time \( t \) and from this calculate the probability for delinquency where \( x_s < 1 \). As a simple example, assuming a standard Gaussian process, the probability distribution at a later time is described by a “distance to delinquency” variable:

\[
Z_S = \frac{\ln(x_s) + (\mu_s - \frac{1}{2} \sigma_s^2)t}{\sigma_s \sqrt{t}}.
\]  

(5)

The probability of delinquency is given by \( N(-z_s) \) where \( N(\cdot) \) is the standard Normal cumulative probability function. There are two parameters: the drift rate \( \mu_s \) and the volatility \( \sigma_s \), which can be used to describe future conditions and risk.

A similar discussion applies to the insolvency variable \( x_i \) and an analogous expression for \( z_i \) can be written by replacing of subscripts in equation (5). We reject the equilibrium assumptions of the Merton model as being too restrictive, since we are concerned with other non-traded markets as well as traded bond markets. In general, we have an evolution of the insolvency variable described by two parameters: the drift rate \( \mu_i \) and the volatility \( \sigma_i \), whereas in the Merton model the drift rate is replaced by a riskless interest rate due to the absence of arbitrage at market equilibrium.

In the case of secured loans, the two causal stochastic variables for delinquency and insolvency may be correlated, a priori. If we assume Gaussian processes then the default probability is determined by a bi-variate Normal probability density function, with a given correlation coefficient (Sy, 2007).

In the uncorrelated case, we have a product of two independent uni-variate Normal probability density functions, with a set of four parameters: \( \mu_s, \sigma_s, \mu_i, \sigma_i \). The probability of default is then given simply by \( N(-z_s)N(-z_i) \). In the case of the Hong Kong SAR mortgage market (Fan and Peng, 2005) referred to above, we would have \( N(-z_i) \) being essentially equal to one. For this case and for the cases of Grameen micro-credit and unsecured loans, the probability of default is determined by \( N(-z_s) \), which is the probability of delinquency.

The puzzle of Hong Kong SAR housing market has been resolved (Sy, 2007) by showing that the falling interest rate environment of period had made loans more easily serviceable leading to low probabilities of delinquency. With reasonable selected values for the parameters: \( \mu_s, \sigma_s, \mu_i, \sigma_i \), we are able to reproduce the main observed features of the Hong Kong SAR property market between 1997 and 2003, when interest rates and property prices were both falling. Similar models can be applied to any property market, provided we can model the factors affecting the cash flow situation of the borrower. Of particular interest for us is the case of Australia.
Australian residential mortgages

To illustrate the application of the causal framework, we assess residential mortgage default risk using a simple model for typical households with wage earners. For the delinquency variable in (4), we assume a loan serviceability ratio (LSR) defined by

\[
x_S = \frac{\text{After tax income} - \text{Living costs} - \text{Other payments}}{\text{Mortgage payment}}.
\]

The insolvency variable in (3) is assumed to be the reciprocal of the loan-to-value ratio (LVR). The loan approval process captures the relevant data from the borrower to provide estimates of LSR and LVR at origination for each loan.

Given microeconomic and macroeconomic assumptions about how wages, inflation, consumer credit usage, interest rates and property prices are likely to change in the period ahead, we can estimate the model parameters \(\mu_S, \sigma_S, \mu_V, \sigma_V\) to predict how LSR and LVR will evolve over time. From their time-dependent probability distributions, we can calculate for any time ahead the probability of default, loss given default and expected loss for any given loan.

In 2006, Australian Prudential Regulation Authority (APRA) collected data on 112,000 housing loans approved by 44 largest lenders in September 2006 worth $27.6 billion in loan commitments. A statistical report has recently been published by APRA (2008). We also found in other studies on this and other datasets that past increases in house prices running well ahead of wage increases had led to decreased serviceability of many loans. At the same time, increased housing lending to more and more households was facilitated by lower lending standards relative to traditional criteria.

We define “traditional” loans as those satisfying the criteria:

- Debt service to gross income ratio not exceeding 30%
- Loan-to-value ratio not exceeding 80%
- Loan approval directly by lending institution, rather than through mortgage brokers
- Full documentation, rather than low-documentation

Traditional loans of our sample have a median LSR of 2.5 and have low probabilities of default as calculated from our model even though interest rates were rising in the 18 months since approval. Bank standard variable rate from RBA increased from 7.8% to 9.45% in April 2008. But this was counter-balanced by average property prices still rising over the period. Data from Australian Bureau of Statistics show weighted average index of 8 capital cities for established homes rose from 109.3 to 128.1 in December 2007.

The non-traditional loans have a lower median LSR of 1.6, leading to higher probabilities of default due to increased difficulties in servicing the mortgages in the same environment, particularly at higher values of LVR, though the average LVR of the sample was around 70%. The parameters \(\mu_S, \sigma_S, \mu_V, \sigma_V\) used in Figure 2 have respective values: 12%, 20%, 11%, 25%. The values have been estimated (RBA statistical data) from actual changes since September 2006 in average wage, inflation, mortgage interest rates and property prices. The calculation predicts what would happen to non-traditional housing loans since September 2006, if the assumed environmental conditions remain constant over the forecast period.
Because of rising property prices, the Merton model would have predicted falling default rates, contrary to our predictions and actual experience. The example selected is not meant to indicate anything general about the situation of Australian mortgages, as this would require a comprehensive study of the contents of all housing loan portfolios. It is selected to illustrate the ability of our deductive model to make predictions based on limited general assumptions about the environment in the period ahead. An inductive method cannot be used to model an anticipated new environment, particularly if it has no historical precedence, since there is no empirical data available to use to estimate the new model and old models would be inappropriate.

Conclusion

Credit risk models need to be based on causal frameworks, such as the one suggested in this paper. Only through an understanding of the causality of the credit default process, can we build deductive models which are capable of making predictions in a changing environment.

References


Keynes, J. M. (1925), review of *Common Stocks as Long Term Investments*, Edgar Lawrence Smith.

Knight, Frank (1921), ‘Risk, Uncertainty and Profit’, Dover Publications, USA.


Notes

1Early examples of failure include Northern Rock in U.K., IKB Deutsche Industriebank in Germany, Bear Stearns in U.S. and RAMS in Australia. All were either bailed out by government or taken over by much larger entities.

2Other parts of the market include other non-conforming mortgage markets, the municipal bond market and auction-rate securities market, to name a few.

3It is the calculation of additional interest rate spread above the riskless rate to compensate for the risk of credit default, which leads to total or partial loss of loan principal.

4Minsky (1992, p.22) wrote: “The securitization of standard mortgages was a technique by which Savings and Loans and Mortgage companies originated mortgages which were then packaged as securities for the portfolios of holders such as pension funds, life insurance companies, mutual trusts and various international holders. Because of the way the mortgages were packaged it was possible to sell off a package of mortgages at a premium so that the originator and the investment banking firms walked away from the deal with a net income and no recourse from the holders. The instrument originators and the security underwriters did not hazard any of their wealth on the longer term viability of the underlying projects. Obviously in such packaged financing the selection and supervisory functions of lenders and underwriters are not as well done as they might be when the fortunes of the originators are at hazard over the longer term. All that was required for the originators to earn their stipend was skill avoiding obvious fraud and in structuring the package.”

5The President’s Working Group on Financial Markets found in March 2008 (Paulson, 2008, p.11): “Originators had weak incentives to maintain strong underwriting standards, particularly at a time when securitisers, credit rating agencies, and mortgage investors did not conduct due diligence sufficient to align originator incentives with the underlying risks.”

6By contrast, only 25% of total outstanding mortgages (A$893 billion) are securitised in Australia (RBA, 2008).

7Prince said: “Last fall, it became apparent that the risk models which Citigroup, the various rating agencies, and the rest of the financial community used to assess certain mortgage-backed securities were wrong. As CEO, I was ultimately responsible for the actions of the company, including the risk models that eventually proved inadequate”.

8He was concerned with the breakdown in the stability of the empirical Phillips relationship between inflation and unemployment found from econometrics.

9His association with the theory of rational expectations may confuse some readers in understanding the basic simplicity and validity of his argument, which was never claimed to be original, as acknowledgement was made in his paper to earlier thinkers such as Knight (1921).

10A recent paper by Campbell et al. (2005) contains a large number of other references to reduced form models.

11The functional form of $F$ is usually assumed linear, while some of variables may be non-linearly transformed by logarithmic, logit or other functions.

12Keynes J. M. (1925) wrote: “It is dangerous to apply to the future inductive arguments based on past experience unless one can distinguish the broad reasons for what it was.” See also Knight (1921, p.7).

13Letter from Alfred Marshall to J. M. Keynes, on file in the Marshall Library, Cambridge, as Keynes 3 (letter no. 66) cited by Coase (1975, p.26). Still earlier expression for the need of both inductive and deductive methods can be found in Knight (1921).

14The lack of predictive ability in current economics (Blaug, 1998) is compared to Newtonian physics which is now fully predictive of a vast array of physical phenomena on intermediate scales (excluding the very large and the very small scales).