Wildcat Banking, Banking Panics, and Free Banking in the United States

Gerald P. Dwyer Jr.

From 1837 to 1865, banks in the United States issued currency with no oversight of any kind by the federal government. Many of these banks were part of "free banking" systems in which there was no discretionary approval of entry into banking.[1] A note received in a transaction might indicate that it was issued by, say, the Atlanta Bank. This banknote was used for payments in transactions and was redeemable on demand at the Atlanta Bank for a specified quantity of specie, gold or silver. These notes were used in transactions just as checks are today. In important respects, though, they were quite different from today's checks. Notes were passed from one person to another and yet another before being returned to the bank. They were the bank's obligations, not bank customers' obligations. Because there was no central bank and no government insurance, the ultimate guarantee of a banknote's value was the value of the bank's assets.

Free banking in the United States sometimes has been equated with "wildcat banking," a name that suggests that opening a bank has much in common with drilling for oil. Drilling for oil is not an obvious analogy for a sound banking system. Use of the word *wildcat* to mean "reckless" or "financially unsound" apparently arose in Michigan in the 1830s, when bankers supposedly established free banks in inaccessible locations "where the wildcats roamed."[2] In the free banking period such locations benefited banks because they hampered noteholders' attempts to redeem notes, and banks with fewer notes redeemed could hold less specie and generate higher net revenue for their owners.

More generally, when banks issue notes, a major issue for banking laws and holders of banknotes is enforcement of banks' contracts to redeem the notes. If a bank issues notes in good faith that they can be redeemed as promised, the issue is simply contract enforcement. If a bank issues notes

The author is a visiting scholar in the financial section of the Atlanta Fed's research department and a professor in the Department of Economics at Clemson University. *He thanks Robert Bliss, R. Alton Gilbert, R.W. Hafer, George G. Kaufman, Will Roberds, Hugh Rockoff, Arthur Rolnick, George Selgin, Larry Wall, and Eugene N. White for helpful comments.*

with no prospect that they can be redeemed, the issue becomes prevention of fraud, or what is "essentially counterfeiting." Milton Friedman (1960, 6) suggests,

> It so happens that the contracts in question are peculiarly difficult to enforce and fraud peculiarly difficult to prevent. The very performance of its central function requires money to be generally acceptable and to pass from hand to hand. As a result, individuals may be led to enter into contracts with persons far removed in space and acquaintance, and a long period may elapse between the issue of a promise and the demand for its fulfillment. In fraud as in other activities, opportunities for profit are not likely to go unexploited.

Free banks did not always redeem their notes as promised, and there are fabulous stories of fraudulent activities, stories that appear frequently in histories of free banking and general histories of banking. For example, in an examination report for Jackson County Bank in Michigan in 1938, the state bank commissioners report that they found the account books had accountholders' names written in pencil and their balances written in pen. In addition, they examined the bank's specie.

> Beneath the counter of the bank, nine boxes were pointed out by the teller, as containing one thousand dollars each. The teller selected one of the boxes and opened it; this was examined and appeared to be a full box of American half dollars. One of the commissioners then selected a box, which he opened, and found the same to contain a superficies only of silver, while the remaining portion consisted of lead and ten penny nails. The commissioner then proceeded to open the remaining seven boxes; they presented the same contents precisely, with a single exception, in which the substratum was window glass broken into small pieces. (U.S. Congress 1839-40, 1109)

Whether or not this story is typical of Michigan's free banks, free banking in that state in the 1830s was a failure, with noteholders suffering heavy losses. In fact, in his influential history of banking, Bray Hammond concludes that people in states where banking was prohibited "were better off than the people of Michigan, Wisconsin, Indiana, and Illinois," who had free banking (Hammond 1957, 626).

Was free banking in the United States so bad that people would have been better off with no banks at all? One way of approaching this question is to ask, Did noteholders suffer substantial losses from holding free banks' notes? If those losses were substantial, were they generally associated with difficulties in enforcing the contract between noteholders and banks, fraud, or both? This issue is not of only historical interest because, as discussed in Box 1, free banks' notes have similarities to some forms of electronic money. Recent research makes it possible to provide more informed answers to these questions than was possible even as recently as 1980.[3]

The Spread of Free Banking

After the Second Bank of the United States' federal charter expired in 1836, the various states provided the legal framework for banking, and there was no banking system operating under federal government law. Prior to free banking, limited-liability organizations were permitted to issue notes if the legislature granted a charter for that specific purpose. Free banking opened up note issuance to limited-liability organizations without discretionary approval by a legislature, as in earlier years, or by a banking regulator, as in later years (Gerald C. Fischer 1968, 177-84). Free banking ended in 1865 when the federal government imposed a tax on state banknotes.

Chart 1 shows a map of the United States in 1860 and the years that the states adopted free banking. Three states adopted free banking in the 1830s: Michigan, Georgia, and New York. The rest that adopted free banking did so in 1849 and later years.

Overview of Free Banking

The free banking systems in the various states had many things in common.[4] The defining characteristic of free banking is that if the requirements of a given state's free banking law are met, any person or group of persons is permitted to open a bank.

Opening a Bank. Prior to opening a bank, subscriptions for a minimum amount of capital funds were required. When subscribing for stock, a person commonly paid in some of the funds and promised to provide additional funds up to the amount subscribed. The law generally required that a minimum amount of funds be paid in before the bank could begin operation. Some states also limited the maximum amount of capital in any one bank. An abbreviated balance sheet for a free bank, shown in Table 1, is helpful for understanding free banks' note issuance. The capital funds appear on the balance sheet as equity capital, a liability. The balance sheet also includes an asset, loans to stockholders, to illustrate the offset for capital subscribed but not paid in.[5] In addition to being liable for their subscribed capital, the bank's stockholders often were subject to double liability: when the bank closed, each stockholder could also be required to pay an additional amount equal to the stockholder's subscribed capital.

Issuing Currency. Banks were permitted to issue banknotes that circulated from hand to hand much as Federal Reserve notes do today. In order to issue notes, banks were required to make a security deposit with the state banking authority. The state banking authority then signed the notes and provided them to the bank.

Box 1
Free Banks' Notes and Electronic Money

Will electronic money resemble the banknotes circulated in the U.S. free banking period?[1] Walter Wriston (1995, 1996; Bass 1996), a former Chairman of Citicorp, and others have suggested that money used for transactions on the Internet may resemble nineteenth-century banknotes more than it will today's money.

Actually, only a subset of what often is called electronic money is "money" in the economic sense, and most of that subset is more similar to money orders or cashier's checks than banknotes. The confusion between money and other means of payment arises even in the economics literature, so confusion in the popular literature is not surprising. Friedman and Schwartz (1970, chaps. 2 and 3) provide an accessible discussion of the definition and measures of money. For example, even though a credit card can be used to make purchases, neither a credit card nor its unused balance is money. When someone uses a credit card to buy a dinner, the purchaser is promising to pay later with money.

Some electronic payment schemes, such as one run by a company called First Virtual, make no pretense at introducing electronic money. First Virtual holds buyers' credit card numbers on a computer inaccessible from the Internet and verifies the authenticity of a purchase. In effect, First Virtual adds an intermediary to transactions. Several other payment schemes focus on preserving anonymity for the buyer but do not introduce the equivalent of currency that can be received and spent repeatedly without involving the money issuer or another third party.

The electronic payment schemes closest to electronic currency are the use of "electronic wallets" and "money modules." These schemes, which require hardware not now widely available in the United States, make it possible to transfer balances from one person's wallet or module to another without another party to the transaction.

Compared with paper currency, checks, and credit cards, such electronic currency would have some advantages for buyers and sellers who want to conduct transactions on the Internet. One advantage is that electronic currency can preserve the anonymity of a transaction in the same way that paper currency does. Probably more important to most people, electronic currency could be used for simple transactions on the Internet between people who do not have enough transactions or the reputation to acquire a merchant credit card account. It also is possible that electronic money could be simpler for international transactions than money denominated in local currency, partly because it is relatively expensive to convert from one currency to another in small amounts.

As of this writing, an institution located in the United States attempting to issue private money would confront substantial legal issues and taxes that might make such an issue impractical. The costs of domestic and international communication on the Internet, however, are effectively the same. Hence, despite impediments in the United States, developing private money offshore in a less restrictive jurisdiction could create a viable alternative even for transactions between U.S. residents.

There may be similarities between electronic money and free banks' notes. Electronic money is likely to consist of uninsured liabilities of private individuals or companies. If so, perhaps the most immediate lessons from free banking are that (1) consumers are not sheep waiting to be sheared (2) attention must be paid to the importance of the assets into which the electronic money is convertible and to the issuer's reputation for making the conversion as promised.

Note

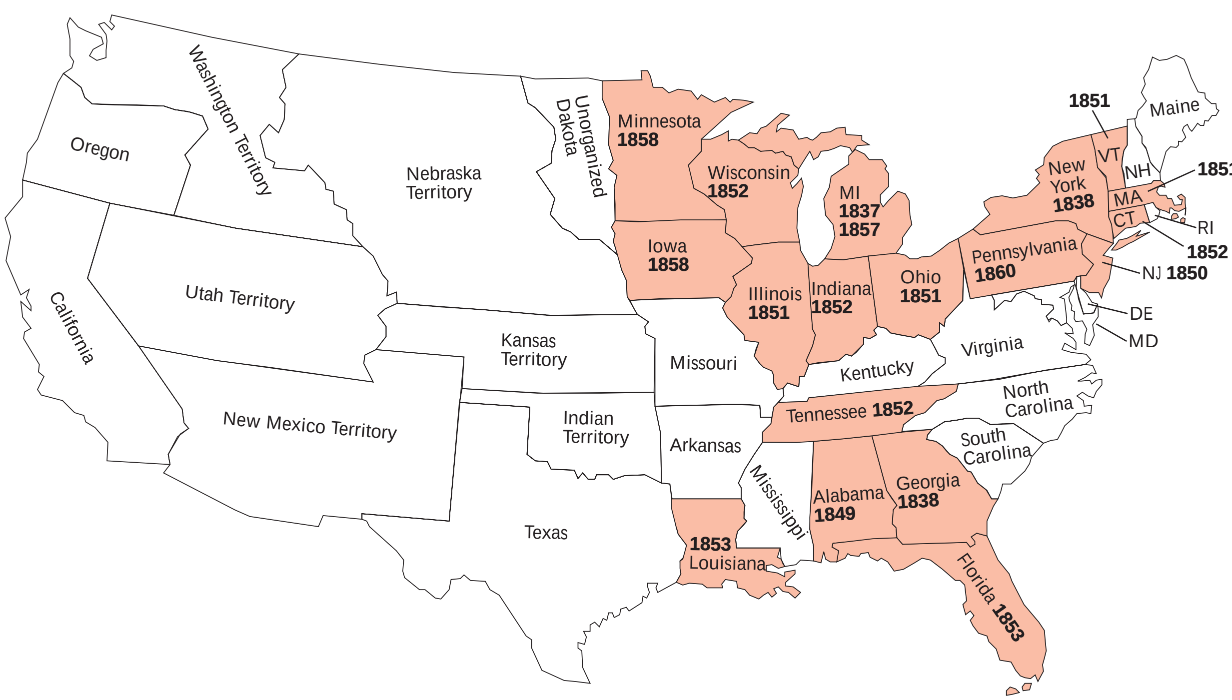
1. Levy (1994) and Flohr (1996) provide accessible introductions to electronic money. Schneier (1996, 139-47), the standard nontechnical source of information on cryptography, provides some details about how one form of electronic money can be securely implemented and references to discussions of other forms of electronic money.

The free banking laws specified acceptable assets that could be deposited. They generally allowed deposits only of selected state bonds, known as state stocks at the time, and U.S. government bonds, both of which traded on the New York Stock Exchange. The marketability of these bonds simplified valuing the notes, which contributed to their widespread use. As Eugene N. White (1995) indicates, it also contributed to eliminating earlier legal restrictions on low-denomination banknotes. As long as the security's value was at least as high as the security deposit required for the outstanding notes, the bank received the interest on the bonds. As Table 1 shows, the bonds deposited were an asset of the bank and the notes were a liability. States required that bonds be valued at the lesser of par value (face value) or market value, and some states permitted banks to issue notes only up to a fraction, for example 90 percent, of the bonds' par or market value.

If the security deposit's value fell below the notes' value, banks were required to add bonds to the security deposit or to decrease their notes outstanding. These changes in bonds or notes were necessary until the security's value was at least as high as the notes' value. If the bank failed to make up the deficiency in its security deposit within a limited time, the bank was closed, and the bank's security deposit was used to pay noteholders on a pro rata basis. The bank's bonds were sold, and noteholders received the lesser of the proceeds or the notes' par value. Any excess of the bonds' value above the value of outstanding notes was returned to the bank's owners. If the proceeds from selling the security deposit were less than the notes' par value, the noteholders could file suit against the bank and its stockholders for the deficiency up to the limits of the bank's and stockholders' liability. This procedure was used for winding up the security deposit if the bank closed for any other reason as well.

Par Conversion on Demand of Banknotes Required. Banks were required by law to convert their notes into specie at the notes' face value on demand. As shown on the balance sheet in Table 1, banks held some specie in order to honor this legal requirement.[6] The free banks were fractional-reserve institutions: they held specie that was a fraction of their outstanding notes.

Chart 1
Free Banking in the United States, 1860



The eighteen states shaded had adopted free banking by 1860. Only Michigan, Georgia, and New York did so in the 1830s, with the rest starting in 1849 or later.

Sources: Rockoff (1975, 3) and Thorndale and Dollarhide (1987, 8).

Banks were penalized for failing to convert notes into specie at par value on demand. If a bank failed to redeem its notes at par on demand, a noteholder could formally protest to the banking authority. The bank had a grace period within which it could redeem the protested notes. Otherwise, the banking authority closed the bank and wound up the bank's security deposit. Even if the bank redeemed the notes during the grace period, some states required the bank to make an additional payment to the protesting noteholder for the time and trouble of protesting the notes.

Bank Runs Possible. The requirement that banks convert notes into specie at par on demand created the possibility of a bank run. Because banks held fractional reserves of specie, they could not instantaneously honor all noteholders' requests for specie. Banks could honor such requests only over time as they reduced their outstanding loans or exchanged assets for specie. Hence, noteholders' demand for converting notes into specie could create a liquidity problem for a bank. If noteholders thought it likely that a bank would not be able to continue to convert its notes into specie at par, they had an incentive to exchange the bank's notes for specie. The noteholders then could hold the specie or banknotes issued by another bank and wait to see whether the bank kept its notes convertible into specie.

A more widespread event possible with required convertibility of notes at par is a run on a banking system or, more traditionally and colorfully, a banking panic. For banknotes, a banking panic is a decrease in the demand for banks' notes associated with an increase in noteholders' estimated probability that banks will temporarily or permanently fail to redeem the notes at par value. While "panic" is the traditional name for such an event, people did not generally, if ever, panic in the sense of having "unreasoning fear." Rather, people had good reason to be apprehensive about whether the banks could continue redeeming their notes at par. In a banking panic, unlike a run on an individual bank, noteholders did not simply exchange their specie for notes issued by other banks. Instead, they held the specie or exchanged it for notes issued by banks not in the banking system.

Locations of Banks. Free banks were permitted to have an office at only one location. This restriction did not prevent individuals from owning or operating more than one bank if they so wished, however. Scattered records indicate that some people owned shares in several banks in one or more states, but there is no systematic evidence on how common such ownership was.

Some states required that banks locate their offices in an area with a minimum number of people. For example, Illinois in 1857 adopted a law requiring that banks be located in cities, towns, or villages with at least 200 people, and Wisconsin in 1858 adopted a similar restriction. Apparently, these laws were adopted to prevent banks from locating in out-of-the-way places, thereby hampering redemption of their notes.

Information about Banks' Activities. Free banking laws required that information be made available to the public about the banks' activities. The laws required that banks submit periodic reports, at least annually and sometimes quarterly, and the state banking authority publish the reports in selected newspapers. The state banking authority also had the power to examine the banks and determine the veracity of their reports.

Table 1
Abbreviated Balance Sheet for a Free Bank

Assets		Liabilities	
Bonds deposited with state banking authority	$50,000	Notes	$50,000
Loans to stockholders	15,000	Equity capital	50,000
Specie	5,000		
Loans	30,000		
Total	$100,000		$100,000

In addition to this required information, various trade publications known as banknote reporters specialized in reporting the values of banks' notes.[7] Under typical circumstances, notes could be exchanged for specie at the issuing bank itself at zero discount: one dollar of notes could be exchanged for one dollar of specie. A bank's notes might well be used in transactions at locations far from the bank, though, and rather than trading at par value the notes would trade at discounts from par at such locations. For example, discount rates for Indiana banks' notes were quoted in New York City. If the discount rate was 1.5 percent for a bank's notes, a person in New York City with $100 in Indiana notes could exchange them for $98.50 of specie. A person holding banknotes in New York City or elsewhere could, if it was the more advantageous course, send the banknotes back to the Indiana bank for redemption at their face value in specie.

Usually, the discount reflected the transportation cost and interest forgone due to the time required to return the notes (Charles W. Calomiris and Larry Schweikart 1991; Gary Gorton 1996). As Gorton (1996) shows, new banks had higher discounts on average because new banks had not yet established a reputation for reliably redeeming their notes. In addition, notes issued by banks that were ceasing operations traded at discounts that reflected the interest forgone due to waiting for redemption and the amount that noteholders expected to receive (Gerald P. Dwyer Jr. and Iftekhar Hasan 1996). Notes issued by banks that were likely to fail traded at discounts greater than usual even before closing. These excess discounts reflected the probability of failure, the payment that noteholders expected to receive, and the interest forgone while waiting for the payment.

Early Experience with Free Banking

Of the early adopters of free banking, Michigan and New York provide an informative contrast. Michigan's experience commonly is regarded as a fantastic failure of free banking and New York's as a solid success.[8]

A Fiasco—Michigan in 1837 and 1838. In 1837, shortly after it became a state, Michigan adopted the first free banking law in the United States.[9] This law was based on a proposed bill in New York and generally was similar to the typical free banking environment outlined above.

There were important differences, though, between Michigan's law and later ones. In free banking systems in the United States, the security deposit was the minimal guarantee of the notes' value. Michigan's free banking law provided that this security deposit consist "either of bonds and mortgages upon real estate within this state or in bonds executed by resident freeholders of the state" (Michigan 1837a, "An Act to organize and regulate banking associations," Section 11). Possible problems with bonds and mortgages on real estate, though, include inflated appraisals and depreciated real estate values when selling large amounts of real estate in security deposits. In Michigan, as bank commissioner Alpheus Felch (1880, 120) indicates, real estate values were far below appraised values when the free banks were closing en masse. The subterfuges possible with personal bonds, which are guarantees by individuals, are obvious.[10]

As events unfolded, Michigan's problems were compounded by a nationwide suspension of specie payments shortly after the state adopted free banking. This suspension of payments occurred for reasons unrelated to free banking in Michigan (Richard H. Timberlake 1993), but it affected free banking in the state. In a suspension of specie payments, banks did not redeem their notes, and under Michigan law, such a suspension by any bank implied that the bank must be closed. In 1838, Michigan amended its law to permit banks to suspend specie payments and after the number of banks quickly doubled further amended its law to prohibit new banks from suspending payments.[11] This suspension was especially problematic for the new Michigan banking system, which did not have an established reputation for reliably redeeming its notes. In addition, it probably did not help that Michigan was a frontier state at the time. Before the advent of the telegraph, let alone modern communications, acquiring information was a slower and more expensive process than today, which would compound the lack of information about the new banks and banking system.

The increase in the number of banks in Michigan was large, and their openings were followed quickly by their demises.[12] In January 1837 there were nine banks in Michigan. By December 1837 there were eighteen banks, and two months later, forty. By September 1839, only nine remained (U.S. Congress, 1840-41, 1449). These numbers are only estimates because it was hard for Michigan's bank commissioners to be sure how many banks ever opened (U.S. Congress, 1839-40, 1107, 1128). In a preamble to recommending repeal of the free banking law, the bank commissioners waxed eloquent, claiming,

> The singular spectacle was presented, of the officers of the State seeking for banks in situations the most inaccessible and remote from trade, and finding at every step an increase of labor by the discovery of new and unknown organizations. Before they could be arrested, the mischief was done; large issues were in circulation, and no adequate remedy for the evil. (U.S. Congress 1839-40, 1129)[13]

Commissioners' reports on some banks are readily available along with many accompanying depositions (U.S. Congress 1839-40). In at least a few cases, according to depositions by available bank officials, the banks were started without any intent of ever redeeming notes. In fact, notes were put into circulation without meeting legal requirements such as having the signature of a bank commissioner on the notes or providing the security deposit for the notes. Such activities were illegal under the law, and the simplest interpretation is that the banknotes were fraudulent if not counterfeit.[14]

It is easy to see how issuing notes and absconding with the proceeds could increase the wealth of a bank's organizers. For the cost of printing up notes, the issuer could use the notes to buy other assets and then skip town with those assets. The balance sheet in Table 2 illustrates the strategy. None of the bank's capital is paid in. The bank's capital is exactly offset by a loan to the owners. The notes are created and issued by making a loan to the owners. If the owners provide personal bonds to start the bank or an inflated appraisal on real property, and if they dispose of the notes and avoid their legal liability after the bank closes, they gain by the full amount of the notes' value.

As long as the person initially receiving the notes does not realize that the notes soon will be worthless—otherwise they will not take them—creating the banknotes increases the owners' wealth. Such a situation cannot be expected to persist, however. Receivers of such notes soon will notice their rapid decrease in value and will accept the notes at a discount if they may have a positive value or will not accept them at all if they certainly will be worthless. Free banking's rapid demise in Michigan itself suggests the promptness of such responses.

Estimates of noteholders' losses from these extraordinary developments are rough at best. In January 1839 the bank commissioners estimated that free banks were authorized to issue more than $4 million of notes and that "at a low estimate, near a million dollars of the notes of insolvent banks are due and unavailable in the hands of individuals" (U.S. Congress 1839-40, 1128-29).[15] It is not clear how many notes were issued. The commissioners indicate that "about forty banks" began operation. Their discussions of individual banks indicate that they thought that about seventeen banks had sufficiently large security deposits to cover their notes.[16] Discussing unredeemed notes, an 1839 Attorney General's report suggests that about $2 million was outstanding toward the end of 1839, and free banks redeemed these notes at about 39 cents on the dollar (Hugh Rockoff 1985). This redemption rate suggests that noteholders' losses were about 60 percent of the par value of these notes, or on the order of $1 million to $1.2 million. Noteholders' actual losses in these banks were reduced by any discount on the notes from their face value when issued.

Even though the estimate is rough, an approximate estimate of noteholders' total losses is $1.2 million. If the free banks had issued $4 million in notes at par value, noteholders lost about 30 percent of the par value of all free banks' notes. The 60 percent loss rate based on $2 million of notes is an overestimate, and the 30 percent loss rate based on $4 million of notes probably is an underestimate. In either case, noteholders' losses on banknotes are substantial.

A Success—New York. The events in Michigan are spectacular, but besides not lasting very long themselves, they also did not persist in the sense that they did not reappear in other states. In 1838, while Michigan was suffering through its debacle, New York passed the free banking law that its legislature had been debating for several years. New York's free banking system is widely regarded as notably successful.

Table 2
Abbreviated Balance Sheet for a Michigan Wildcat Bank

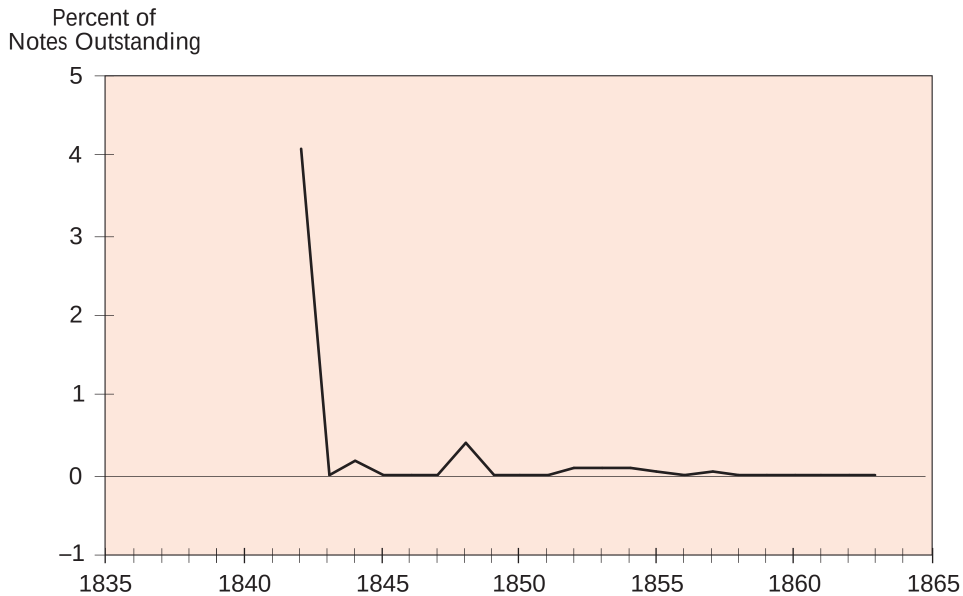
Assets		Liabilities	
Loans to stockholders	$100,000	Notes	$50,000
		Equity capital	50,000
Total	$100,000		$100,000

New York required that banks' security deposits consist of New York state government bonds or bonds and mortgages on real estate. Available evidence suggests that the bonds and mortgages on real estate were less of a problem than in Michigan.

Chart 2 shows losses suffered by noteholders in New York free banks for the years available, 1842 to 1863.[17] The annual loss rates on New York notes were relatively high in the 1840s—4 percent in 1842, 0.2 percent in 1844, and 0.4 percent in 1848—and then never again as high as 0.1 percent. Noteholders' loss rates of less than 0.1 percent in later years are not obviously more than their losses from inadvertently destroying or misplacing notes.

Losses on total notes give a picture of the typical noteholders' losses, but they do not show the losses suffered by those who held notes issued by the banks that failed—banks that ceased operation and paid noteholders less than the par value of their notes. Chart 3 shows the losses per dollar of notes in failed

Chart 2
Loss Rate on Total Notes in New York Free Banks, 1842-63[a]

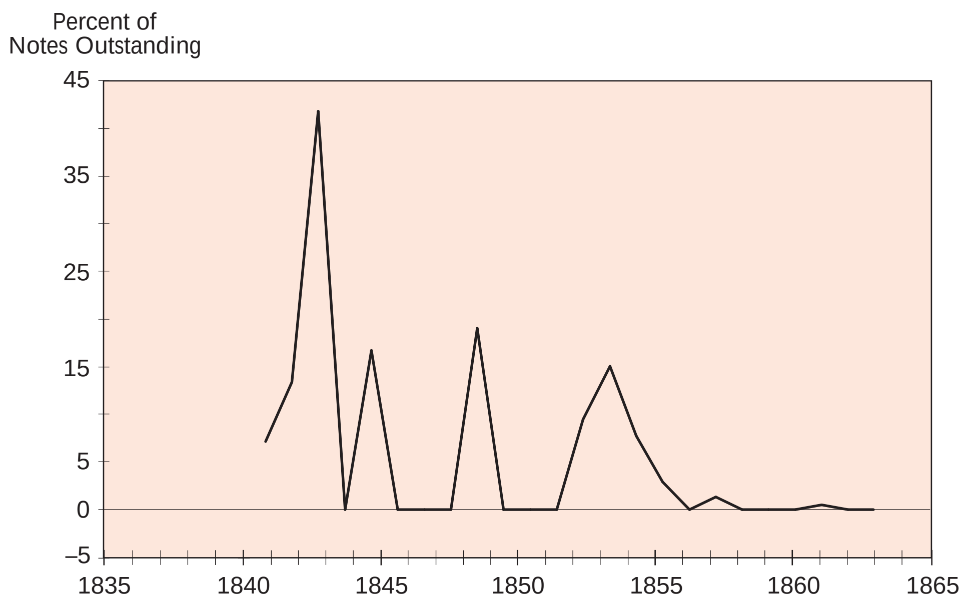


While Michigan's free banking efforts seemed to fail dismally, New York's free banking system is widely regarded as successful.

[a] The only dates for which information is available.

Source: King (1983).

Chart 3
Loss Rate on Notes in Failed New York Free Banks, 1840-63[a]



Losses on total notes (Chart 2) give a picture of the typical noteholders' losses, but they do not show the losses suffered by those who held notes issued by banks that failed, ceasing operation and paying noteholders less than the note's par value.

[a] These data include 1840 and 1841 whereas the aggregate losses shown in Chart 2 do not because data on total notes are not available for those years.

Source: King (1983).

New York free banks from 1840 to 1863.[18] For a few years, noteholders' loss rates on these banks' notes are relatively high. Nonetheless, loss rates on failed banks' notes show the same pattern of declining losses over time as do noteholders' loss rates on all notes. The highest loss rate is 42 percent in 1842, within the range of estimated loss rates for Michigan a few years earlier. In the 1840s, the annual average loss rate is 9.8 percent; in the 1850s, it is 3.7 percent; and in the four years of the 1860s, it is 0.1 percent. Although the loss rate borne by those who held failed banks' notes sometimes is substantial, even this loss rate decreases over time.

It is easy to overstate the significance of these losses. This pattern of zero losses by some and sometimes nontrivial losses by others means that in turn misfortune was borne by some and not by others. During this period, banknote reporters made it relatively low cost to be informed about the value of banks' notes. While the more informed had an incentive to shift these losses to the less informed, in the absence of evidence, it is hard to say more.[19] Even the annual average loss rate on notes in the small proportion of banks that failed in the 1850s is only 3.7 percent. The overall loss rate for that decade is less than 0.1 percent.

Free Banking in Selected States

As Chart 1 indicates, most states that adopted free banking did so in the 1850s, after New York had about fifteen years' experience with the banking system. Besides experiences of the early adopters of free banking, it is informative to examine what happened in selected states that adopted free banking later and apparently had substantial problems.[20] Extreme examples can be the best teachers, but it is important to realize that they are not representative examples. Indiana, Illinois, and Wisconsin had particularly bad experiences with their free banking systems, as indicated by Hammond's conclusion cited earlier that people in these states would have been better off with no banks at all than with free banking.

Bank Entry and Failures. Table 3 summarizes free banks' entry and exit in these three states from the inception of free banking to 1863. In the table, a bank is listed as ceasing operation if it closed and the bank's security deposit was sufficient to redeem all of the bank's notes at their face value. A bank is listed as failing if it closed and the bank's security deposit was not sufficient to redeem all of the bank's notes at their face value.[21] A noticeable aspect of Table 3 is the large amount of entry and exit. Much of this activity simply reflects people starting banks and later closing them because it was optimal to do so. There is no obvious reason to be more concerned about it than to be concerned about grocery stores beginning and ceasing operation.[22]

While the timing differs between the states, the failures in each state are clustered in specific time periods. In Indiana, 68 banks started in the first three years of free banking in the state, and only 38 existed at the end of that period. Eighty-seven percent of the banks in Illinois closed at the start of the Civil War, and most of them failed. Out of 108 banks at the start of 1861 in Wisconsin, 36 failed and 15 ceased operation in the next two years. As mentioned above, these states are not typical. They are chosen for discussion precisely because they have notorious episodes in which many banks closed.

Was free banking in the United States so bad that people would have been better off with no banks at all?

Wildcat Banking. Were the occasional large numbers of banks that failed wildcat banks with reckless or financially unsound operations?[23] By themselves, high failure rates do not mean that banks are operating recklessly. Conversely, banks that remain open for a long period may well be operating recklessly.

Duration before Failure. Although eventual failure rates are unreliable measures of banks' ex ante riskiness, Arthur J. Rolnick and Warren E. Weber (1984) use the duration of a bank's existence as a measure of whether a bank is a wildcat bank. They define a bank as a wildcat bank if it failed within a year after beginning operation. This definition is relatively straightforward, and it is possible to determine whether any particular bank failed shortly after opening. In addition, it focuses on one aspect of wildcat banking: starting a bank and absconding with one-time gains from starting it. Using this definition, Rolnick and Weber (1984) and

Table 3
Free Banks Entering and Exiting in Indiana, Illinois, and Wisconsin, 1853-63

Period	Entry	Ceased Operation	Failed	Closed[a]	Number of Banks
Indiana					
– Dec 53	30				30
Dec 53 – Jul 54	19		1	2	46
Jul 54 – Jan 56	19	7	10	10	38
Jan 56 – Jul 56	2	4		3	33
Jul 56 – Jul 57	5	7	2	3	26
Jul 57 – Jan 58		7			19
Jan 58 – Jan 59	1	4			16
Jan 59 – Jan 60	1				17
Jan 60 – Jan 61	2		1		18
Jan 61 – Jan 62					18
Jan 62 – Jan 63		1			17
Illinois					
– Apr 53	23				23
Apr 53 – Apr 54	8	2			29
Apr 54 – Jan 56	15	9			35
Jan 56 – Nov 56	16	3	2		46
Nov 56 – Jan 58	4	5	2		43
Jan 58 – Oct 58	5	1			47
Oct 58 – Jan 60	29	1			75
Jan 60 – Oct 60	20				95
Oct 60 – Apr 62	5	3	80		17
Apr 62 – Jan 63	8				25
Wisconsin					
– Jan 53	2				2
Jan 53 – Jan 54	8				10
Jan 54 – Jan 55	14	1			23
Jan 55 – Jan 56	10	3			30
Jan 56 – Jan 57	17	2			45
Jan 57 – Jan 58	26	3			68
Jan 58 – Jan 59	34	3			99
Jan 59 – Jan 60	16	8			107
Jan 60 – Jan 61	7	5	1		108
Jan 61 – Jan 62	2	12	35		63
Jan 62 – Jul 62		3	1		59

[a] Unknown whether bank ceased operation or failed. In addition, there are six Indiana banks for which dates of operation are not available. Three of these ceased operation and one failed, and it is unknown whether the other two ceased operation or failed.

Sources: Rolnick and Weber (1982); Economopolous, unpublished data.

Andrew Economopolous (1988, 1990) clearly have shown that wildcat banking was unimportant if not irrelevant in Indiana, Wisconsin, and Illinois.

Rockoff (1975) suggested, for reasons outlined in Box 2, that wildcatting might be due to bonds in the security deposit being valued at par rather than market value. Subsequent research into state laws and regulators' operations in states, though, has found bonds being valued at the lesser of par or market value in every case except New York from 1838 to 1840 (Rolnick and Weber 1984; Economopolous 1988). Legislators in New York and other states learned from New York's problems in its earliest years and did not repeat the mistake.

Bank Owners Highly Leveraged in Their Ownership. Rolnick and Weber's measure of wildcat banking is not informative about whether a bank was operating in a highly leveraged and possibly reckless manner. The president of a competitor of free banks suggested that anyone with relatively little funds could organize a free bank (Hugh McCulloch 1888, 125-26). If a bank's organizer has some funds, borrows more, and uses the funds to buy state bonds, then the organizer can use the state bonds as security for notes. In exchange for the deposited bonds, the state banking authority sends notes to the organizer, and the organizer uses the notes to buy more bonds. This process continues until the organizer uses the notes to pay off the original loan. A possible end result of this process is the balance sheet in Table 4. Nothing in the balance sheet suggests a bank with reckless operations. The bank's ratio of notes to capital is one, and the bank has sufficient bonds deposited to pay off noteholders and a substantial loan portfolio. The loans just happen to be loans to the bank's owners.

Why would anyone organize such a bank? As long as it is solvent, the bank receives the interest on the bonds held by the state banking authority. In the example in Table 4, starting with, say, $5,000, the bank's owner is receiving interest on $50,000 in bonds. The bank's owner is highly leveraged in this transaction, but it is not apparent on the bank's balance sheet.[24] There are three aspects of this operation that are particularly pertinent. First, the bonds held by the state banking authority are available to pay noteholders.

Second, if the loans are collectible, noteholders are covered even against relatively large losses on the state bonds. The owner is liable for the loan to the bank, and, generally speaking, the owner also is liable for

Box 2
Valuation of Bonds Deposited as Security

The valuation of bonds as security for banknotes had important effects on how free banking worked. As Rockoff (1975) points out, if the bond security was valued at more than its market value, individuals had an incentive to buy bonds, issue notes, and abscond with the proceeds. For example, if someone could buy $80,000 worth of bonds at current market prices and the bonds were valued as security at their face value of, say, $100,000, and the notes could be passed for more than $80,000, say $90,000, there is a one-time gain of $10,000 in starting the bank. If the owner could avoid being sued for noteholders' losses, for example by leaving the court's jurisdiction, this difference between the amount received for the notes and the market value of the bonds created an incentive to start a bank and let it fail quickly.

After a few years of free banking's operation, legislators were aware of this incentive. Initially, from 1838 to 1840, bond security in New York was valued at its par value, which can be and was greater than some bonds' market value. In 1840, New York amended its law to require that the bond security be valued at the lesser of par or market value, a requirement followed by other states.

While addressing one problem, this provision of free banking laws was associated with another problem. Because bonds were valued at the lesser of par or market value, everything else the same, banks found it in their interest not to buy bonds trading at prices much above their par value. Bonds purchased at prices above par value could be used to support notes only equal to the bonds' par value. A smaller issue of notes decreased the bank's loans and its income. If banks are attempting to maximize expected income, other things the same, they prefer not to buy bonds trading well above their par value. Banks' risk aversion can, of course, cause banks to buy bonds trading well above par if such bonds are less risky than bonds trading closer to par value. In effect, banks face a trade-off between their risk and their return, which is absent if bonds are valued at market price, no matter what their par value. This provision may explain why Illinois and Wisconsin banks held large amounts of southern bonds, which had unfortunate consequences when prices of those bonds fell at the start of the Civil War.

the amount of equity capital again should it fail. In the example in Table 4, the owner is liable for an additional $95,000 over and above the $5,000 of personal funds invested in the bank. If the bank's owner has substantial additional assets that are difficult to move beyond the jurisdiction of the state's courts, the owner can be forced to make these payments with the result that the noteholders are unlikely to suffer losses.

Third, this banking operation has substantial value to the bank's owners *as long as the bank continues to operate* because the owner continues to receive interest on the state bonds. The owner has no incentive to abscond with funds because the bank's positive present value is due solely to continuing to receive the interest on the bonds, not because of any one-time gain from starting the bank.

Table 4
Abbreviated Balance Sheet for a Free Bank,
With the Owner Highly Leveraged in Its Organization

Assets		Liabilities	
Bonds deposited with state banking authority	$50,000	Notes	$50,000
Loans to stockholders	45,000	Equity capital	50,000
Specie	5,000		
Total	$100,000		$100,000

Nonetheless, there is a sense in which the bank is a risky venture for noteholders: the ultimate funds available to noteholders are the security deposit and the owner's assets, not the security deposit and a more diversified loan portfolio. Unfortunately, little direct can be said about the value of loan diversification to noteholders without detailed data on banks' loan portfolios and bank owners' assets.

Remote Location. A quite different but simple way of thinking about wildcatting is in terms of the word's apparent origin: remote locations that hamper note redemption. Such locations can be associated with reckless operations because outside knowledge about such banks' operations might be quite limited. Did free banks locate in remote and inaccessible places?

Chart 4 shows maps of Indiana, Illinois, and Wisconsin in 1860, indicating the population and the number of banks in each county. There is no obvious pattern to the location of Indiana banks, other than perhaps some tendency for them to be along the Ohio River on the southern boundary and along state borders generally. Each county with three banks has a major town: from north to south, Indianapolis, Evansville, and Terre Haute. Banks in Wisconsin generally are located in the more populous and more accessible downstate counties. Banks in Illinois generally are located in the southern part of the state near the Ohio River, across the Mississippi River from St. Louis and in Bloomington. A striking aspect of the map for Illinois is the almost complete absence of free banks in the most populous county, Cook County. There were many private banks in Cook County, and these banks made loans there. Illinois free banks themselves made fewer loans than other banks because a usury law applied to free banks but not other lenders, including private banks (F. Cyril James 1938, 233). Free banks could circumvent the usury law by lending their notes to affiliated private banks that made loans at higher interest rates. Hence, while Illinois banks located in accessible locations, they apparently found it expedient to issue notes from offices in less populous locations than Chicago.

Episodic Factors External to the Banking Systems. If wildcat banking is not the explanation of why so many banks closed, what is? In the case of Indiana, a change in a law in Ohio was the initiating factor in the Indiana free banks' problems. Indiana banknotes circulated in other states, and evidence suggests that a relatively large amount of Indiana banknotes was used in transactions in Ohio, partly because of relatively high taxes on banks in Ohio (Hasan and Dwyer 1994, 275-78). Indiana's free banks encountered difficulties when Ohio passed a law in May 1854 that made it illegal as of October 1, 1854, for anyone in Ohio to use small banknotes issued by banks in other states. This decrease in the demand for Indiana banknotes resulted in the return of the notes for redemption and decreases in the prices of Indiana bonds, which were about two-thirds of the banks' security deposits.

Chart 5 shows prices of Indiana bonds with a 5 percent coupon for this period. For comparison, Chart 5 also shows the prices of U.S. government bonds and other state government bonds with data available for at least half of the period. As the chart shows, Indiana bond prices were above 96 percent of par through the middle of August 1854, after which they fell about 10 percent for two months. The trough in bond prices is in December 1854. This decrease is after the change in

Ohio's law and coincides with the organized expulsion of notes from Ohio. In the absence of any other developments concerning Indiana's debt, the timing suggests that the decrease in the demand for Indiana banknotes and consequently Indiana bonds was a result of the Ohio law.[25]

In 1861, however, the decrease in bond prices occurred before the Illinois and Wisconsin free banks' difficulties and is an important factor in those difficulties.[26] Chart 6 shows prices of bonds that were 10 percent or more of the aggregate portfolio of banks in either state in 1861; it also shows U.S. bond prices for comparison.[27] All of these bonds have 6 percent coupon rates.[28] The prices of southern and border state bonds fell before the Civil War and then fell dramatically the same week that Confederates fired on Fort Sumter and Lincoln responded by ordering a blockade and calling up troops.[29] The low prices occurred in June 1861, and bond prices increased thereafter to the end of 1862. While bond sales by banks may have affected the bond prices' movements, the Civil War itself is the initiating factor that resulted in many banks failing in Illinois and Wisconsin.

Banking Panics. Even though the initiating factors are different in 1854 in Indiana and in 1861 in Illinois and Wisconsin, subsequent events are strikingly similar. In each of these instances, there was a banking panic that affected banks in the banking system.

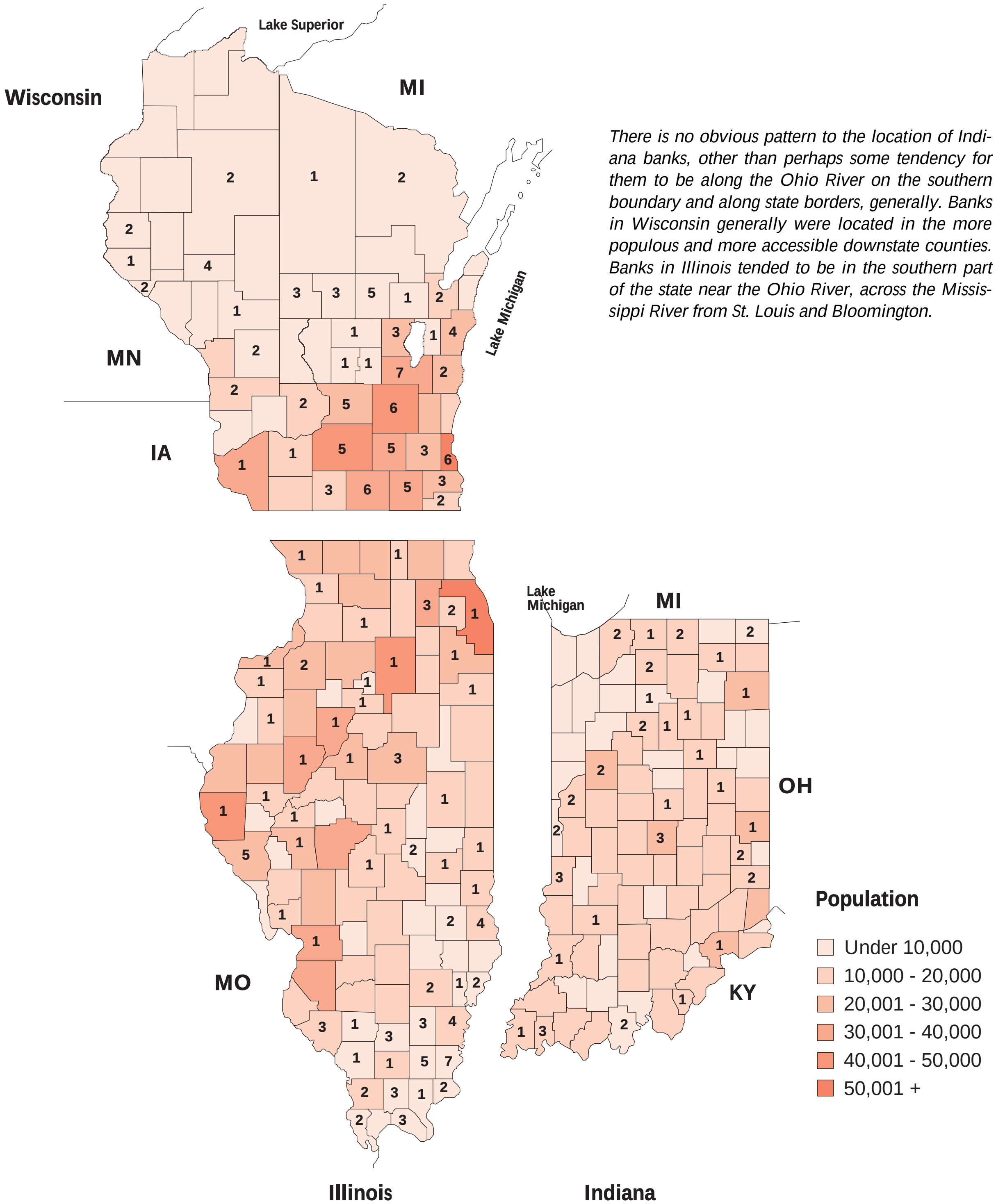
In all three states, discount rates on notes in banknote reporters indicate that the market value of all banks' notes fell quite substantially. Table 5 shows discount rates for banknotes in each of these states during these episodes. In Indiana at the end of 1853, the discount rates on banknotes were 1.5 percent. By December 1854, almost 90 percent of the Indiana free banks had discount rates of 25 percent or more. A typical New York City holder of an Indiana bank's notes lost almost 25 percent of the notes' value. This loss reflected a change from a situation with expected redemption on demand at face value to a nonzero probability of the bank closing, with delayed redemption of the notes and the possibility of receiving less

Table 5
Discount Rates on Notes and Changes in Notes Outstanding in Indiana in 1854, Illinois in 1861, and Wisconsin in 1861

	Indiana	Illinois	Wisconsin
		Discount Rates	
Date	12/53	6/60	6/60
Discount rate	1.50	2.25	2.75
Percent of banks with this discount rate and higher	100	97.5	97.1
Date	12/54	6/61	6/61
Discount rate	25	60	20
Percent of banks with this discount rate and higher	89	100	100
		Percentage Change in Banknotes Outstanding	
Period	10/54 to 1/56	1/60 to 1/62	1/60 to 1/62
Percentage change	–44.7	–84.2	–59.8

Sources: Discount rates in 1853 and 1854 are from *Thompson's Bank Note and Commercial Reporter*, December 15, 1853, and December 1, 1854. Discount rates in 1860 and 1861 are from *Hodge's Journal of Finance and Bank Reporter*, June 9, 1860, and June 22, 1861. The data on Indiana, Illinois, and Wisconsin banknotes are from U.S. Congress (1863-64, Table 2, 216-17).

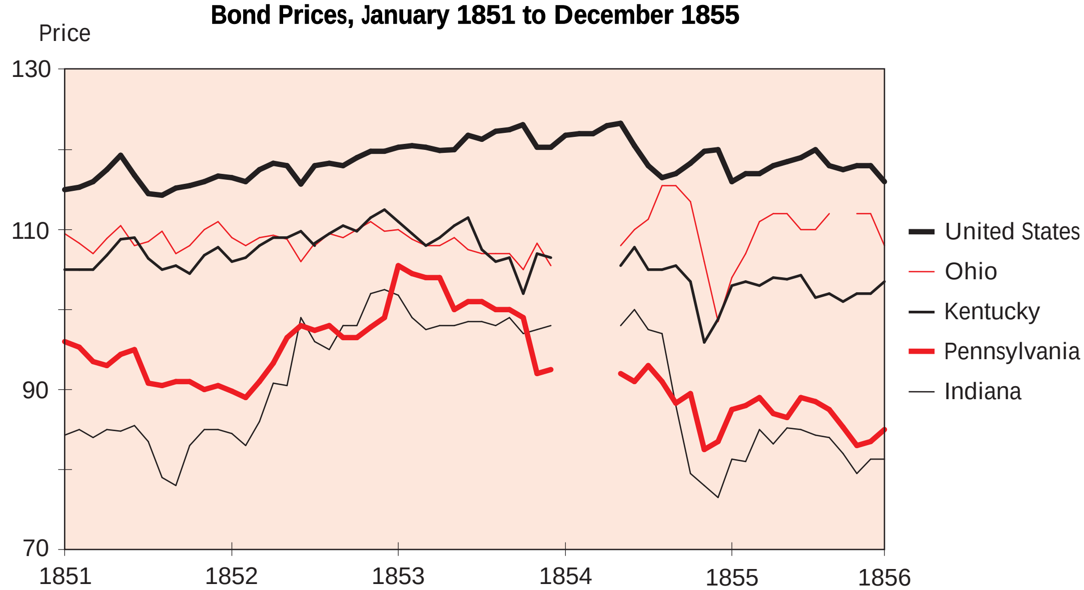
Chart 4
Population and Number of Free Banks by County in Wisconsin, Illinois, and Indiana, 1860



There is no obvious pattern to the location of Indiana banks, other than perhaps some tendency for them to be along the Ohio River on the southern boundary and along state borders, generally. Banks in Wisconsin generally were located in the more populous and more accessible downstate counties. Banks in Illinois tended to be in the southern part of the state near the Ohio River, across the Mississippi River from St. Louis and Bloomington.

Source: Thorndale and Dollarhide (1987, 381 [Wisconsin], 105 [Illinois], and 112 [Indiana]).

Chart 5
Bond Prices, January 1851 to December 1855

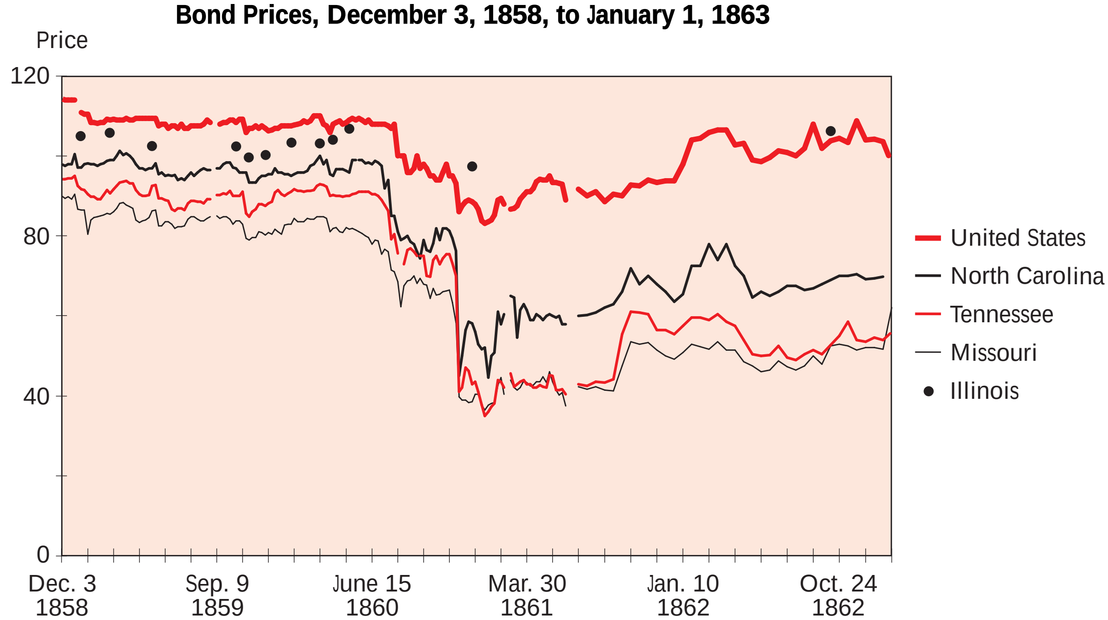


Indiana's free banks encountered difficulties when Ohio passed a law in May 1854 that made it illegal as of October 1,1854, for anyone in Ohio to use small banknotes issued by banks in other states. This decrease in demand for Indiana banknotes resulted in the return of the notes for redemption and decreases in the prices of Indiana Bonds.

Note: Gaps in state data indicate that data were unavailable.

Source: See data appendix (available on request).

Chart 6
Bond Prices, December 3, 1858, to January 1, 1863



The prices of southern and border state bonds fell before the Civil War and then fell dramatically the same week that the Confederates fired on Fort Sumter.

Note: Only fragmentary data on Illinois bond prices are available.

Source: See data appendix (available on request).

than the notes' face value. In Illinois and Wisconsin in 1861, quite different initiating developments—the onset of the Civil War—had similar effects.

These discount rates are greater than noteholders' losses. While loss rates are not known for all Indiana banks that ceased operations in 1854 and 1855, noteholders' average loss rate even on notes issued by a typical bank known to have failed is 12 percent, and the maximum known loss rate on notes issued by an Indiana bank that failed in 1854 and 1855 is 20 percent. This average loss rate in failed banks is far less than the discount rates of at least 25 percent on almost all banks' notes and also is small in comparison with losses in the 1830s in Michigan and losses in 1842 in New York.[30] Holders of notes from a typical bank in Wisconsin suffered losses of about 7.2 percent, and holders of Illinois notes suffered larger losses, about 22.2 percent.

These developments in all three states also are followed by substantial contractions in the amount of notes outstanding. From October 1854 to January 1856, Indiana banknotes outstanding fell by about 45 percent. From January 1860 to January 1862, Wisconsin banknotes outstanding fell by about 60 percent and Illinois banknotes fell by an even larger 84 percent.

In response to these developments, bankers attempted to coordinate their responses and reassure noteholders that some banks were solvent. In Indiana in 1854 and in Wisconsin in 1861, the free banks suspended payments.[31] A detailed comparison of Illinois and Wisconsin indicates that the suspension of payments had substantial effects (Dwyer and Hasan 1996). The suspension of payments explains much of the difference between 87 percent of Illinois banks closing and 47 percent of Wisconsin banks closing. In addition to decreasing the number of banks that ceased operations, the joint suspension decreased noteholders' losses by about 20 percentage points. Besides being similar in the 1854 and 1861 episodes, bankers' coordinated responses, including the suspensions of payments, are similar to bankers' responses to runs on the banking system in the National Banking period.[32]

Conclusion

Free banking in the United States was not the disaster portrayed by some, but it also was not problem-free. The early years of free banking were troubled. Holders of Michigan notes lost 30 to 60 percent of the notes' value. In 1842 holders of New York notes lost 4 percent of the notes' value and holders of failed banks' notes lost 42 percent. With the exception of episodic events that generated atypical losses, free banking's performance improved over time. This improvement is associated with, and possibly due to, adjustments in the laws in response to problems that arose. In the 1850s, a substantial number of states adopted free banking laws.

Free banking in Indiana, Illinois, and Wisconsin are alleged later instances of reckless banking. There is no evidence that free banks in these states generally were characterized by continuing fraud to transfer wealth from passive noteholders to shrewd bankers. There also is little evidence supporting a generalization that these free banks were imprudent, let alone financially reckless. The episodic difficulties faced by free banks were not self-induced implosions. In these instances, banks' losses occurred sporadically when developments outside the banking systems decreased the demand for the banks' notes or decreased the value of the banks' assets. These episodic difficulties resulted in banking panics, and bankers, legislators, and bank regulators dealt with the panics in ways that anticipated developments in the subsequent National Banking period.

Free banking disappeared when it was taxed out of existence by the federal government in 1865. This action was not due to apparent dissatisfaction voiced by citizens of free banking states. In fact, the national banking law adopted during the Civil War included many provisions similar to the free banking laws. Nonetheless, it is an open question whether some feasible banking system other than free banking would have improved people's well-being in free banking states.

Notes

1. "Free banking" is the name used for these banking laws at the time, and this usage of the term is clear in context. This period was not one of laissez-faire banking, in which the only laws applied to banks are those applying to similar firms whether or not they are financial institutions. Free banking laws in the United States included many detailed provisions of the laws that applied to banking and not other businesses, some of which had unfortunate effects.

   It is ironic that the banks in the United States most similar to laissez-faire banks, private banks, have received little study. It is difficult to know even how many private banks there are at any time, let alone anything about them. Because private banks are not incorporated, do not have limited liability, and are subject only to general laws, there is very little documentary evidence, and none of it is readily available. Individuals or partnerships in the United States long have been unable to issue notes, but private banks face the same issues in the deposit and loan business as do today's commercial banks.
2. According to the *New Shorter Oxford English Dictionary*, the use of the word *wildcat* for a reckless or unsound operation arose in the early 1800s. The usual basis of the name, as in Hammond (1957, 600-601), for example, is the explanation in the text. Dillistin (1949, 60-63) argues for a different, strained interpretation.
3. Rockoff (1975) was the first economist in many years to examine U.S. free banking. L. White (1984) explored free banking in Britain, including its intellectual history, and Rolnick and Weber (1983; 1984; 1985; 1988) wrote an influential series of papers investigating U.S. free banking. In recent years, there has been a torrent of research on free banking all over the world. Dowd (1992) includes nine papers on some of these free banking episodes. Selgin and White (1994) survey much of the research on free banking. This research into free banking is part of an examination of basic issues concerning monetary and banking systems analyzed in recent years by Hayek (1978), Friedman and Schwartz (1986), Goodhart (1988), and others. Other studies include a classic analysis by Smith ([1936] 1990) and more recent analyses by Bordo and Schwartz (1995), Goodhart (1994), Roberds (1995), Schwartz (1993), and Selgin (1993; 1994).
4. This summary is based on Dewey (1910), Hammond (1957), Rolnick and Weber (1984), Hasan and Dwyer (1994), and Dwyer and Hasan (1996).
5. Loans to stockholders generally are not so obvious on available free banks' balance sheets.
6. Although they legally could demand it, noteholders did not necessarily require specie in exchange for the banknotes. They often accepted notes issued by other banks.
7. Dillistin (1949) provides detailed information on the reporters, and Gorton (1996) provides an economic analysis of the discount rates.
8. Georgia is the remaining state that adopted free banking in the 1830s. Georgia never had more than two free banks, however; hence, the history of free banking in Georgia is not particularly informative and is not examined in this paper. Schweikart (1987) and Scott (1989) provide overviews of banking in Georgia before the Civil War.
9. There are no histories of banking in Michigan that include this period. The available information is limited because fire destroyed the Michigan bank commissioners' records (Rolnick and Weber 1983, 1089). Felch's (1880) recollections of this period, during which he was a legislator and a bank commissioner, provide an informative but prejudiced overview. The reports by the bank commissioners printed in the House Executive Documents (U.S. Congress 1837-38, 1839-40) also are informative. Shade (1972) examines the relationship between banks and politics in the Old Northwest: Ohio, Indiana, Illinois, Michigan, and Wisconsin.
10. These problems apparently became clear quickly. The original banking bill including personal bonds in the security deposit was approved March 15, 1837, but was amended to include only bonds and real estate mortgages on December 30, 1837 (Michigan 1838, "An Act to amend an act entitled 'An Act to organize and regulate banking associations' and for other purposes," Section 6).
11. The laws are "An Act suspending, for a limited time certain provisions of law, and for other purposes," approved June 22, 1837 (Michigan 1837b), and "An Act to amend an act entitled 'An act suspending for a limited time certain provisions of law, and for other purposes'," approved December 28, 1837 (Michigan 1838).
12. Shade (1972, 36-37) indicates that the Michigan legislature granted nine new charters in 1836 in addition to the existing charters and passed the free banking law after receiving eighteen requests for new charters in its 1837 session.
13. Given today's banking laws or, for that matter, later free banking laws, it is natural to suppose that banks were required to inform the bank commissioners before opening. This was not the case, though. Free banks in Michigan were required to file applications with the treasurer and clerk of the county in which they intended to open their office, not with the bank commissioners (Michigan 1837a, "An Act to organize and regulate banking associations," Section 1).
14. Dillistin (1949, chap. 2) is the best single source on counterfeiting of free banks' notes.
15. At least one of the commissioners, Alpheus Felch, was not favorably disposed to free banking. He was one of four legislators out of thirty-nine to vote against the original free banking law (Felch 1880, 115; Shade 1972, 37). He also was one of the Supreme Court justices who ruled in litigation in 1844 that the free banking law was unconstitutional (Rockoff 1985, 886). This $1 million estimate seems to be the estimate that Felch (1880) relies on, contrary to Rockoff's supposition (1985, 887).
16. These evaluations range from tentative ones of "hope no loss" to definite ones of "no possible loss."
17. These losses are the difference between the par value of the notes and the dollar amount received from the banking

regulator and do not allow for the forgone interest in the meantime or later recoveries from the banks or their stockholders.

18. These data include 1840 and 1841, whereas the aggregate losses do not, because reliable data on total notes are not available for 1840 and 1841 (King 1983, 147).
19. In Wisconsin in 1861, banks decided not to accept ten banks' notes at par and announced it only after some businesses had paid workers in those banks' notes (Krueger 1933, 82-85). The result was a riot.
20. At the start of the Civil War, Tennessee free banks had problems similar to those in Illinois and Wisconsin, but the surviving data do not include noteholders' losses (Pierce and Horning 1991).
21. Noteholders may have been paid the face value of their notes even if the bank's security deposit was insufficient to redeem the notes at face value. The available information from the states' archives is on note redemption by the security deposit, which does not include information on payments from other sources. Even if a noteholder was paid face value, the payment was delayed and the present value would have been less than the face value. There is insufficient information available to reliably calculate such present values. Not having such present values, though, is a second-order problem compared with not having information on all payments to noteholders.
22. Increased entry, though, can be associated with increased competition, which is desirable. On the basis of raw numbers, Ng (1988) suggests that free banking did not increase bank entry. Using an economic model, though, Economopolous and O'Neill (1995) provide evidence that free banking did increase entry. Bodenhorn (1990) presents evidence that free banking was also associated with more changes in banks' market ranks. Kahn's (1985) computations indicate that free banks had a shorter life expectancy than chartered banks, which is not obviously undesirable anyway. These computations are vitiated, though, by an assumption that the probability of closing is the same every year, an assumption grossly at variance with the data.
23. Rockoff (1975, 4-5) defines a wildcat bank as a free bank that cannot continuously redeem its notes. He later (1991, 96-103) elaborates on his views of wildcat banking and distinguishes them from Rolnick and Weber's views.
24. Interestingly, Bodenhorn and Haupert (1995) provide evidence that free banks issued too few notes to maximize their net revenue.
25. The relative price of Indiana bonds rose in 1852 at least partly due to a change in the way interest was paid. Before 1853, one-fifth of the interest on the bonds was paid in bonds on which interest would not be paid until 1853. Indiana began paying all of the promised interest in 1853.
26. Details are provided in Rolnick and Weber (1984), Economopolous (1988), Hasan and Dwyer (1994), Dwyer and Hasan (1996), and the earlier work referenced in these papers.
27. Movements of northern states' bond prices generally are similar to movements of U.S. bond prices.
28. Only fragmentary data on Illinois bonds prices are available in *Bankers' Magazine,* the source of the bond prices. There is no evidence of changes in bond prices specific to Illinois in Chart 6, although nontrivial transitory changes could be concealed by the paucity of observations. There is no evidence, though, of events other than the bank failures that might have affected prices of Illinois bonds.
29. It is less surprising that Missouri bonds fell as much as southern bonds when one recalls that Missouri was under martial law with a provisional state government for the duration of the war (Brownlee 1958). Ratchford (1941, 124-25) indicates that Missouri paid no interest on its bonds from the outbreak of the Civil War until the ratification of a Reconstruction-era constitution in 1866.
30. The loss rates for Michigan and New York are weighted-average loss rates for all banks, and the loss rates for Indiana, Illinois, and Wisconsin are simple average loss rates across banks.
31. As in Michigan in 1838, the state legislatures suspended the provision of the free banking laws that would have revoked banks' charters because they failed to redeem their notes.
32. Dwyer and Gilbert (1989) and Calomiris and Gorton (1991) summarize these later episodes. Sprague (1910) and Friedman and Schwartz (1963) provide detailed information and analysis.

References

Bass, Thomas A. "The Future of Money." Interview with Walter Wriston. *Wired* 4 (October 1996): 140-43, 200-05.

Bodenhorn, Howard. "Entry, Rivalry, and Free Banking in Antebellum America." *Review of Economics and Statistics* 72 (November 1990): 682-86.

Bodenhorn, Howard, and Michael Haupert. "Was There a Note Issue Conundrum in the Free Banking Era?" *Journal of Money, Credit, and Banking* 27 (August 1995): 702-12.

Bordo, Michael D., and Anna J. Schwartz. "The Performance and Stability of Banking Systems under 'Self-Regulation': Theory and Evidence." *Cato Journal* 14 (Winter 1995): 453-79.

Brownlee, Richard S. *Gray Ghosts of the Confederacy: Guerrilla Warfare in the West, 1861-1865*. Baton Rouge, La.: Louisiana State University Press, 1958.

Calomiris, Charles W., and Gary Gorton. "The Origins of Banking Panics: Models, Facts, and Bank Regulation." In *Financial Markets and Financial Crises,* National Bureau of Economic Research Project Report, edited by R. Glenn Hubbard, 109-73. Chicago and London: University of Chicago Press, 1991.

Calomiris, Charles W., and Larry Schweikart. "The Panic of 1857: Origins, Transmission, and Containment." *Journal of Economic History* 51 (December 1991): 807-34.

Dewey, Davis R. *State Banking before the Civil War*. National Monetary Commission. Washington: Government Printing Office, 1910.

Dillistin, William H. *Bank Note Reporters and Counterfeit Detectors*. New York: American Numismatic Society, 1949.

Dowd, Kevin. *The Experience of Free Banking*. London: Routledge, 1992.

Dwyer, Gerald P., Jr., and R. Alton Gilbert. "Bank Runs and Private Remedies." Federal Reserve Bank of St. Louis *Review* 71 (May/June 1989): 43-61.

Dwyer, Gerald P., Jr., and Iftekhar Hasan. "Suspension of Payments, Bank Failures, and the Nonbank Public's Losses." Federal Reserve Bank of Atlanta Working Paper 96-3, May 1996.

Economopolous, Andrew. "Illinois Free Banking Experience." *Journal of Money, Credit, and Banking* 20 (May 1988): 249-64.

——. "Free Bank Failures in New York and Wisconsin: A Portfolio Analysis." *Explorations in Economic History* 27 (1990): 421-41.

Economopolous, Andrew, and Heather O'Neill. "Bank Entry during the Antebellum Period." *Journal of Money, Credit, and Banking* 27 (November 1995): 1071-85.

Felch, Alpheus. "Early Banks and Banking in Michigan." *Pioneer Collections: Report of the Pioneer Society of the State of Michigan* 2 (1880): 111-24.

Fischer, Gerald C. *American Banking Structure*. New York: Columbia University Press, 1968.

Flohr, Udo. "Electric Money." *Byte* 21 (June 1996): 74-84.

Friedman, Milton. *A Program for Monetary Stability*. New York: Fordham University Press, 1960.

Friedman, Milton, and Anna J. Schwartz. *A Monetary History of the United States, 1867-1960*. Princeton, N.J.: Princeton University Press, 1963.

——. *Monetary Statistics of the United States*. New York: National Bureau of Economic Research, 1970.

——. "Has the Government Any Role in Money?" *Journal of Monetary Economics* 17 (January 1986): 37-62.

Goodhart, Charles. *The Evolution of Central Banks*. Cambridge, Mass.: MIT Press, 1988.

——. "The Free Banking Challenge to Central Banks." *Critical Review* 8 (Summer 1994): 411-25.

Gorton, Gary. "Reputation Formation in Early Bank Notes." *Journal of Political Economy* 104 (April 1996): 346-97.

Hammond, Bray. *Banks and Politics in America from the Revolution to the Civil War*. Princeton, N.J.: Princeton University Press, 1957.

Hasan, Iftekhar, and Gerald P. Dwyer Jr. "Bank Runs in the Free Banking Period." *Journal of Money, Credit, and Banking* 26 (May 1994): 271-88.

Hayek, F.A. *The Denationalization of Money—The Argument Refined*. 2d ed. London: Institute of Economic Affairs, 1978.

Hodge's Journal of Finance and Bank Reporter, June 9, 1860, and June 22, 1861.

James, F. Cyril. *The Growth of Chicago Banks*. New York: Harper and Brothers Publishers, 1938.

Kahn, James A. "Another Look at Free Banking in the United States." *American Economic Review* 75 (September 1985): 881-85.

King, Robert G. "On the Economics of Private Money." *Journal of Monetary Economics* 12 (July 1983): 127-58.

Krueger, Leonard Bayliss. *History of Commercial Banking in Wisconsin*. University of Wisconsin Studies in the Social Sciences and History, no. 18. Madison: University of Wisconsin, 1933.

Levy, Steven. "E-Money (That's What I Want)." *Wired* 2 (December 1994): 174-79, 213-19.

McCulloch, Hugh. *Men and Measures of Half a Century*. New York: Charles Scribners Sons, 1888.

Michigan. *Acts of the Legislature of the State of Michigan Passed at the Annual Session of 1837*. Detroit, Mich.: John S. Bagg, 1837a.

——. *Acts of the Legislature of the State of Michigan Passed at the Special Session of 1837*. Detroit, Mich.: John S. Bagg, 1837b.

——. *Acts of the Legislature of the State of Michigan Passed at the Adjourned Session of 1837 and the Regular Session of 1838*. Detroit, Mich.: John S. Bagg, 1838.

Ng, Kenneth. "Free Banking Laws and Barriers to Entry in Banking, 1838-1860." *Journal of Economic History* 48 (December 1988): 877-89.

Pierce, Sarah B., and Bruce C. Horning. "The Free Banking Era in Tennessee." Vanderbilt University. Unpublished paper, 1991.

Ratchford, B.U. *American State Debts*. Durham, N.C.: Duke University Press, 1941.

Roberds, William. "Financial Crises and the Payments System: Lessons from the National Banking Era." Federal Reserve Bank of Atlanta *Economic Review* 80 (September/October 1995): 15-31.

Rockoff, Hugh. *The Free Banking Era: A Reexamination*. New York: Arno Press, 1975.

——. "New Evidence on Free Banking in the United States." *American Economic Review* 75 (September 1985): 886-89.

——. "Lessons from the American Experience with Free Banking." In *Unregulated Banking: Chaos or Order?*, edited by Forrest H. Capie and Geoffrey E. Wood, 73-109. New York: St. Martin's Press, 1991.

Rolnick, Arthur J., and Warren E. Weber. "The Free Banking Era: New Evidence on Laissez-Faire Banking." Federal Reserve Bank of Minneapolis Research Department Staff Report 80, 1982.

——. "New Evidence on the Free Banking Era." *American Economic Review* 73 (December 1983): 1080-91.

——. "The Causes of Free Bank Failures." *Journal of Monetary Economics* 14 (October 1984): 269-91.

——. "Banking Instability and Regulation in the U.S. Free Banking Era." Federal Reserve Bank of Minneapolis *Quarterly Review* 9 (Summer 1985): 2-9.

——. "Explaining the Demand for Free Bank Notes." *Journal of Monetary Economics* 21 (January 1988): 47-71.

Schneier, Bruce. *Applied Cryptography*. 2d ed. New York: John Wiley and Sons, Inc., 1996.

Schwartz, Anna J. "Are Central Banks Necessary?" *Critical Review* 7 (Spring-Summer 1993): 355-70.

Schweikart, Larry. *Banking in the American South from the Age of Jackson to Reconstruction*. Baton Rouge, La.: Louisiana State University Press, 1987.

Scott, Carole E. "Antebellum Banking in Georgia and South Carolina." In *Essays in Economic and Business History*, edited by Edwin J. Perkins. Selected Papers from the Economic and Business History Historical Society, vol. 8. Los Angeles: Department of History, University of Southern California, 1989.

Selgin, George. "The Rationalization of Central Banks." *Critical Review* 7 (Spring-Summer 1993): 335-54.

——. "Are Banking Crises Free-Market Phenomena?" *Critical Review* 8 (Fall 1994): 591-608.

Selgin, George, and Lawrence H. White. "How Would the Invisible Hand Handle Money?" *Journal of Economic Literature* 32 (December 1994): 1718-49.

Shade, William Gerald. *Banks or No Banks: The Money Issue in Western Politics, 1832-1865*. Detroit, Mich.: Wayne State University Press, 1972.

Smith, Vera C. *The Rationale of Central Banking*. 1936. Reprint. Indianapolis: Liberty Press, 1990.

Sprague, O.M.W. *History of Crises under the National Banking System*. U.S. National Monetary Commission. Washington: Government Printing Office, 1910.

Thompson's Bank Note and Commercial Reporter, December 15, 1853, and December 1, 1854.

Thorndale, William, and William Dollarhide. *Map Guide to the U.S. Federal Censuses, 1790-1920*. Baltimore, Md.: Genealogical Publishing Co., Inc., 1987.

Timberlake, Richard H. *Monetary Policy in the United States: An Intellectual and Institutional History.* Chicago: University of Chicago Press, 1993.

U.S. Congress. House. Executive Document 79. 25th Cong., 2d sess., 1837-38.

——. Executive Document 172. 26th Cong., 1st sess., 1839-40.

——. Executive Document 111. 26th Cong., 2d sess., 1840-41.

——. Executive Document 20. 38th Cong, 1st sess., 1863-64.

White, Eugene N. "Free Banking, Denominational Restrictions, and Liability Insurance." In *Money and Banking: The American Experience,* 99-117. The George Edward Durell Foundation. Fairfax, Va.: George Mason University Press, 1995.

White, Lawrence H. *Free Banking in Britain: Theory, Experience, and Debate: 1800-1845.* Cambridge: Cambridge University Press, 1984.

Wriston, Walter. "Money: Back to the Future." *Wall Street Journal*, November 24, 1995, A8.

——. "Money—Back to the Future?" Lecture given at Claremont-McKenna College, November 20, 1996.

Options and Volatility

Peter A. Abken and Saikat Nandi

Volatility is a measure of the dispersion of an asset price about its mean level over a fixed time interval. Careful modeling of an asset's volatility is crucial for the valuation of options and of portfolios containing options or securities with implicit options (for example, callable Treasury bonds) as well as for the success of many trading strategies involving options. The problem of pricing options is one confronting not only option traders but also, increasingly, a broad spectrum of investors. In particular, institutional investors' portfolios frequently contain options or securities with embedded options. More and more, risk-management practices of financial institutions as well as of other corporate users of derivatives require frequent valuation of securities portfolios to determine current value and to gauge portfolios' sensitivities to market risk factors, including changes in volatility (see Peter A. Abken 1994). A model of volatility is needed for managing portfolios containing options (including derivatives and other securities containing options) for which market quotes are not readily available and that consequently must be marked to model (that is, valued by model) rather than marked to market. Accurate assessments of volatility are also key inputs into the construction of hedges, which limit risk exposures, for such portfolios.

Because of the central role that volatility plays in derivative valuation and hedging, a substantial literature is devoted to the specification of volatility and its measurement. Modeling volatility is challenging because volatility in financial and commodity markets appears to be highly unpredictable. There has been a proliferation of volatility specifications since the original, simple constant-volatility assumption of the famous option pricing model developed by Fischer Black and Myron S. Scholes (1973). This article gives an overview of different specifications of asset price volatility that are widely used in option pricing models.

Abken and Nandi are senior economists in the financial section of the Atlanta Fed's research department.

The Effect of Volatility

A simple example will illustrate the importance of volatility for options. Consider a call option that gives the holder of the option the right to buy one unit of a stock at a future date T at a particular price (also called the strike or the exercise price of the option). Let the strike price of the option, denoted by K, be equal to \$50. Note that the value of a call option at maturity is given by $\max(S_t - K, 0)$, where S_t denotes the stock price at the maturity of the option. Thus, the call option at its maturity has a value equal to the difference between the stock price at maturity and the strike price if $S_t > K$ and zero otherwise. If the stock price at maturity of the option is less than the strike price, the optionholder would rather buy the stock from the market than exercise the option and pay the higher strike price K for the stock. Now consider the following two stock price scenarios in which an option exists that has a strike price of \$50:

High-Volatility Scenario

Stock Price	\$30	\$40	\$50	\$60	\$70
Option Payoff	0	0	0	\$10	\$20

Low-Volatility Scenario

Stock Price	\$40	\$45	\$50	\$55	\$60
Option Payoff	0	0	0	\$ 5	\$10

The average stock price (across the five states of the world) is \$50 in both scenarios, but volatility is higher in the first scenario because of the wider dispersion of possible stock prices. In contrast to the constant average stock price in each scenario, the average option payoff is \$7.50 in the low-volatility scenario and \$15.00 in the high-volatility scenario. The reason is that the downside of the payoff to the optionholder is limited to zero because the option does not have to be exercised if the stock price at maturity is less than the strike price. The optionholder merely loses the price paid to the option writer (or seller) for the purchase of the option. However, the optionholder gains if the stock price at maturity is greater than the strike price. The higher the volatility, the higher is the probability that the option payoff at maturity will be greater than the strike price and consequently will be of value to the optionholder. Very high stock prices can increase the value of the call option at maturity without limit. However, very low stock prices cannot make the value of the option payoff at maturity less than zero. Thus the asymmetry of the payoffs due to the nature of the contract implies that volatility is of value to the optionholder at the maturity of the option. In general, since the price of the option prior to its maturity is the expectation of the option payoff at maturity (discounted at an appropriate rate), an increase in the volatility of the underlying asset increases the expectation—and consequently the price of the option today.

In general, future volatility is difficult to estimate. While the historical volatility of an asset return is readily computed from observed asset returns (see Box 1), this measure may be an inaccurate estimate of the future volatility expected to prevail over the life of an option. The future volatility is unobservable and may differ from the historical volatility. Hence, unlike the other parameters that are important for pricing options (namely, the current asset price, the strike price, the interest rate, and time to maturity), the volatility input has to be modeled. For example, the Black-Scholes option pricing model is a simple formula involving these five variables that prices European options. (Such options can be exercised only upon maturity. See John C. Cox and Mark Rubinstein 1985 for an exposition of the Black-Scholes formula.) The Black-Scholes model assumes that volatility is constant, the simplest possible approach. However, a preponderance of evidence (see Tim Bollerslev, Ray Y. Chou, and Kenneth F. Kroner 1992) points to volatility being time-varying. In addition, that variation may be random or, equivalently, "stochastic." Randomness means that future volatility cannot be readily predicted using current and past information.

Before proceeding with an overview of the various approaches for treating time-varying volatility, the discussion examines a frequently documented phenomenon known as the volatility smile to motivate the consideration of different volatility specifications. The existence of the smile is an indication of the inadequacy of the constant-volatility Black-Scholes model. A common feature of all time-varying volatility models reviewed below is that they have the potential to give prices that are free of the Black-Scholes biases, such as the smile.

For the Black-Scholes model, the only input that is unobservable is the future volatility of the underlying asset. One way to determine this volatility is to select a value that equates the theoretical Black-Scholes price of the option to the observed market price. This value is often referred to as the implied (or implicit) volatility of the option. Under the Black-Scholes model, implied volatilities from options should be the same regardless of which option is used to compute the volatility. However, in practice, this is usually not the case. Different options (in terms of strike prices and maturities) on the same asset yield different implied volatilities, outcomes that are inconsistent with the Black-Scholes model. The pattern of the Black-Scholes implied volatilities with respect to strike prices has become known as the volatility smile. The existence of a smile also means that if only one volatility is used to price options with different strikes, pricing errors will be systematically related to strikes. The smile has also been shown to depend on options' maturities.

The existence of pricing biases for the Black-Scholes model has been well documented. These biases have varied through time. For example, Rubinstein (1985) reports that short-maturity out-of-the-money calls on equities have market prices that are much higher than the Black-Scholes model would predict. On the other hand, since the stock market crash of 1987, the volatility smile has had a persistent shape, especially when derived from equity-index option prices—as the strike price of index-equity options increases, their implied volatilities decrease. Thus, an out-of-the-money put (or in-the-money call) option has a greater implied volatility than an in-the-money put (or out-of-the-money call) of equivalent maturity.

Buying an out-of-the-money put can serve as insurance against market declines. The surprising severity of the market crash of 1987 increased the cost of crash protection, as manifested by a relatively high cost for out-of-the-money put options. Because the option price, for calls or puts, increases as volatility rises, higher option prices are associated with higher implied volatilities. Thus, relatively high out-of-the-money put prices are mirrored in high implied volatilities for those options.

The smile in equity index options is often referred to as a skew because the high implied volatilities for out-of-the-money puts (or, equivalently, for in-the-money calls) progressively decline as puts become further in-the-money (or as calls become further out-of-the-money). Box 2 gives more detail about the volatility skew in the S&P 500 index-equity options.

This article reviews two overarching approaches to generalizing the constant-volatility assumption of the Black-Scholes model that have appeared in the option pricing literature. Both lines of research have developed concurrently. The first approach assumes that variations in volatility are determined by variables known to market participants, such as the level of the asset price. Models of this type are referred to as deterministic-volatility models. This approach contrasts with the second, more demanding one, commonly called stochastic volatility, in whicππh the source of uncertainty that generates volatility is different from, although possibly correlated with, the one that drives

Box 1
Measurement of Historical Volatility

The standard way to measure volatility from asset prices is straightforward. Assuming no intermediate cash flows like dividend payments, suppose $r_t = (P_t - P_{t-1})/P_{t-1}$ represents the return of an asset as measured by buying the asset at time $t-1$ at P_{t-1} and selling it at t at P_t, that is, over a single period of time. Further assume that these returns are not dependent on each other: the fact that today's return is high or low reveals nothing about tomorrow's return. A statistical description of the behavior of returns is that they are just different realizations of a random variable that is evolving through time. If these single-period asset returns are calculated from time $t = 1$ to time $t = T$, then the mean or average return over this period, denoted by $\bar{r}$, is estimated by

$$\bar{r} = \frac{\sum_{t=1}^{T} r_t}{T}. \tag{1}$$

The above equation is the symbolic representation for summing the returns from $t = 1$ to $t = T$ and then dividing by the length of the time interval (T) over which the returns are measured. Having estimated the mean, the standard deviation, a measure of volatility, is estimated by

$$\sigma = \sqrt{\frac{\sum_{t=1}^{T} (r_t - \bar{r})^2}{T-1}}. \tag{2}$$

Variance, denoted by σ^2, is the square of the above quantity and is also a measure of volatility. In other words, in order to estimate the standard deviation, one sums the squared deviations of the individual returns from the mean return, divides by $T - 1$, and takes the positive square root of the resultant quantity. This measure is usually referred to as historical volatility. The relevant volatility for pricing options is not that which occurred in the past but that which is expected to prevail in the future. However, historical volatility may be useful in forming that expectation inasmuch as volatility is correlated through time.

asset prices. Therefore, knowledge of past asset prices is not sufficient to determine volatility (using discretely observed prices). For reasons that will be explained more fully below, the first approach has been the most popular as a modeling strategy because of its relative simplicity.

The models under consideration in this article have been developed for equity, currency, and commodity options. Practitioners have also used the Black-Scholes model to price and hedge such options. Stochastic volatility is even more challenging to incorporate into models of fixed-income securities because of the complexities of modeling the term structure of interest rates. The Black-Scholes model has not been the benchmark model for pricing options in fixed-income markets, and less work has been done on stochastic-volatility bond pricing models. Thus, this topic is beyond the scope of this article.

Within the first approach, three types of models have been proposed. These are (1) implied binomial tree models, (2) general autoregressive conditional heteroscedasticity (GARCH) models, and (3) exponentially weighted moments models. Although somewhat arcane sounding on first reading, each will prove to have its own intuitive appeal. Each type of model also has the potential to closely or exactly match model option prices with actual market option prices. For the second approach involving stochastic volatility, models may be divided into those that have closed-form solutions for option prices and those that do not. Closed-form solutions refer to pricing formulas that are readily computed, given current computer technology. The distinction is a matter of practicality because the time it takes to compute prices is relevant to practitioners who trade options or hedge positions using options. Advances in computer technology will gradually blur the distinction among current stochastic-volatility models as time-consuming computations become less so in the future.

Deterministic Volatility

The first approach that has been used to address the deficiencies of the benchmark Black-Scholes model and the volatility smile involves deterministic-volatility specifications. The Black-Scholes formula for pricing European options is predicated on the assumption of constant volatility. The simplest relaxation of the constant volatility assumption is to allow volatility to depend on its past in such a way that future volatility can be perfectly predicted from its history and possibly other observable information. As an example, suppose the variance of asset returns σ^2_{t+1} is described by the following equation:

$$\sigma^2_{t+1} = \theta + \kappa\sigma^2_t.$$

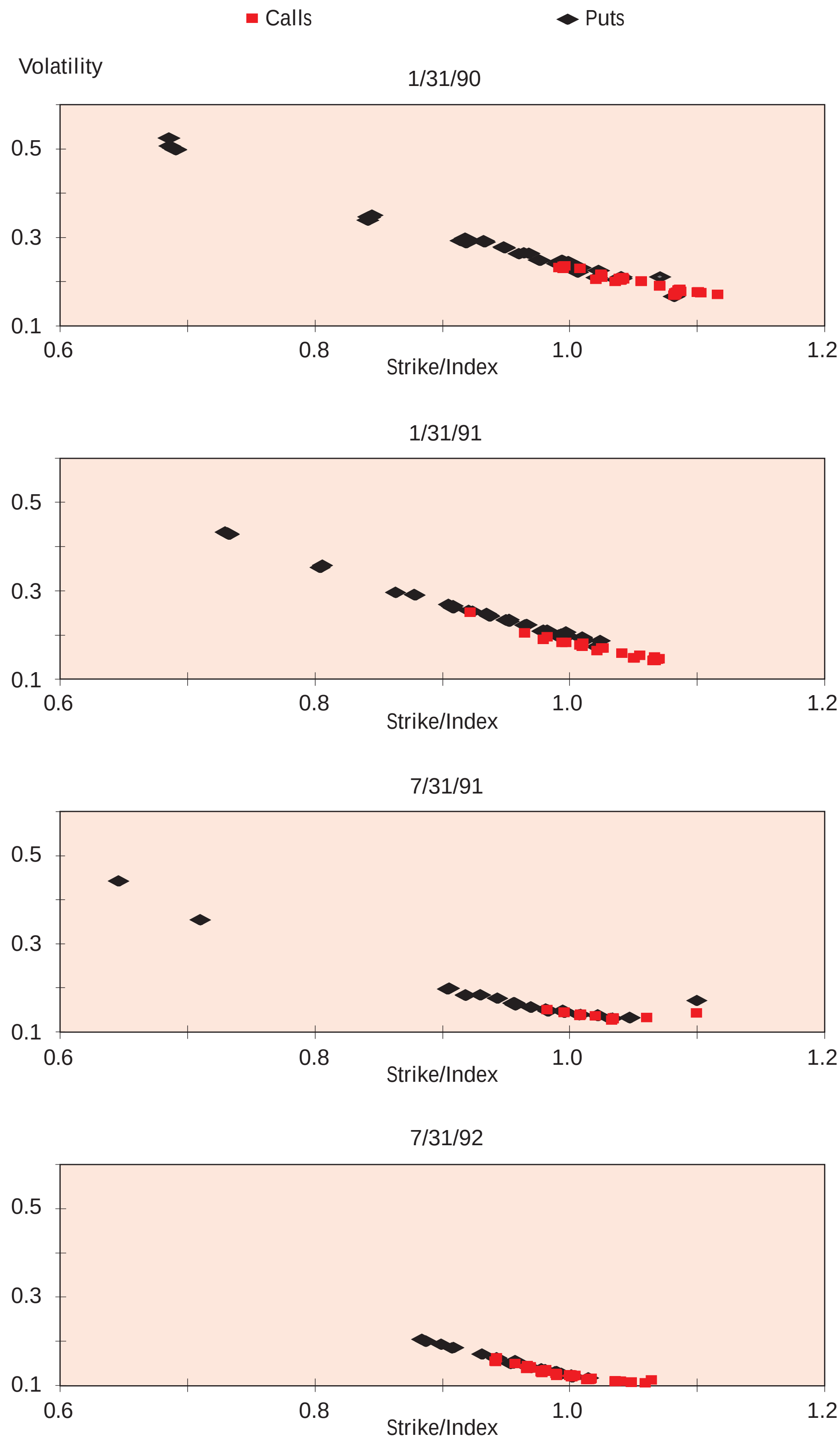
Box 2
The Volatility Skew in S&P 500 Index Options

The chart illustrates the skew on four different days.[1] The skew is computed from S&P 500 index options that are traded at the Chicago Board Options Exchange (CBOE). These are standard European options, for which exercise can occur only on the option expiration date, and their payoffs are determined by the level of the S&P 500 index on the option maturity date. The Black-Scholes equation is used to infer the volatility using the other option formula inputs and the quoted option price. Each chart contains implied volatilities from puts and calls that were traded between 10:00 A.M. and 2:30 P.M. All options had forty-five days to maturity. Diamonds are the implied volatilities derived from individual put transactions, and squares are implied volatilities from individual call transactions. The volatilities are plotted against the ratio of the option strike to the index level. Thus, a value of one corresponds to puts or calls being at the money. Ratios less than one represent strike prices that are out-of-the-money for puts and in-the-money for calls. The most noticeable feature of each of these plots is that the deep-out-of-the-money puts have implied volatilities substantially above the volatilities of other options. These volatilities decline almost linearly as the strike-index ratio increases. Similar smile effects have been observed in interest rate option markets (see Amin and Morton 1994 for Eurodollar futures options and Abken and Cohen 1994 for Treasury bond futures options) and in foreign exchange markets (Bates 1995).

Note

1. These prices are from a CBOE data base that covers the years 1990-92. More recent smiles computed from settlement price data have the same shape as those illustrated in the charts.

Volatility Smiles
(*Options Quotes from 10:00-14:30*)



The future volatility depends on a constant and a constant proportion of the last period's volatility. In this case, the constant variance of the asset returns in the Black-Scholes formula can be replaced by the average variance that is expected to prevail from time t until time T (the expiration time), which is approximately given by

$$\frac{1}{T-t}\sum_{u=t}^{T}\sigma_u^2,$$

and the Black-Scholes formula can continue to be used.

A more general case specifies volatility as a function of other information known to market participants. One alternative of this kind posits volatility as a function of the level of the asset price: $\sigma(S)$. One particular model of this type, known as the constant elasticity of variance (CEV) model, in which volatility is proportional to the level of the stock price raised to a power, appeared early in the option pricing literature (Cox and Steve Ross 1976). However, the CEV model proved not to be free of pricing biases (David Bates 1994). A more recent variation on this volatility specification was developed by Rubinstein (1994). Instead of assuming a particular form of the volatility function, Rubinstein's method effectively infers the dependence of volatility on the level of the asset price from traded options at all available strike prices. He calls the model "implied binomial trees" because the implied risk-neutral distribution (which depends on the volatility) of the asset price at maturity is inferred from option prices by constructing a so-called binomial tree for movements of the asset price.[1] (See Box 3 for a discussion of risk-neutral valuation.) Related models have been proposed by Emanuel Derman and Iraz Kani (1994), Bruno Dupire (1994), and David Shimko (1993).

In a recent empirical test of deterministic-volatility models, including binomial tree approaches, Bernard Dumas, Jeffrey Fleming, and Robert Whaley (1996) show that the Black-Scholes model does a better job of predicting future option prices. The option delta, which is derived from an option pricing model and measures the sensitivity of the option price to changes in the underlying asset price, can be used to specify positions in options that offset underlying asset price movements in a portfolio. The authors demonstrate that the Black-Scholes model resulted in better hedges than those from models based on deterministic-volatility functions.

For their tests based on using S&P 500 index options prices, they conclude that "simpler is better" (20). The authors note that one reason for the better performance of the Black-Scholes model is that errors, from various sources, in quoted option prices distort parameter estimates for deterministic-volatility models and consequently degrade these models' predictions. However, hedging performance, which is a key consideration for risk managers and traders alike, has not been systematically tested across all option pricing models. As noted below, other research indicates that some versions of stochastic-volatility models may outperform the simple Black-Scholes model in terms of hedging.

ARCH Models. Autoregressive conditional heteroscedasticity (ARCH) models for volatility are a type of deterministic-volatility specification that makes use of information on past prices to update the current asset volatility and have the potential to improve on the Black-Scholes pricing biases. The term *autoregressive* in ARCH refers to the element of persistence in the modeled volatility, and the term *conditional heteroscedasticity* describes the presumed dependence of current volatility on the level of volatility realized in the past. ARCH models provide a well-established quantitative method for estimating and updating volatility.

ARCH models were introduced by Robert F. Engle (1982) for general statistical time-series modeling. An ARCH model makes the variance that will prevail one step ahead of the current time a weighted average of past squared asset returns, instead of equally weighted squared returns, as is done typically to compute variance (see Box 1). ARCH places greater weight on more recent squared returns than on more distant squared returns; consequently, ARCH models are able to capture volatility clustering, which refers to the observed tendency of high-volatility or low-volatility periods to group together. For example, several consecutive abnormally large return shocks in the current period will immediately raise volatility and keep it elevated in succeeding periods, depending on how persistent the shocks are estimated to be. Assuming no further large shocks, the cluster of shocks will have a diminishing impact as time progresses because more distant past shocks get less weight in the determination of current volatility.

Some technical features of ARCH models also make them attractive compared with many other types of option pricing models that allow for time-varying volatility. In an ARCH model, the variance is driven by a function of the same random variable that determines the evolution of the returns.[2] In other words, the random source that affects the statistical behavior of returns and volatility through time is the same.

Box 3
Risk-Neutral Valuation

The risk-neutral approach to option valuation was pioneered by Cox and Ross (1976) and then developed systematically by Harrison and Kreps (1979) and Harrison and Pliska (1981). It was motivated by the observation that the Black-Scholes option pricing formula does not depend on any parameters that reflect investors' preferences toward risk—that is, their risk-return trade-offs. The key assumption is merely that investors prefer more wealth to less wealth. In particular, the option price does not depend on the expected return of the asset, which is determined by investor preferences. Since the option price does not depend on investors' attitudes toward risk, the same option price will result irrespective of the form of investor preferences. A very convenient preference is "risk neutrality." A risk-neutral investor cares only about the average level of wealth that can be attained by trading in a risky asset and pays no attention to the associated risk. If investors are risk-neutral, then in equilibrium the expected returns on all assets in the economy have to equal the risk-free rate; otherwise, investors would attempt to buy (sell) those securities that have expected returns greater (less) than the return on the risk-free rate, driving the expected return to equality with the risk-free rate. Therefore, under risk neutrality, the dynamics of the returns process—that is, the statistical behavior of returns through time—has to be adjusted to make the mean return on the risky asset equal to the risk-free rate.

As an example, consider an asset whose returns process is described by the following equation:

$$r_t = \mu + \sigma_t \epsilon_{1,t}, \tag{1}$$

where $\epsilon_{1,t}$ is a random variable that is distributed normally with mean zero and variance of unity (a unit normal random variable). This equation is sometimes called the law of motion or dynamics for the return process. The mean return on the asset is μ. The realizations of the random variable $\epsilon_{1,t}$ make the returns r_t (at time t) different from μ, and these realizations are referred to as innovations. The above equation can be rewritten using a different normal random variable $\omega_{1,t}$, with zero mean and unit variance,

$$r_t = rf_t + \sigma_t\, \omega_{1,t}, \tag{2}$$

where rf_t is the risk-free rate. Thus, under the law of motion governed by the innovation process, $\omega_{1,t}$, the mean return of the asset, equals rf_t. For option pricing, the law of motion of the asset returns that is relevant is (2) and not (1). Since the mean return of the asset under (2) is the risk-free rate, (2) is also known as the law of motion of the asset under the risk-neutral distribution—an environment in which all risky assets have expected returns equal to the risk-free rate.

One of the key results of option pricing theory is that the price of an option, or any financial claim that has an uncertain future payoff, is given by the mathematical expectation of its payoff at its maturity, discounted at the risk-free rate. The computation of this expectation assumes that the returns of the asset follow risk-neutral dynamics, such as the example given by equation (2).

If there is a second random variable that affects the price of the option, then, as in the previous example, the mean of that state variable is adjusted to give the dynamics of the state variable in a risk-neutral world. Suppose the variance σ_t^2 follows the random process

$$\sigma_t^2 = \sigma_{t-1}^2 + \kappa(\theta - \sigma_{t-1}^2) + \gamma\sigma_t\, \epsilon_{2,t}, \tag{3}$$

where $\epsilon_{2,t}$ is a standard normal random variable. The above equation is the discrete-time counterpart of the continuous-time variance process given in Heston (1993), in which the variance "reverts" to its long-term mean θ at rate κ, and the volatility of the variance itself is measured by γ. A risk-neutralized representation of the above process analogous to (2) is

$$\sigma_t^2 = \sigma_{t-1}^2 + \kappa^*(\theta^* - \sigma_{t-1}^2) + \gamma\sigma_t\epsilon_{2,t}^*. \tag{4}$$

The shock $\epsilon_{2,t}^*$ is another standard normal random variable, and κ^* and θ^* are obtained from κ and θ by a risk-adjustment procedure (see Heston 1993). In this case, the value of the option is equal to the mathematical expectation under the risk-neutral distribution as generated by (2) and (4), although the statistical behavior of returns and variance in the real world is generated by (1) and (3).

The risk-neutral distribution itself can be inferred from traded option prices. See Abken (1995) for a basic illustration and Abken, Madan, and Ramamurtie (1996) and Aït-Sahalia and Lo (1995) for advanced approaches.

As a result, volatility can be estimated directly from the time series of observed returns on an asset. In contrast, the direct estimation of volatility from the returns process is very difficult using stochastic-volatility models.

There are many different types of ARCH models that have a wide variety of applications in macroeconomics and finance. In finance, the two most popular ARCH processes are generalized ARCH (GARCH) (Bollerslev 1986) and exponential GARCH (EGARCH) (Daniel B. Nelson 1991). The technical distinctions are beyond the scope of this article; however, researchers have tended mostly to use the GARCH process and its variations for option pricing.[3] Although GARCH captures the evolution of the variance process of asset returns quite well, it turns out that there is no easily computable formula, like the Black-Scholes formula, for European option pricing under a GARCH volatility process. Instead, computer-intensive methods are used to simulate the returns and the volatility under the risk-neutral distribution in order to compute European option prices and hedge ratios. (Recent examples include Kaushik Amin and Victor Ng 1993 and Jin C. Duan 1995.)

Owing to the lack of efficient pricing and hedging formulas for GARCH models, practitioners—and some researchers—often substitute the expected average variance from a GARCH model for the variance input in the Black-Scholes formula (see Engle, Alex Kane, and Jaesun Noh 1994). However, the Black-Scholes formula does not hold if the variance of asset returns follows a GARCH process; such a substitution is theoretically inconsistent but may work in practice. Another problem with using the extant GARCH option pricing models is that they do not value American options, which account for most of all traded options. American options can be exercised at any time before maturity, and consequently their prices equal or exceed the prices of comparable European options by the value of this extra flexibility, termed the early-exercise premium. A simple approximation is achieved by adding an estimate of the early-exercise premium to the European price derived from a GARCH model. (There are numerical methods, such as Monte-Carlo simulations, that can value American options, but these methods are currently impractical because of the enormous number of computations required.) The value of the early-exercise premium is often evaluated using the Barone-Adesi-Whaley (1987) formula for the Black-Scholes model.

An early test of a GARCH option pricing model is Engle and Chowdhury Mustafa (1992), who examined S&P 500 index options. Their results show that the GARCH pricing model cannot account for all of the pricing biases observed in the option market. Engle, Kane, and Noh (1994) compared the trading profits resulting from a particular trading rule by using two alternatives for the variance forecasts needed for Black-Scholes: the variance forecast from a GARCH model and the variance forecast in the form of the Black-Scholes implied volatility from a previous period. As noted above, plugging a GARCH forecast into the Black-Scholes formula is ad hoc; however, in an experiment using S&P 500 index options, Engle, Kane, and Noh produced greater hypothetical trading profits using the GARCH volatility forecast than they did using the Black-Scholes implied volatility.

To summarize, although GARCH is a good description of the evolution of the variance process of the asset returns, option pricing models based on GARCH are computationally demanding and may not be very useful for many practitioners given current computing technology. In addition, only a limited number of empirical tests have been done to date on GARCH option pricing models; as a consequence, it is hard to say how well the model does in pricing options and evaluating hedge ratios.[4]

Exponentially Weighted Moments Models. David G. Hobson and L.C.G. Rogers (1996) propose a new type of option pricing model for time-varying volatility that also has the potential to match the observed volatility smile. Their mathematical specification allows past asset-price movements to feed back into current volatility. This characteristic has some of the flavor of a GARCH model in terms of a similar feedback effect; however, the type of feedback can be much more general than encountered in standard GARCH models. Also like GARCH, but unlike standard stochastic-volatility models, there is only one source of uncertainty that drives both the asset price and its volatility.[5]

The Hobson-Rogers model captures past asset price volatility through a so-called offset function. The feedback relationship is primarily embodied in the functional dependence of the volatility on the offset function. The intuition behind the offset function is apparent from its form:

$$S_t^{(m)} = \sum_{u=1}^{\infty} \lambda e^{-\lambda u} (Z_t - Z_{t-u})^m,$$

where $S_t^{(m)}$ is the value of the function at time t and m is the order of the function.[6] This function simply weights deviations of a transformed current price Z_t (a "discounted" logarithm of the price) from its value u

periods ago, $(Z_t - Z_{t-u})$, raised to the power m. The power applied to the deviation, or order of the offset function, is technically the statistical moment of the offset that is employed. For example, a first-order offset function (m = 1) considers the deviation itself, whereas a second-order offset function takes the squares of those deviations and therefore consists of a measure related to the variances of those deviations. The weighting is done by an exponential function that through the parameter λ places more or less importance on the past relative to the present. A high value for λ implies that recently experienced changes in the asset price have a much greater impact on volatility (and the drift) than more distant past shocks. This weighting is similar to the treatment of past return shocks in ARCH modeling. A low λ gives relatively more weight to the past shocks. The persistence of past shocks λ can be estimated indirectly from options prices.

The feedback mechanism in this model works primarily through the asset price volatility, which can take any number of functional forms. Hobson and Rogers consider one simple form in detail in their paper. They show that even a simple version of the offset function, with m = 1, can give option prices that when substituted into the Black-Scholes equation generate a volatility smile in implied Black-Scholes volatilities evaluated at different strike prices, mimicking the smile observed in actual markets.

The impact of the Hobson-Rogers assumption about the volatility specification and the persistence of volatility on option prices needs to be evaluated empirically to see how it compares with Black-Scholes or any other model. The model's ability to trace out a smile is suggestive and may indicate the model's potential to match actual prices well; an empirical evaluation of this model has not been performed to date.

Stochastic Volatility

Stochastic volatility implies that the future level of the volatility cannot be perfectly predicted using information available today. The popularity of stochastic volatility in option pricing grew out of the fact that distributions of the asset returns exhibit fatter tails than those of the normal distribution (Benoit Mandelbrot 1963 and Eugene F. Fama 1965). In other words, the observed frequency of extreme asset returns is much higher than would occur if returns were described by a normal distribution. Stochastic-volatility models can be consistent with fat tails of the return distribution. The occurrence of fat tails would imply, for example, that out-of-the-money options would be underpriced by the Black-Scholes model, which assumes that returns are normally distributed. However, the fat-tailed asset return distributions can also come from ARCH-type volatility as well as from jumps in the asset returns (Robert C. Merton 1976). Stochastic-volatility models could also be an alternative explanation for skewness of the return distribution. Despite the relative complexity of stochastic-volatility models, they have been popular with researchers, and additional justification for these models has recently come to light in the literature on asymmetric information about the future asset price and its impact on traded options.[7]

In a stochastic-volatility model, volatility is driven by a random source that is different from the random source driving the asset returns process, although the two random sources may be correlated with each other. In contrast to a deterministic-volatility model in which the investor incurs only the risk from a randomly evolving asset price, in a stochastic-volatility environment, an investor in the options market bears the additional risk of a randomly evolving volatility. In a deterministic-volatility model, an investor can hedge the risk from the asset price by trading an option and a risk-free asset based on a risk exposure computed using an option pricing formula (see Cox and Rubinstein 1985). (Equivalently, the option's payoff can be replicated by trading the underlying asset and a risk-free asset.) However, with a random-volatility process, there are two sources of risk (the risk from the asset price and the volatility risk); a risk-free portfolio cannot be created as in the Black-Scholes model. After hedging, there is a residual risk that stems from the random nature of the volatility process. Since there is no traded asset whose payoff is a known function of the volatility, volatility risk cannot be perfectly hedged. In order to bear this volatility risk, rational investors would demand a "volatility risk" premium, which has to be factored into option prices and hedge ratios.[8]

A feature of stochastic-volatility models that is not shared by deterministic-volatility models is that the price of an option can change without any change in the level of the asset price. The reason is that the option price is driven by two random variables: the asset price and its volatility. In stochastic-volatility models, these two variables may not be perfectly correlated, implying that the expected volatility over the life of the option may change without any change in the asset price. The change in volatility alone can cause the option price to change.

Most stochastic-volatility models assume that volatility is mean reverting. That is, although volatility varies from day to day, there is a presumed long-run level toward which volatility settles in the absence of additional shocks. Market participants refer to this feature as “regressing to the mean” of the volatility. (The evidence for this phenomenon is especially strong in markets for interest rate derivatives. See, for example, Robert Litterman, Jose Scheinkman, and Laurence Weiss 1991 and Amin and Andrew Morton 1994.)

Stochastic-volatility models can be classified into two broad categories: those that lack closed-form solutions for European options and those that have closed-form solutions.[9] Even if a model’s parameters are known, most stochastic-volatility option pricing models are computationally demanding for pricing European options and especially so for pricing American options. A notable exception is the model of Steven Heston (1993) that gives closed-form solutions for prices and hedge ratios of European options. All other models compute option prices either by numerically solving a complicated partial differential equation or by Monte Carlo simulation. However, many key parameters are not readily estimated from data, particularly those of the volatility process, because, unlike the returns process of the underlying asset, the volatility process is not directly observable. Since parameter estimation is often time-consuming, the lack of readily computed solutions for option prices in many stochastic-volatility models can compound the difficulties of estimation.

The modeling of volatility and its dynamics is a difficult task because the path of volatility during the life of an option is highly unpredictable.

Although stochastic-volatility pricing models give only closed-form solutions for European options, a good approximation for the price of an American option can be obtained by adding an early exercise premium using the Barone-Adesi-Whaley approximation in the same way as for ARCH models. Examples of this practice are in Hans J. Knoch (1992) and Bates (1995). At present, the only other way to price American options under stochastic volatility is by solving a second-order partial differential equation (Angelo Melino and Stuart Turnbull 1992), which is extremely computationally burdensome.

Stochastic-Volatility Option Models without Closed-Form Solution. John C. Hull and Alan White (1987), Louis O. Scott (1987), and James B. Wiggins (1987) were among the first to develop option pricing models based on stochastic volatility. Hull and White as well as Scott made the questionable assumption that the risk premium of volatility is zero—that is, the volatility risk is not priced in the options market—and that volatility is uncorrelated with the returns of the underlying asset. Wiggins, who also assumed a zero-volatility risk premium, found that the estimated option values under stochastic volatility were not significantly different from Black-Scholes values, except for long maturity options. For equity options, Christopher Lamoureux and William Lastarapes (1993) offer evidence against the assumption of a zero-volatility risk premium. For currency options, Melino and Turnbull (1992) found that a random-volatility model yields option prices that are in closer agreement with the observed option prices than those of the Black-Scholes model. While the numerical methods and computers currently available allow computation of these stochastic-volatility option prices, they are still largely impractical for determining hedge ratios, which are vital to market-makers, dealers, and others. As a result, these stochastic-volatility models may not currently be useful for practitioners. Nevertheless, development of stochastic-volatility models continues as researchers attempt to find more tractable models.

Stochastic-Volatility Models with Closed-Form Solutions. Elias M. Stein and Jeremy C. Stein (1991) develop a European option pricing model under stochastic volatility that is somewhat easier to evaluate than the models described above.[10] Although less computationally expensive than the other models, the authors make the unrealistic assumption of zero correlation between the volatility process and the returns of the underlying asset.

Heston (1993) was the first to develop a stochastic-volatility option pricing model for European equity and currency options that can be easily implemented, is computationally inexpensive, and allows for any arbitrary correlation between asset returns and volatility.[11] The model gives closed-form solutions not only for option prices but also for the hedge ratios like the deltas and the vegas of options. (Delta and vega measure the sensitivity of the option price to changes in the asset price and to changes in the volatility, respectively. Knowledge of these measures enables the construction of hedges for options or for portfolios containing embedded options.)

In this model, the asset returns r_t and the variance σ_t^2 are assumed to evolve through time as

$$r_t = \mu + \sigma_t \epsilon_{1,t}$$

and

$$\sigma_t^2 = \sigma_{t-1}^2 + \kappa(\theta - \sigma_{t-1}^2) + \gamma\sigma_t\epsilon_{2,t},$$

respectively, where $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are two standard normal random variables that could be correlated with each another, either positively or negatively, with a correlation coefficient, ρ. Equivalently, this coefficient also measures the correlation between the return of the asset and the volatility process.

In this model, the variance evolves through time in such a way that its long-run average level is measured by θ and the speed with which it is pulled toward this long-run mean is measured by κ, also known as the mean-reversion coefficient. The variable γ is a measure of the volatility of variance. If γ is zero, the model simplifies to a time-varying deterministic-volatility model. In the finance literature, this process for the volatility is also known as a square-root volatility process. The particular nature of the process ensures that volatility "reflects" away from zero: if volatility ever becomes zero, then the nonzero κ ensures that volatility will become positive.

Note that σ_t^2 in this model is not directly comparable to the implied variance from the Black-Scholes model. The reason is that σ_t^2 represents the instantaneous variance (at time t), whereas the implied variance in the Black-Scholes model is the average expected variance through the life of an option and need not equal the instantaneous variance if the model is not true. In Heston's model, the average expected variance during the life of an option is a function of the instantaneous variance, the long-run average variance, the speed with which the instantaneous variance adjusts, and the time to expiration of the option.

The option price and hedge ratios in Heston's model are functions not only of the parameters that appear in the Black-Scholes formula but also of κ, θ, ρ, γ, and an additional parameter, λ. The parameter λ is a constant such that $\lambda\sigma_t^2$ measures the risk premium of volatility. The volatility risk premium is assumed to be directly proportional to the level of the volatility. The need for an assumption about the form of the volatility-risk premium is a weakness of any stochastic-volatility model because the form of the volatility-risk premium cannot be deduced from the weak assumption that all investors prefer more wealth to less wealth, as discussed in Box 3, but requires assumptions on investor tolerance toward risk that in general are difficult to justify. In this model, the form of the volatility-risk premium is crucial because it enables the derivation of the closed-form solutions for option prices and hedge ratios. However, it should not be interpreted as a weakness of this model vis-à-vis other stochastic-volatility models of option prices because others make the even stronger and less plausible assumption that the risk premium of volatility is zero.

The parameters ρ and γ are very important for determining the form of the risk-neutral distribution of the asset price at the time of the option's expiration (the terminal asset price) and hence the current option price. In other words, they may be important for accounting for the smile effects seen in the chart. For example, consider the probability that a European call option will finish in the money. Ceteris paribus, an increase in γ (an increase in the volatility of volatility) makes the tails of the risk-neutral distribution fatter: the occurrence of extreme returns is more likely.[12] The sign and magnitude of ρ determines the sign and extent of skewness in the risk-neutral distribution of the terminal asset price. Positive correlation implies that an increase in the returns of the underlying asset is associated with an increase in the volatility, tending to make the right tail of the distribution thicker and the left tail thinner than those of a normal distribution of asset returns. In other words, the frequency of extreme positive outcomes is higher and the frequency of extreme negative outcomes is lower than in the Black-Scholes model—that is, the returns have positive skewness. As a result, prices of out-of-the-money calls, which benefit from this scenario of positive skewness, are higher in the stochastic-volatility model than corresponding Black-Scholes call prices, and those of out-of-the-money puts (that lose under this scenario) are lower. On the other hand, a negative correlation implies that a decrease in the returns of the underlying asset is associated with an increase in the variance. Therefore, the left tail would be thicker and the right tail thinner than assumed for the Black-Scholes model. Since out-of-the-money puts benefit from a thicker left tail, market prices for these options would be higher than in the Black-Scholes model (underpricing by the Black-Scholes model), and, similarly, out-of-the-money calls that lose from a thicker left tail would be overpriced by the Black-Scholes model.

This last scenario is consistent with observations in the market for S&P 500 index options since the crash of 1987. As noted earlier, out-of-the-money puts have tended to command much higher prices than can be explained by the Black-Scholes model, whereas out-of-the-money calls are overpriced by the Black-Scholes model. According to the stochastic-volatility model, the underpricing of the out-of-the-money puts and overpricing of out-of-the-money calls by the Black-Scholes model—the volatility skew—could be the result of a negative correlation between index returns and a random volatility process.

A likely cause of financial market volatility is the arrival of information and its subsequent incorporation into asset prices through trading.

The empirical work done on Heston's model includes that by Knoch (1992), Saikat Nandi (1996), and Bates (1995). In order to take into account the possibility of sudden large price movements, such as the crash of 1987, Bates generalizes Heston's model by allowing for jumps in asset prices. While Knoch and Bates study the pricing issues of this model for options on foreign currencies, Nandi examines both pricing and hedging issues using the S&P 500 index options. All of these studies find that Heston's model is able to generate prices that are in closer agreement with market option prices than those of the Black-Scholes model. However, it is not the case that this model is able to explain all biases of the Black-Scholes model. While it is true that the remaining pricing biases are of smaller magnitude than those of the Black-Scholes model, Nandi finds that there are still substantial biases for out-of-the-money puts and calls in the S&P 500 index options market. In particular, the model underprices out-of-the-money puts and overprices out-of-the-money calls. It is possible that the square-root volatility process and therefore the model itself are misspecified. This misspecification would be unfortunate because the particular form of the volatility process is what makes this stochastic-volatility model tractable.

If the Black-Scholes assumption of constant volatility were true, a hedge portfolio (hedged against the risk from the asset price) would simply earn the risk-free rate of return. Such a portfolio would typically consist of a position in the underlying asset and an option. The position would be altered through time by trading, based on the formulas for hedge ratios determined by the Black-Scholes model (see Cox and Rubinstein 1985) or other option pricing models. When volatility is stochastic, as it probably is in the real world, hedging using the Black-Scholes model does not result in risk-free positions. A stochastic-volatility model may do a better job of hedging against price and volatility risks. Nandi (1996) finds that for S&P 500 index options the returns of a hedge portfolio constructed using Heston's stochastic-volatility model come closer to matching a risk-free return through time better than hedge portfolio returns obtained using the Black-Scholes model.

Volatility Jumps. All the time-varying volatility models that have been discussed so far assume that the volatility of the underlying asset as well as its price evolves "smoothly," though randomly, through time: there are no jumps in the volatility process. However, a likely cause of financial market volatility is the arrival of information and its subsequent incorporation into asset prices through trading. To the extent that information—"news"—arrives in discrete lumps, it is possible that volatility shifts between episodes of low and high volatility. For example, uncertainty about an impending news release (concerning some macroeconomic variable, like an anticipated change in the fed funds rate by the Federal Open Market Committee) may cause the volatility of an asset price to rise. However, after a few rounds of trading, with the information having been incorporated into asset prices, volatility may revert back to its previous level.

To account for jumps like those in the example, Vasantlilak Naik (1993) develops a pricing model for European options in which volatility switches between low and high levels. Each level or "regime" is expected to last for a certain period of time that is not known a priori. One tractable version of his model assumes that the risk from the volatility jumps is not priced by market participants. The model takes the same parameters that enter the Black-Scholes formula as well as additional parameters such as the probabilities of jumps from one regime to another regime, given that volatility is currently in a particular regime. Naik finds that short-maturity options are much more sensitive to volatility shifts than long-maturity options. The reason is that, over a long period of time, expected upward and downward jumps in volatility are canceled by each other, resulting in a volatility that is close to the normal level.

This model has not been empirically tested and therefore cannot yet be evaluated against other stochastic-volatility models. In general, jump models can be difficult to verify empirically because jumps occur infrequently. The parameters of such models may be imprecisely estimated using relatively small historical data series on option prices or underlying asset prices.

Conclusion

Since volatility of the underlying asset price is a critical factor affecting option prices, the modeling of volatility and its dynamics is of vital interest to traders, investors, and risk managers. This modeling is a difficult task because the path of volatility during the life of an option is highly unpredictable. Clearly, the Black-Scholes assumption of constant volatility can be improved upon by incorporating time variation in volatility. While deterministic-volatility models can capture the dynamics of the volatility reasonably well, many of these option pricing models, such as ARCH models, are computationally expensive, especially for American options. Deterministic-volatility option pricing models have the advantage that most parameters can be estimated directly from the observable time series of returns data. However, superior hedging performance of such models relative to that of the Black-Scholes model has not been demonstrated. On the other hand, there is evidence that some stochastic-volatility option pricing models provide better hedges than Black-Scholes, although for stochastic-volatility option pricing models and volatility-jump models, parameter estimation is typically demanding and problematic.

The development of tractable stochastic-volatility models as well as more efficient methods of model parameter estimation are currently an area of intensive research. For both academic researchers and market practitioners, no consensus exists regarding the best specification of volatility for option pricing. Although a number of alternative approaches can account, at least partially, for the pricing deficiencies of the Black-Scholes model, none dominates as a clearly superior approach for pricing options.

Notes

1. Instead of taking a wide range of values as in the real world, a binomial tree restricts stock price movements at any moment in time to be either up with one probability or down with another (see Cox and Rubinstein 1985).
2. Although there is one source of uncertainty that drives both the asset returns and the volatility in a GARCH model, which is a special case of ARCH, the asset returns are distributed continuously—that is, one out of an infinite number of possible uncertain returns will be realized over the next period. Therefore, with discrete trading (as in a GARCH model), it is not possible to replicate all possible uncertain returns outcomes (see Duffie and Huang 1985) by trading in the option and a risk-free asset (or, equivalently, a unique risk-free portfolio cannot be created by trading in the underlying asset and an option). Hence, a risk premium associated with the returns of the underlying asset is required in a GARCH model.
3. The NGARCH of Engle and Ng (1993) is one such variation.
4. GARCH can capture the volatility smile. In a GARCH model, such as Duan's (1995), the price of an option, besides being a function of the variables that appear in the Black-Scholes formula, is also a function of variables that describe the time variation in volatility as well as a variable that accounts for the risk premium of the asset returns, that is, the excess return over a risk-free asset. Since the risk premium summarizes investor preferences, the GARCH option pricing model is not preference-free—a key attribute of the Black-Scholes model. Duan shows that under the risk-neutral distribution, the value of the GARCH variance at a point in time is negatively correlated with past asset returns if the risk premium of the asset is greater than zero. Such a negative correlation can give rise to negative skewness in the risk-neutral distribution, which seems to be a feature of the empirical data, as discussed in Bates (1995). GARCH models can therefore potentially generate option prices that are consistent with the observed volatility skew.
5. The Hobson-Rogers model is also preference-free. This model, unlike GARCH, is set in continuous time. There being a single source of uncertainty and continuous trading, all possible uncertain returns outcomes of the underlying risky asset over the next period can be replicated by trading in an option and a risk-free asset (Duffie and Huang 1985), and there is no need for any risk premium of returns.

6. The Hobson-Rogers equation actually is written with an integral rather than a summation.
7. Back (1993) shows how stochastic volatility might be introduced endogenously in asset markets due to asymmetric information about the future price of an underlying asset on which an option is traded.
8. In an ARCH option pricing model the risk premium that enters is the risk premium of asset returns and not the risk premium of volatility.
9. For American options, a closed-form solution in a stochastic-volatility model has not yet been derived.
10. Their model requires the numerical evaluation of a two-dimensional integral (that is computationally easier) rather than the solution of a second-order partial differential equation. However, the volatility process is allowed to become negative, an undesirable feature.
11. Heston's (1993) paper gives the closed-form solution for prices of call options. The price of a put option can be easily obtained using the put-call parity for European options.
12. A tail of a probability distribution is the area under the distribution that assigns probabilities to extreme outcomes. For example, in the typical bell-shaped normal distribution, there are two tails, the right tail and the left tail, that slowly taper off.

References

Abken, Peter A. "Over-the-Counter Financial Derivatives: Risky Business?" Federal Reserve Bank of Atlanta *Economic Review* 79 (March/April 1994): 1-22.

——. "Using Eurodollar Futures Options: Gauging the Market's View of Interest Rate Movements." Federal Reserve Bank of Atlanta *Economic Review* 80 (March/April 1995): 10-30.

Abken, Peter A., and Hugh Cohen. "Generalized Method of Moments Estimation of Heath-Jarrow-Morton Models of Interest-Rate Contingent Claims." Federal Reserve Bank of Atlanta Working Paper 94-8, August 1994.

Abken, Peter A., Dilip B. Madan, and Sailesh Ramamurtie. "Estimation of Risk-Neutral and Statistical Densities by Hermite Polynomial Approximation: With an Application to Eurodollar Futures Options." Federal Reserve Bank of Atlanta Working Paper 96-5, June 1996.

Aït-Sahalia, Yacine, and Andrew W. Lo. "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices." National Bureau of Economic Research Working Paper 5351, 1995.

Amin, Kaushik, and Andrew Morton. "Implied Volatility Functions in Arbitrage-Free Term Structure Models." *Journal of Financial Economics* 35 (1994): 141-80.

Amin, Kaushik, and Victor Ng. "ARCH Processes and Option Valuation." University of Michigan Working Paper, 1993.

Back, Kerry E. "Asymmetric Information and Options." *Review of Financial Studies* 6 (1993): 435-72.

Barone-Adesi, Giovanni, and Robert Whaley. "Efficient Analytic Approximation of American Option Values." *Journal of Finance* 42 (1987): 301-20.

Bates, David. "The Skewness Premium: Option Pricing under Asymmeric Processes." University of Pennsylvania Working Paper, 1994.

——. "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutschemark Options." *Review of Financial Studies* 9 (1995): 69-107.

Black, Fischer, and Myron S. Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81 (May/June 1973): 637-54.

Bollerslev, Tim. "Generalized Autoregressive Conditional Heteroscedasticity." *Journal of Econometrics* 31 (1986): 307-27.

Bollerslev, Tim, Ray Y. Chou, and Kenneth F. Kroner. "ARCH Modeling in finance: A Review of the Theory and Empirical Evidence." *Journal of Econometrics* 5 (1992): 5-59.

Cox, John C., and Steve Ross. "The Valuation of Options for Alternative Stochastic Processes." *Journal of Financial Economics* 3 (1976): 145-66.

Cox, John C., and Mark Rubinstein. *Options Markets*. Englewood Cliffs, N.J.: Prentice Hall, 1985.

Derman, Emanuel, and Iraz Kani. "Riding on the Smile." *Risk* 7 (February 1994): 32-39.

Duan, Jin C. "The GARCH Option Pricing Model." *Mathematical Finance* 5 (1995): 13-32.

Duffie, Darrell, and Chi-Fu Huang. "Implementing Arrow-Debreu Equilibria by Continuous Trading of Few Long-Lived Securities." *Econometrica* 53, no. 6 (1985): 1337-56.

Dumas, Bernard, Jeffrey Fleming, and Robert Whaley. "Implied Volatility Functions: Empirical Tests." Rice University Working Paper, 1996.

Dupire, Bruno. "Pricing with a Smile." *Risk* 7 (January 1994): 18-20.

Engle, Robert F. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica* 50 (1982): 987-1008.

Engle, Robert, Alex Kane, and Jaesun Noh. "Forecasting Volatility and Option Prices of the S&P 500 Index." *Journal of Derivatives* (1994): 17-30.

Engle, Robert F., and Chowdhury Mustafa. "Implied ARCH Models from Option Prices." *Journal of Econometrics* 52 (1992): 289-311.

Engle, Robert, and Victor Ng. "Measuring and Testing the Impact of News on Volatility." *Journal of Finance* 48 (1993): 1749-78.

Fama, Eugene F. "The Behavior of Stock Market Prices." *Journal of Business* 38 (1965): 34-105.

Harrison, J. Michael, and David Kreps. "Martingales and Arbitrage in Multiperiod Securities Markets." *Journal of Economic Theory* 20 (1979): 381-408.

Harrison, J. Michael, and Stanley Pliska. "Martingales and Stochastic Integrals in the Theory of Continuous Trading." *Stochastic Processes and Their Applications* 11 (1981): 215-60.

Heston, Steven. "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options." *Review of Financial Studies* 6 (1993): 327-43.

Hobson, David G., and L.C.G. Rogers. "Complete Models with Stochastic Volatility." University of Bath, School of Mathematical Sciences. Unpublished paper, 1996.

Hull, John C., and Alan White. "The Pricing of Options on Assets with Stochastic Volatilities." *Journal of Finance* 42 (1987): 281-300.

Knoch, Hans J. "The Pricing of Foreign Currency Options with Stochastic Volatility." Ph.D. dissertation, Yale School of Organization and Management, 1992.

Lamoureux, Christopher, and William Lastarapes. "Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities." *Review of Financial Studies* 6 (1993): 293-326.

Litterman, Robert, Jose Scheinkman, and Laurence Weiss. "Volatility and the Yield Curve." *Journal of Fixed Income Research* (June 1991): 49-53.

Mandelbrot, Benoit. "The Variation of Certain Speculative Prices." *Journal of Business* 36 (1963): 394-419.

Melino, Angelo, and Stuart Turnbull. "The Pricing of Foreign Currency Options with Stochastic Volatility." *Journal of Econometrics* 45 (1992): 239-65.

Merton, Robert C. "Option Pricing with Discontinuous Returns." *Journal of Financial Economics* 3 (1976): 125-44.

Naik, Vasantlilak. "Option Valuation and Hedging Strategies with Jumps in the Volatility of Asset Returns." *Journal of Finance* 48 (1993): 1969-84.

Nandi, Saikat. "Pricing and Hedging Index Options under Stochastic Volatility: An Empirical Examination." Federal Reserve Bank of Atlanta Working Paper 96-9, August 1996.

Nelson, Daniel B. "Conditional Heteroscedasticity in Asset Returns: A New Approach." *Econometrica* 59 (1991): 347-70.

Rubinstein, Mark. "Nonparametric Tests of Alternative Option Pricing Models Using All Reported Trades and Quotes on the 30 Most Active Option Classes from August 23, 1976, through August 31, 1978." *Journal of Finance* 40 (1985): 455-80.

——. "Implied Binomial Trees." *Journal of Finance* 49 (1994): 771-818.

Scott, Louis O. "Option Pricing When the Variance Changes Randomly: Theory, Estimation, and Application." *Journal of Financial and Quantitative Analysis* 22 (1987): 419-38.

Shimko, David. "Bounds on Probability." *Risk* 6 (April 1993): 33-37.

Stein, Elias M., and Jeremy C. Stein. "Stock Price Distributions with Stochastic Volatility: An Analytic Approach." *Review of Financial Studies* 4 (1991): 727-52.

Wiggins, James B. "Option Values under Stochastic Volatilities." *Journal of Financial Economics* 19 (1987): 351-72.