The Economics of Financial Privacy: To Opt Out or Opt In?

Jeffrey M. Lacker

A consumer's financial transactions give rise to a wealth of very personal data. Every credit card purchase, every ATM withdrawal, every loan payment, every paycheck deposit leaves an electronic trace at a person's bank. Advances in information technology now allow firms to collate information from disparate sources and compile comprehensive profiles of individual behavior. The resulting databases can allow businesses to target very specific consumer categories—high-income, gun-owning dog lovers, for example—in ways that were never before possible.

When should a bank be able to share information about you with other businesses? Some consumer advocates want to protect consumers' financial privacy by restricting such information sharing. New technologies, they say, have encouraged increased intrusions on consumer privacy, leading to more junk mail, more telemarketing calls, and a heightened risk of identity theft. They argue for tough "opt-in" laws that would require financial institutions to obtain a consumer's explicit consent before sharing personal information about them.

Banks and other financial service providers point out that information sharing provides benefits to consumers by allowing for more targeted marketing and services. The new technologies make it easier for businesses to find consumers that would be interested in buying their specialized products and services—hunting-dog training supplies, for example. Such marketing directly benefits consumers when it results in a voluntary purchase. In addition,

■ The author is Senior Vice President and Director of Research. This article first appeared in the Bank's *2001 Annual Report.* It benefited from the comments of the author's colleagues in the Bank's Research Department, especially John Weinberg, Marvin Goodfriend, Laura Fortunato, Ned Prescott, Aaron Steelman, and John Walter, and from the assistance of Elise Couper. The views expressed are the author's and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

greater information sharing can reduce wasteful marketing to consumers that are likely to be uninterested. With these benefits in mind, financial service providers argue for "opt-out" laws that merely require them to give consumers the right to request that their information not be shared.

After vigorous debate, Congress adopted an opt-out requirement for banks and other financial institutions as part of the Gramm-Leach-Bliley Act of 1999 (GLBA), legislation that was designed to encourage financial modernization. Any financial institution that intends to share nonpublic customer information with third parties (companies not related by ownership ties) must give customers an opportunity to deny them permission to do so, or opt out. In addition, financial institutions are required to provide customers with an annual statement of their privacy policy. Consumers received a blizzard of notices in the mail when those provisions were fully implemented in the summer of 2001.[1]

The controversy did not end with the passage of the GLBA. The Act allows individual states to adopt privacy provisions that are stricter than the federal standard if they so desire. California's legislature recently considered an opt-in law that would have required financial institutions to obtain customer permission before sharing information with third parties. Moreover, banks would have been required to give consumers the right to opt out of information sharing with affiliated companies (companies related by ownership ties).

This essay examines the opt-out/opt-in debate from the perspective of the economics of financial privacy. The premise is that a financial institution's privacy policy is a *characteristic* of the products and services the institution offers. We can therefore apply the well-understood principles governing how markets work when there are important differences in product characteristics. The result is surprising for both sides of the issue: it doesn't seem to matter whether opt-out or opt-in is adopted as the standard. Either way, competitive forces should bring about an economically efficient amount of information sharing. In fact, even in the absence of opt-out or opt-in laws, the amount of information sharing should be economically appropriate. Opt-out/opt-in laws will be irrelevant as long as financial institutions are not prevented from offering customers a range of desirable privacy options.

The broad and multifaceted issues that surround privacy go well beyond the opt-out/opt-in debate. Although this essay is narrowly focused on the latter, the general principles outlined here have a much wider application. At a fundamental level, opt-out versus opt-in is really a question about the proper

[1] The deadline for compliance was July 1, 2001. For more information on the financial privacy provisions of the GLBA, see the Federal Trade Commission's Web site (Federal Trade Commission 2002). The privacy provisions of the GLBA apply to any institution engaged in activities that have been deemed "financial in nature or incidental to such financial activities" under the Bank Holding Company Act. This means that whenever the Fed and the Treasury determine that an activity is financial in nature and therefore a permissible activity for a financial holding company, the entire financial industry is brought under the privacy provisions of the GLBA.

allocation of "rights" in contractual relationships—a customer's right to privacy versus the right of a financial institution to share its information. The answer economics provides is that whether rights are allocated in accord with opt-out or opt-in is irrelevant, as long as consumers and financial institutions are free to agree to an alternative arrangement if it suits them. Most financial privacy questions concern the specification of rights of various parties in contractual relationships. The irrelevance result of this essay thus should carry over to other related settings; laws and regulations providing more (or less) "privacy rights" should generally have little effect on consumers' financial privacy.[2]

1. PRIVACY IN THE FINANCIAL MARKETPLACE

Financial privacy can be thought of as a bundle of characteristics associated with a particular financial service. A bank that does not share nonpublic customer information with third parties is providing its customers a service with different characteristics from a bank that does share such information. How do markets work when products or services differ in their characteristics?

In well-functioning competitive markets, consumers selecting among products with different bundles of characteristics are willing to pay more for products with characteristics they value. Some characteristics make a product more costly to provide. Producers are willing to supply products with more costly characteristics only if they are compensated for the additional cost. One would expect to see products with characteristics for which a customer's willingness to pay exceeds the incremental production cost. For example, some people are willing to pay more for a car with a built-in CD player, but CD players are costly. It is logical then that consumers whose willingness to pay exceeds the cost of the CD player would own cars with CD players.

Well-functioning markets generally provide goods and services that are appropriate when judged against the benchmark of economic efficiency. With regard to product characteristics, economic efficiency means that a given product characteristic is supplied if and only if the value of that characteristic to consumers exceeds its cost to society. When markets function smoothly, the incentives of producers and consumers are aligned with economic efficiency. Suppliers find it profitable to provide products with the appropriate characteristics, since consumers are willing to pay at least the additional cost. Characteristics for which consumers' valuations fall short of the cost of production cannot be profitably supplied.

Financial privacy is a service characteristic that some consumers prefer. Many consumers harbor deep concerns about privacy in general and financial privacy in particular. According to one recent poll, 56 percent of consumers say

[2] For other economic analyses of financial privacy, see Kahn, McAndrews, and Roberds (2000) and Bauer (forthcoming).

they are "very concerned" about potential loss of privacy.[3] Overall, consumers seem to have three main fears.[4] They fear being robbed or cheated by criminals that obtain personal information. They fear embarrassing revelations due to the disclosure of sensitive information. And they dislike intrusive marketing in the form of telephone calls or junk mail. When financial institutions share customer information with outside companies, it can erode customer privacy on all three counts.

Providing greater financial privacy can be costly for a financial service provider because it means foregoing the potential economic value of information sharing. Marketers can make better decisions the more information they have about prospective customers and are therefore willing to pay banks to get it. Better information helps marketers find customers who genuinely may be interested in buying their products and saves them the expense of soliciting consumers who are not. These benefits provide genuine economic value by increasing the probability of a successful buyer-seller match and decreasing the probability of wasting marketing efforts on those who would not be interested.

Consumers that place a high value on financial privacy ought to be willing to pay for high-privacy financial services. If consumers prefer that their bank not share nonpublic information about them with unaffiliated companies, they should be willing to pay for this service characteristic implicitly through lower deposit interest rates, higher loan interest rates, or higher account-related fees. More directly, banks could offer direct inducements—a bonus payment, coupon, or sweepstakes entry, for example—to customers that agree to information sharing. Many nonfinancial firms offer such enticements to customers that return "product registration cards" filled out with their name, address, and other information. Consumers that value financial privacy would pay by foregoing their bank's offer. Similarly, many grocery stores offer cards to customers that qualify them for discounts when they present the cards at checkout stations. In exchange, stores gather data on customer purchases.

Along the same lines, if sharing nonpublic customer information with third parties is economically beneficial, financial institutions should be willing to compensate their customers who allow them to do so.[5] The outside firms with which the information is shared should be willing to pay an amount up to the information's value to them. The financial institution should then be willing to pass this along to their customers in the form of higher interest rates on savings, lower interest rates on loans, or lower fees. More directly, they should be willing to simply pay those customers who agree to share an amount up to the incremental value of the information.

Ideally, the economic benefits of financial privacy should be balanced against the economic costs. When the economic value of sharing nonpublic

[3] National Consumers League (2000).
[4] Research by Alan Westin, as cited in Paul (2001).
[5] See Kovacevich (2000).

customer information with third parties falls short of the value consumers place on preventing that information sharing, economic efficiency would dictate that no information sharing takes place. Similarly, when the economic value of sharing nonpublic customer information with third parties exceeds the value consumers place on preventing it, economic efficiency would dictate that information sharing should take place. If the market for financial privacy is well functioning, then we should see an economically efficient amount of financial privacy.

2. DOES THE MARKET FOR FINANCIAL PRIVACY WORK WELL?

Is there anything different about financial privacy? Are the markets for financial privacy poorly functioning in the sense that they deliver outcomes that are not economically efficient? There does not appear to be any plausible reason to think so.

For markets to misfunction in this sense, one of two conditions must exist: either a divergence between the value of a product characteristic to consumers and their willingness to pay it, or a divergence between the cost to suppliers of providing that characteristic and the overall cost to society. Divergences could be caused by externalities, monopoly power, or verification problems.

An externality occurs when an action by one group affects the well-being of others that do not transact with that group. For example, burning leaves in my front yard raises the risk of fire for my suburban neighbor.[6] Externalities are often invoked to explain a broad range of government laws and regulations—prohibiting suburban leaf burning, for example.

Is there an externality in the market for financial privacy? No, it doesn't appear so. Sharing nonpublic customer information about a consumer affects that consumer's privacy but not the privacy of other consumers. The sharing institution is a counterparty of the affected customer, and either can withdraw from the relationship. The two of them have ample opportunity to take information sharing into account when setting the terms of their relationship. Thus no parties are affected by the information sharing except those who are participants in the transaction.

"Public goods" are a type of externality that can result in inefficiency and are defined by two properties. They are nonrivalrous, meaning that one person's use does not detract from the ability of another to use it. And they are nonexcludable, meaning that one cannot prevent people from using it. A lighthouse is a classic example of a public good: one ship's use does not prevent

[6] One could argue that the two parties could negotiate an efficient solution to this problem; my neighbor can simply pay me not to burn leaves, or can sue me if the fire spreads. For additional explanation see the section on the Coase Theorem.

another ship's use, and you cannot prevent a ship from using it.[7] Information is nonrivalrous because one person's use does not prevent another from using the same information. But information is excludable because you can prevent people from obtaining it. Therefore financial information is not a public good.

Monopoly power is another possible cause of market misfunction. When a firm is sheltered from competitive pressures it can raise prices and restrain supply. Similarly, a protected monopolist may find it profitable to supply too little of a desired product characteristic when customers are prevented from seeking preferred characteristics from other suppliers. This problem may have been relevant to the banking industry decades ago when competition was severely limited by regulatory restrictions on pricing, entry, and geographic expansion, but these restrictions have been largely dismantled. As a consequence, the market for financial services is now widely judged to be relatively competitive. Thus it seems unlikely that banks or other financial institutions are manipulating privacy policies because of significant monopoly power.[8]

A third potential cause of market misfunction stems from the difficulty of verifying whether a financial institution is living up to its stated privacy policy. A customer that receives junk mail or telemarketing calls may have a hard time discerning where the marketer obtained the information. The spelling of a name or address can be altered slightly in order to trace information sharing, but this technique is obviously limited. In cases of identity theft it is often impossible to determine exactly how the identity was stolen after the fact.

Do verification problems interfere with the efficiency of the market for financial privacy? Not necessarily. Note that there are a number of mechanisms to help ensure that an institution lives up to its privacy commitments, despite the difficulty of observing whether or not it has done so. First, an institution that fails to comply with its stated financial privacy policy may be liable for "unfair and deceptive trade practices." If caught, the institution would be subject to civil litigation as well as regulatory action by the Federal Trade Commission. The potential legal costs can deter noncompliance, even if the probability of detection is small. There is nothing particularly unique about financial privacy in this regard. Consumers often rely on hard-to-verify commitments by the firms they patronize—a commitment to product quality, for example.

[7] Coase (1974) pointed out, however, that coastal lighthouses are often funded from fees charged to ships using nearby ports, so even the services of lighthouses are at times excludable. A lighthouse is therefore only a public good when ships cannot be excluded from using its services if they do not pay—for example, in settings where most ships are on long-distance voyages.

[8] If financial institutions were exercising market power and this resulted in inefficient financial product characteristics, a more appropriate remedy would be for regulators to ensure effective competition rather than regulate service characteristics. Moreover, it would appear inconsistent to regulate service characteristics on the grounds of impediments to competition while not regulating service prices.

Second, institutions that wish to attract customers for whom privacy is important will want to convince those customers of their organization's commitment to its privacy policy. Such institutions will have an incentive to cultivate and safeguard their reputation as a high-privacy entity. At least one prominent bank has advertised a "no telemarketing" promise, indicating that banks are capable of actively competing on the basis of their privacy policies.[9] Third parties can evaluate a financial institution's compliance, just as *Consumer Reports* independently assesses the quality of consumer products. The potential for embarrassing media publicity also motivates an institution to live up to its commitments. Standard industry practice is for a firm that rents its mailing list to approve every mailing or telemarketing script that is used. Evidently firms believe that at least some consumers could trace marketing contacts to them, with possibly detrimental effects on their customer relationships.

While reputational considerations and laws on trade practices can go partway toward ensuring that a firm is faithful to its stated privacy policy, some would argue that these mechanisms are inherently limited and imperfect. Enforcement is often costly and compliance is rarely 100 percent. Do these imperfections warrant legislative restrictions aimed specifically at information sharing? No. Any entity attempting to verify and enforce a financial firm's privacy commitments will confront the same imperfections. A governmental effort to enforce a ban on information sharing, for example, will face the same verification difficulties—costly enforcement and incomplete compliance—as would any private parties. So a government ban on information sharing would have no advantage; in fact, it would have the disadvantage of possibly preventing economically useful information sharing.

The market for financial privacy therefore appears to work fairly well. This means that we should expect economically efficient outcomes: information will be shared if and only if the economic benefits of information sharing exceed the value consumers place on preventing information sharing.

3. OPT-OUT VERSUS OPT-IN

Provided the market for financial privacy works fairly well, it should not make much difference whether we adopt an opt-out law or an opt-in law. Either way, an economically efficient level of information sharing will result. Why is this so?

Under an opt-out law, banks that value information sharing will be willing to provide inducements to get high-privacy customers not to opt out because

9 The phrase appeared in television advertising for Capital One during November 2001. As of this writing (January 17, 2002), the company's home page prominently features the following description of their "New No-Hassle Card": "9.9% Fixed APR on Everything, No Telemarketing, No Annual Fee."

information sharing can lower the cost of providing banking services. Similarly, automakers are willing to discount the price of cars without CD players, since these cars are less costly to build. Banks will be willing to pay an amount up to the incremental value of sharing the customer's nonpublic information. If that falls short of the value the customer implicitly places on privacy, then the customer will decline the inducement and opt out. In that case, the economic value of the information sharing is less than the cost to the customer of yielding this bit of privacy, and information sharing is not economically efficient. Alternatively, the customer may feel that the value of the inducement exceeds the value of preventing information sharing, in which case the inducement is accepted and the customer does not opt out. Here, the economic value of the information sharing exceeds the cost to the customer of yielding this bit of privacy, and information sharing is economically efficient.

Under an opt-in law, the reasoning and the result are exactly the same. Banks will be willing to provide the same inducement to get a customer to opt in as they would have provided to get a customer to refrain from opting out—up to the economic value of the information sharing. If that amount exceeds the value that the customer places on preventing information sharing, then information sharing will take place and is economically efficient. Otherwise the customer will refuse the enticement; in this case information sharing is not economically efficient and will not take place.

In fact, the same reasoning applies in the absence of opt-out or opt-in laws. If the law is silent on whether banks need to seek permission to share nonpublic information with third parties, banks nonetheless could decide to do so on their own. If some customers truly care about information sharing with third parties, they will seek out banks that give them the option of preventing it. If information sharing is economically useful, banks will find it more costly to serve customers that insist on preventing it. Competition will force banks to pass along the increased cost to high-privacy customers. Ultimately, an economically appropriate amount of information sharing will take place, with or without opt-out or opt-in laws.

The difference between opt-out and opt-in standards is like the difference between treating CD players in cars as standard equipment or as an add-on option. If CD players are an option, one would expect the price of the option to reflect the incremental cost. If instead CD players are standard equipment, the discount for cars without CD players should reflect the incremental cost. It should not make a difference whether car buyers have to ask to get a CD player in their car or ask not to have one. Either way we should see a market-clearing quantity of cars with CD players.

The debate between proponents of opt-out and opt-in seems predicated on the view that the choice would affect how many consumers would prevent information sharing. The hypothesis seems to be that fewer consumers would opt out under an opt-out standard than would fail to opt in under an opt-in standard. This could well be the case, but it would be evidence that many consumers are relatively indifferent about information sharing by their financial institution; they would not bother to opt out, nor would they bother to opt in. If this is true, then little is at stake for these consumers. Those who would neither opt out nor opt in evidently place little value on preventing their financial institution from sharing nonpublic information about them. The economic efficiency implications of the choice between opt-out and opt-in would therefore be negligible for them as well, even if participation rates differed significantly.

4. AN ALTERNATIVE LINE OF REASONING: THE COASE THEOREM

The knowledgeable reader may have noticed that the logic of this essay is closely related to the insights that Ronald H. Coase presented in his celebrated paper "The Problem of Social Cost."[10] (This paper was cited by the Royal Swedish Academy of Sciences in awarding him the 1991 Nobel Prize in Economics.) Coase wrestled with the issue of externalities, the same issue as in my leaf-burning example. Before Coase's paper economists generally believed that, absent government intervention, externalities would result in inefficient outcomes because one party (I, for example) would ignore the cost (increased fire hazard) that his action (leaf burning) imposed on another party (my neighbor). The contribution of Coase was to notice that the two parties could negotiate an efficient solution to the externality problem as long as the relevant rights were clearly assigned. For example, if I am entitled to burn leaves, my neighbor could offer to pay me not to, or could offer to help me dispose of them by some other method. Alternatively, if I am required to obtain my neighbor's permission to burn leaves, I could offer to pay my neighbor. If the value to me of burning leaves is less than the value to my neighbor of my not burning leaves, then my neighbor will pay me not to do so in the first case. In the second case, I will be unwilling to offer my neighbor enough money to get permission to burn leaves. Either way we get an efficient outcome; I don't burn leaves. The general proposition is that (under certain conditions) any well-defined allocation of property rights leads to efficient outcomes. This result is often called the Coase Theorem.

The application to financial privacy should be clear. Opt-out and opt-in are just different allocations of property rights. Opt-out means financial institutions have the right to share information; customers can ask them to stop. Opt-in means customers have the right to no-information-sharing; financial

[10] Coase (1960).

institutions can ask them for permission to share. Either way, according to Coase, the prediction is an efficient amount of information sharing.

The Coase Theorem has its limitations, however. It is said to hold only if "transaction costs" are zero; in other words, any agreement that is in the mutual interest of the parties is actually agreed upon. Transaction costs are the difficulties associated with actually reaching an agreement among the affected parties. It may be costly to communicate and coordinate among a large number of parties, for example. When transaction costs are significant, the assignment of property rights can affect efficiency. One premise of this essay, as I discuss later, is that the costs of opting out are negligible, in which case the Coase Theorem applies.[11]

The logic of this essay, however, differs subtly from Coase's analysis. Coase envisioned bargaining between affected parties. As a result, the assignment of property rights could alter the distribution of net benefits, even if that assignment had no effect on efficiency. For example, if I have the right to burn leaves, I get paid not to burn them; yet if I need permission, I earn nothing when I don't burn them. I am better off in the first case, while my neighbor is better off in the second case. The assignment of rights thus alters the relative well-being of my neighbor and me, even though either assignment leads to efficient leaf-burning decisions. In competitive markets, in contrast, the assignment of contractual rights generally does not affect people's well-being. The choice between opt-out and opt-in determines which rights are, by default, bundled together with financial services. Under either regime, competition and free entry implies that both high-privacy and low-privacy financial services will be available at prices reflecting their true cost. In competitive markets, the choice of regime should have no effect on the net cost of financial services with particular characteristics, just as a law mandating that CD players be sold separately should have no effect on the total price of cars with CD players. The efficiency implication of Coase's famous theorem carries over to competitive markets, however, and buttresses the case made here: market mechanisms should work well at providing an efficient level of financial privacy.

5. OPT-OUT IN PRACTICE: FEW CONSUMERS DO

During the first half of 2001, many banks began mailing out the privacy notices required by the GLBA. Those that share nonpublic customer information with unaffiliated companies are required to give their customers the opportunity

[11] The costs are negligible in part because of the regulations that require financial institutions to provide customers with a "reasonable means" of opting out. In a sense, then, this part of the allocation of property rights has efficiency implications consistent with the Coase Theorem. The reasonable-means provision appears to be an efficient choice since it minimizes the "transaction costs" of opting out. Friedman (2000) applies Coase's approach to a broad array of privacy issues in which transaction costs are nonnegligible.

to opt out of third-party information sharing. Although there is only limited evidence so far, press reports suggest that the response rate is rather low. According to the trade publication *American Banker,* industry estimates of the number of consumers who have opted out "hover around 5 percent."[12] One survey of savings banks showed that more than half were experiencing an opt-out rate of one percent or less.[13]

Opting out does not appear to be very hard. The financial privacy regulations require that financial institutions give customers a "reasonable means" of exercising their right to opt out. The regulations even offer examples of acceptable and unacceptable methods. Providing a toll-free number to call or supplying a mail-in card for a check-box response are deemed reasonable means. Requiring a customer to write his or her own letter is not deemed reasonable.

Despite these requirements, critics claim that opting out is difficult because privacy notices are complex, confusing, and hard to read.[14] Food labels are often cited, in contrast, as a simple, well-understood notice system. Some financial institutions, however, are actively working toward simpler and clearer privacy notices.[15] Apparently, they view that it is in their business interest to make their notices as agreeable to their customers as possible. Many institutions sent privacy notices for the first time in 2001, and some experimentation and learning seem to be taking place. Perhaps opt-out rates will rise as GLBA privacy notices are refined and consumers learn about what they contain.

Nevertheless, the fact that so few bank customers are currently taking the relatively easy step of opting out seems to indicate that most consumers now place a negligible value on preventing financial institutions from sharing nonpublic information about them with third parties. A small fraction of consumers feel strongly enough to take advantage of the opt-out option. This group appears to place a significant value on guarding their financial privacy. But for a broad majority of Americans, the value they place on financial privacy does not exceed the inconvenience of exercising their right to opt out.[16]

This pattern—about 5 percent of people willing to take action to protect their privacy—is consistent with other evidence on consumers' privacy preferences. The Direct Marketing Association, a marketing industry trade group,

---

12 Lee (2001).

13 America's Community Bankers (2001).

14 See transcripts and supporting documentation from the workshop on effective privacy notices hosted by the Federal Trade Commission and the federal financial regulatory agencies (Federal Trade Commission 2001).

15 See the presentations by Marty Abrams, John Dugan, Patricia Faley, and David M. Klaus at the privacy notices workshop along with the public comments submitted by Walter Kitchenman, Vance Gudmundsen, and Steve Bartlett in connection with the event (Federal Trade Commission 2001).

16 One could argue that consumers are just lazy, but this reasoning leads to the same conclusion; the value they place on financial privacy is not enough to motivate them to opt out.

offers consumers the ability to opt out of telephone or mail marketing by their members. The 4.2 million participants in their telephone opt-out program represent about 4.2 percent of U.S. households with telephone service. The 4.0 million participants in their mail opt-out program represent about 3.8 percent of total U.S. households.[17]

A very low opt-out rate is also consistent with other choices consumers make with regard to privacy. Few consumers disable cookies when browsing the Internet. (Cookies are small files that a Web site places on a user's computer to enable tracking the user on subsequent visits.) Few consumers read privacy notices. Many consumers readily provide their credit card number over the phone or to a waiter.[18] The picture that emerges, then, is that a few consumers place significant value on preventing information sharing by their financial institutions, but the broad majority of consumers are relatively indifferent.

6. OPT-OUT IN PRACTICE: FEW BANKS PAY

Financial institutions do not appear to be offering inducements to customers to get them to refrain from opting out. This suggests that the economic value of sharing nonpublic customer information is relatively low. Otherwise financial institutions would find it worthwhile to compensate their customers for their cooperation. In fact, not all institutions are even engaged in information sharing that would trigger the opt-out requirement. A survey of savings banks found that fewer than one-third needed to send out opt-out notices.[19]

Banks do not lack opportunities to share customer information. There is an active market for consumers' names, addresses, and other personal information. Individual merchants rent their customer lists to marketers, often through list brokers. Credit bureaus offer selections from their databases based on age, income, occupation, family status, net worth, type of automobile, religion, and so on. According to its Web site, Equifax even offers a selection based on a person's carburetor type. American Express offers customer lists selected on the basis of purchase patterns—shoe buyers that spend more than $1000 annually, for example. Lists are available from magazines, membership organizations, book clubs, and merchants.[20]

---

[17] The three main credit bureaus also offer a program through their trade group that allows consumers to opt out of pre-approved credit offers, but the credit bureaus do not release statistics on the number of consumers opting out.

[18] According to a recent survey, 24 percent of consumers protect their privacy by disabling cookies (Harris Interactive Inc. 2001). An American Bankers Association poll found that 36 percent of consumers said they had read their bank's privacy notice (American Bankers Association 2001).

[19] America's Community Bankers (2001).

[20] For information on lists see Equifax (2001), American List Counsel (2002), and Worldata (2002).

Apparently, the market for consumer information does not provide banks with sharing opportunities that would make it worthwhile to offer material rewards for consumer cooperation. A glance at the prices for such information suggests why—prices are relatively low. Rates for lists of merchandise buyers, for example, appear to be relatively consistent, ranging from 8 cents to 13 cents per name as of early 2001. Base prices at one large credit bureau range from 1.65 to 4 cents per name per mailing, depending on volume, with add-on charges for additional selection criteria ranging from .25 cents per name for length of residence, title, or gender to 2 cents per name for net worth. Thus the value to a financial institution of sharing nonpublic customer information might not be large enough to warrant offering a significant sum to customers.

7. WHY IS FINANCIAL PRIVACY AN ISSUE NOW?

Applying economics to financial privacy leads to the conclusion that financial markets can provide an appropriate balance between consumers' desires for privacy and the economic value of information sharing. If this is true, then why do surveys show widespread consumer concern about privacy yet few consumers taking action to opt out of information sharing? And why has there been such clamor for privacy legislation in the past few years, culminating in the financial privacy provisions of the GLBA?

The dramatic changes in communications and computing technologies in recent years might help explain why so many recent surveys report consumer concern about privacy. Financial institutions have always possessed detailed information about their customers. Moreover, active markets for customer lists have been around for decades.[21] Only recently, however, has the collation and analysis of information from disparate sources become highly automated. This technological advance allows more targeted marketing efforts; a company can solicit high-income, gun-owning dog lovers, for example. The resulting improvement in marketing success rates appears to have led to an increase in the number of mail and telephone solicitations.

Before the technological developments that lowered the cost of manipulating databases, assembling such detailed consumer profiles was not economically feasible. Consumers came to view the limited nature of information sharing by financial institutions as an implicit part of their contractual relationship, relying on the *practical obscurity* of what other firms knew about

[21] I recall my father managing rentals of his company's mailing list in the 1960s. The list was kept on "addressograph plates"—metal strips embossed with names and addresses. While these strips could be linked together for automated addressing of mass mailings, any sorting or selection had to be handled manually. The list was rented out through mailing houses that handled the actual printing and distribution. All rentals had to be approved by list owners. Decoys—false names and addresses—were included in the list to provide a means of verification by the list owner.

them.[22] Since widespread information sharing was impractical then, few surveys asked how consumers felt about it. New technologies have dispersed the fog of practical obscurity that formerly surrounded many consumer transactions. The privacy concerns that appear in consumer surveys could represent ex post regret at the lack of contractual constraints on information sharing. This conflicts, however, with the evidence cited earlier indicating that most consumers do not feel strongly about information sharing. Alternatively, perhaps consumer preferences haven't changed, but consumers are merely asked about them more often today. Now that interfirm information sharing is economically viable, we see surveys on the subject.

Economists are often skeptical of survey evidence on consumer preferences, but it is not the sincerity of consumers' responses that is in doubt. Surveys rarely confront consumers with the cost consequences of their choices. When asked whether they desire greater privacy without reference to cost, they are likely to say "yes"—more of a good is generally preferred to less, after all. But when confronted with real-life choices, many consumers decide that the benefits of greater privacy are outweighed by the costs. One recent study found a dramatic disparity between consumers' stated privacy preferences and their actual online behavior.[23] Participants answered many "highly personal" questions, despite having stated that privacy was important to them. The discrepancy between widespread consumer "concern" and the willingness of many consumers to readily compromise their privacy could well reflect the gap between the artificial choices implicit in survey questions and the real choices consumers actually face.[24]

8. CONCLUSION

The economics of financial privacy is based on the notion that a financial institution's privacy policy is a characteristic associated with the products and services the institution offers. In well-functioning markets, prices reflect product characteristics; consumers are willing to pay more for characteristics they value, and producers charge more for characteristics that are more costly to supply. Consumers that value financial privacy ought to be willing to pay for privacy policies that they prefer. And if it is economically beneficial to share information with other companies, financial institutions ought to be willing to compensate their customers for permission to do so. The fact that few banks seem to be paying customers not to opt out is strong evidence that the economic value of information sharing is relatively small. And the fact that so

[22] Gramlich (1999).

[23] Spiekermann, Grossklags, and Berendt (no date available).

[24] Harper and Singleton (2001).

few consumers are opting out, despite the low cost of doing so, is evidence that few consumers place a significant value on preventing information sharing.

This line of reasoning also leads to a stark and surprising conclusion: the choice between opt-out and opt-in standards is irrelevant. Under an opt-out standard, banks could pay customers to refrain from opting out, while under an opt-in standard banks could pay customers to opt in. Either way, financial markets should deliver an efficient amount of information sharing. One puzzle remains, however: Why is financial privacy such a controversial issue if few consumers care enough about preventing information sharing to take simple steps to prevent it? Nevertheless, the economics of the issue is clear—financial privacy laws like the GLBA accomplish less than either privacy advocates or their critics presume.

REFERENCES

American Bankers Association. 2001. "ABA Survey Shows Nearly One Out of Three Consumers Read Their Banks' Privacy Notices." News Release (7 June).

American List Counsel. 2002. http://www.amlist.com [17 January].

America's Community Bankers. 2001. "ACB Privacy Compliance Survey." Manuscript (November).

Bauer, Paul. "Consumer's Financial Privacy and the Gramm-Leach-Bliley Act." Federal Reserve Bank of Cleveland *Economic Commentary* (forthcoming).

Coase, Ronald H. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3 (October): 1–44.

______________. 1974. "The Lighthouse in Economics." *Journal of Law and Economics* 17 (October): 357–76.

Equifax. 2001. TotalSourceXL. Consumer Database (Fall). http://www.equifax.com/business_solutions/information_services/documents/Fall_2001_TotalSource_XL_Rate_Card.pdf [17 January 2002].

Federal Trade Commission. 2001. Interagency public workshop entitled Get Noticed: Effective Financial Privacy Notices, 4 December, at The Ronald Reagan Building and International Trade Center, Washington, D.C. http://www.ftc.gov/bcp/workshops/glb/index.html [17 January 2002].

______________. 2002. “Gramm-Leach-Bliley Act: Financial Privacy and Pretexting.” http://www.ftc.gov/privacy/glbact/index.html [17 January].

Friedman, David. 2000. “Privacy and Technology.” *Social Philosophy & Policy* 17 (Summer): 186–212.

Gramlich, Edward M. 1999. “Statement to the U.S. House Subcommittee on Financial Institutions and Consumer Credit of the Committee on Banking and Financial Services.” 21 July 1999. *Federal Reserve Bulletin* 85 (September): 624–26.

Harper, Jim, and Solveig Singleton. 2001. “With a Grain of Salt: What Consumer Privacy Surveys Don’t Tell Us.” Manuscript, Competitive Enterprise Institute (June).

Harris Interactive Inc. 2001 “A Survey of Consumer Privacy Attitudes and Behaviors.” Manuscript.

Kahn, Charles M., James McAndrews, and William Roberds. 2000. “A Theory of Transactions Privacy.” Working Paper 2000-22. Federal Reserve Bank of Atlanta.

Kovacevich, Richard M. 2000. “Privacy and the Promise of Financial Modernization.” *The Region* 14 (March): 27–29.

Lee, W. A. 2001. “Opt-Out Notices Give No One a Thrill.” *American Banker* (10 July).

National Consumers League. 2000. “Online Americans More Concerned about Privacy than Health Care, Crime, and Taxes, New Survey Reveals.” News Release (4 October).

Paul, Pamela. 2001. “Mixed Signals.” *American Demographics* 23 (March): 45–49.

Spiekermann, Sarah, Jens Grossklags, and Bettina Berendt. No date available. “Stated Privacy Preferences versus Actual Behaviour in EC environments: A Reality Check.” Manuscript, Humboldt University.

Worldata. 2002. Worldata & WebConnect Online Datacard Library. Online database. http://www.worldata.com [17 January].

Survey Measures of Expected Inflation: Revisiting the Issues of Predictive Content and Rationality

Yash P. Mehra

The forecasting accuracy, predictive content, and rationality of survey measures of inflation expectations are important for a number of reasons. In monetary policy deliberations, the Federal Reserve needs a reliable measure of inflation expectations to assess the outlook for future inflation and gauge the stance of current monetary policy. Hence it is important to see if the widely available survey forecasts are accurate and useful in predicting actual future inflation.[1] This reliance on direct measures of inflation expectations has become more critical because of the reduced stability of the short-run relationship between monetary aggregates and GDP expenditures since the early 1980s. Furthermore, during the past two decades the Federal Reserve has conducted policy focusing on the behavior of short-term interest rates. Inflation expectations are important in identifying expected real interest rates that determine real spending in the economy.

The rationality of inflation expectations, namely that economic agents do not make systematic errors in making their forecasts of inflation, is also important. The premise that economic agents may have rational expectations is now widely accepted and employed by macroeconomists in building general

I would like to thank Michael Dotsey, Pierre-Daniel Sarte, Andreas Hornstein, and Thomas Humphrey for many helpful suggestions and Elliot Martin for excellent research assistance. The views expressed in this paper are those of the author and do not necessarily represent those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] The forecasting accuracy is measured here by the mean absolute forecast error, or the root mean squared error constructed using prediction errors.

equilibrium models and discussing effects of policy. In such models the effects of monetary policy on output and employment depend in part on whether expectations are rational. It is therefore important to examine whether the popular survey inflation forecasts exhibit rationality.

The most recent work evaluating the forecasting performance of the survey measures of expected inflation appears in Thomas (1999). I extend it in two main directions. In most previous research, the predictive content of survey measures for inflation is not adequately investigated. I examine this issue using the test of Granger-causality, which helps determine whether the survey measures contain additional information about the subsequently realized inflation rates beyond what is already contained in the past history of the actual inflation rates.[2] I allow for the possibility that survey inflation forecasts and actual inflation rates series may be cointegrated (Engle and Granger 1987). If these two series are cointegrated, then such cointegration implies that inflation forecasts and actual inflation series move together in the long run. In the short run, though, these two series may drift apart. This drift property of cointegrated series has important implications for tests of predictive content and rationality. In particular, the forecast error may have serial correlation, suggesting the presence of systematic forecast errors.[3] The fact that in the long run these two series revert to one another—with forecasts adjusting to actual inflation or inflation adjusting to forecasts, or both—implies that the short-run drifts may have predictive content for future movements in inflation. Thus, the presence of serial correlation in forecast errors and the fact that economic agents take these errors into account when they forecast future inflation are not inconsistent with the paradigm of rational expectations.[4]

The other key aspect of the survey measures of inflation examined in previous work concerns their efficiency: whether or not survey respondents employ all relevant information in generating their inflation forecasts. Inflation expectations are said to be efficient if survey respondents employ all relevant information when forecasting. In previous research, this test for efficiency was often conducted using the most recent available information on the past values of the economic variables. But data on some economic variables is subject to significant revisions over time, and so the use of revised data in

[2] This test of predictive content is more rigorous than simply asking whether survey inflation forecasts are more accurate than the naïve inflation forecasts given by the most recent inflation rate known to the respondent at the time forecasts are made. The test for Granger-causality seeks information about the future inflation rate beyond what is already contained in the entire past history of the inflation rate, not just in the most recent inflation rate.

[3] The drift caused by a shock to the fundamentals may be persistent in the short run if economic agents rationally learn the nature of the shock and the resulting true process generating the fundamentals.

[4] A recent paper by Grant and Thomas (1999) uses the cointegration and error-correction methodology in the test for rationality. The authors, however, do not examine the issue of predictive content. Moreover, they consider only the Livingston and Michigan-mean surveys.

the test for efficiency is questionable, since revised data would not have been known to the respondents at the time they made their forecasts. Tests for efficiency conducted using revised data on the relevant economic variables may then yield incorrect inferences on the rationality of survey forecasts. I investigate whether inferences on efficiency reported in previous research are sensitive to the use of real-time data.[5]

In this article, I examine the behavior of three survey measures of one-year-ahead CPI inflation expectations. I evaluate their relative forecasting accuracy and predictive content over a full sample period, from 1961:1 to 2001:3, and two subperiods, 1961:1 to 1980:2 and 1980:3 to 2001:3. The early period is the period of upward-trending inflation, and the later period is the period of downward-trending inflation.[6] The later period also coincides with a major change in the monetary policy regime, when Paul Volcker, appointed Fed Chairman in 1979, put in place a disinflationary policy. In an environment where a central bank must establish credibility for changes in its inflation targets, a rational expectations equilibrium may exist in which inflationary expectations are slow to adjust. Along the transition path, economic agents may continue to expect higher inflation than is actually realized and may thus make systematic forecast errors. In order to assess whether test results for unbiasedness and predictive content for the later period are robust to this phenomenon, I also examine the period that begins with the appointment of Alan Greenspan as Fed Chairman. I assume that the transition to a low inflation environment was credible by the end of the Volcker regime.

The three survey measures considered here are the Livingston Survey of Professional Economists (denoted hereafter Livingston); the Michigan Survey of U.S. households (denoted Michigan-mean or Michigan-median); and the Survey of Professional Forecasters (denoted SPF).[7] The Livingston and Michigan-mean forecasts are available for the full sample period, whereas the Michigan-median and SPF forecasts are available only for the later subperiod.

---

[5] Zarnowitz (1985) and Keane and Runcle (1989) are among the first to suggest that the use of revised data could affect inferences on rationality. The inference on Granger-causality could also be affected if the price series are revised. However, Consumer Price Index (CPI) inflation data has not been subject to significant revisions, so I focus on the effect of revisions in other economic variables pertinent to the test for efficiency.

[6] Here I follow Thomas (1999) in splitting the sample in the second quarter of 1980, when the CPI inflation rate peaked.

[7] The Livingston survey currently conducted by the Federal Reserve Bank of Philadelphia covers professional economists in academia, in private financial and nonfinancial corporations, and in government. The Michigan Survey currently conducted by the Survey Research Center at the University of Michigan covers U.S. households and is based on a randomly selected sample of at least 500 households. The respondents are asked to provide forecasts of the inflation rate over the next year in the prices of "things you buy." The survey has been conducted quarterly from 1959 through 1977 and monthly since the beginning of 1978. The Survey of Professional Forecasters covers professional forecasters in the business sector for the most part and is currently conducted by the Federal Reserve Bank of Philadelphia. Consumer Price Index inflation forecasts were initiated in the third quarter of 1981.

As a benchmark, I consider one naïve forecast, which is simply the most recent one-year growth rate of CPI inflation known to the survey respondents at the time forecasts are made.[8]

The empirical work presented here supports the following observations. First, all survey measures considered here are more accurate than the naïve forecast. However, as regards their relative forecast accuracy, the results are sensitive to the sample period. While both the Livingston and Michigan-mean forecasts perform equally well over the full period and the period of rising inflation, the Michigan-mean forecasts are the least accurate over the period of downward-trending inflation. For this later period, the Michigan-median forecasts provide the most relatively accurate forecasts of one-year-ahead CPI inflation.

Second, tests for Granger-causality indicate that survey forecasts considered here contain a forward-looking component and can help predict actual future inflation, with the exception of the Livingston forecasts. The Livingston forecasts do not Granger-cause inflation over the full period, implying they have no predictive content for future inflation.

Third, the Michigan-median forecasts are unbiased, but the results of the others are mixed. The Livingston forecasts are unbiased over the full period, but biased over the early and later periods. The Michigan-mean forecasts are biased over the full and later periods, but unbiased over the early.

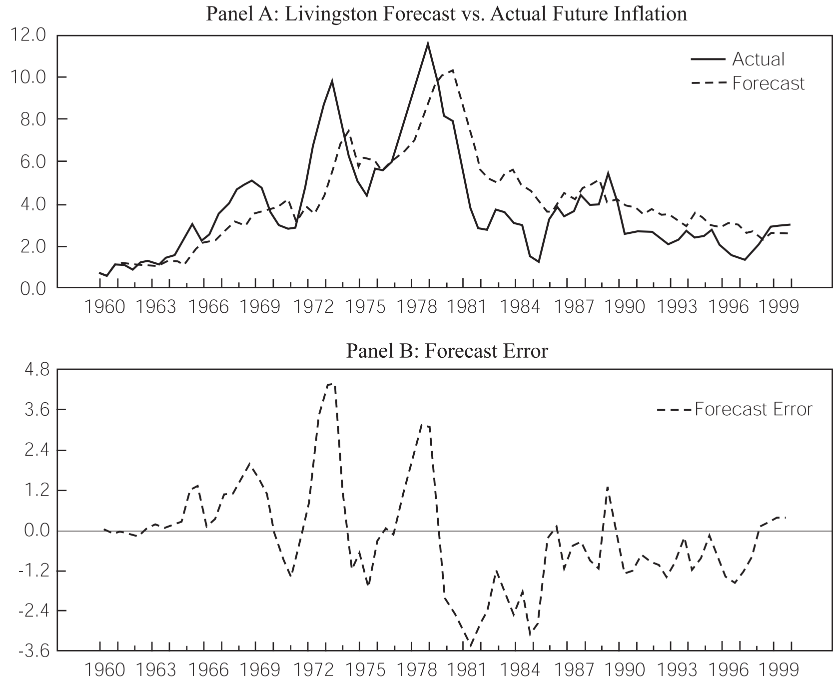
Fourth, tests for efficiency performed using revised data indicate that the forecast error is correlated with past information, including the output gap. This result implies that survey respondents did not take into account past information in making their predictions, a result already reported in Thomas (1999). However, real-time estimates of the output gap differ substantially from those generated using revised data. If tests for rationality are conducted using real-time data, then their results indicate that survey respondents did take into account past information in predicting future inflation.

Finally, excluding the Volcker period from the later period does not dramatically alter the results. There is an increase in forecast accuracy as measured by the mean error or the root mean squared error criterion; however, the Livingston and Michigan-mean forecasts remain biased. The SPF forecasts look much better over this short period, being unbiased and almost as accurate as the Michigan-median forecasts.

Section 1 provides a graphical review of the recent behavior of three survey measures considered here. It also describes the various statistical tests that are

[8] The other benchmark inflation model commonly used in previous work is based on the Fisher model of interest rates. According to the Fisher model, the nominal interest rate at any time can be regarded as the sum of the expected real interest rate and the expected rate of inflation. Given an estimate of the expected real interest rate, one can then recover estimates of the expected inflation rate from the nominal interest rate. This benchmark forecasting model has, however, not done well (see Thomas [1999]).

Figure 1



used to evaluate the survey forecasts. Section 2 presents the empirical results, and concluding observations are in Section 3.

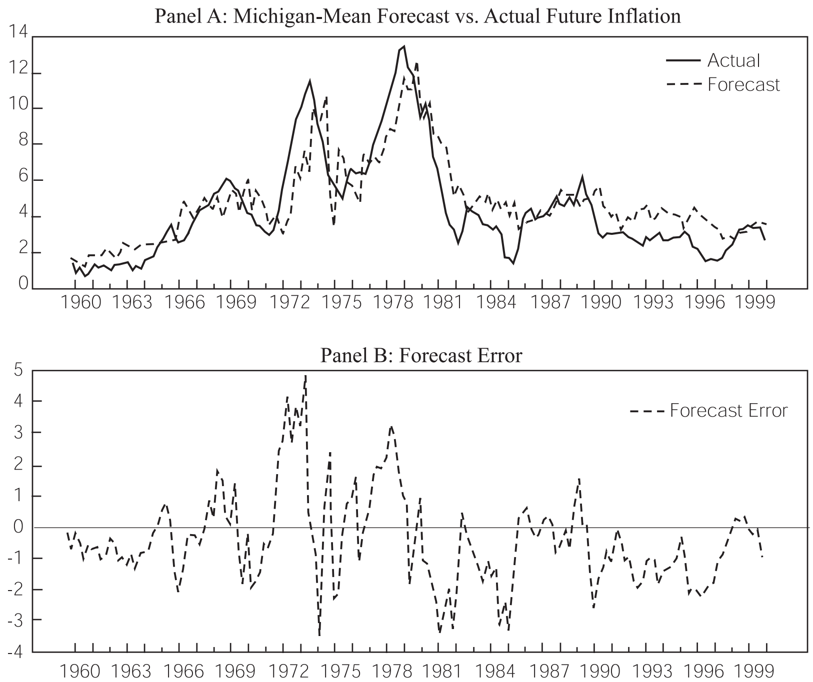
1. EMPIRICAL METHODOLOGY

Various statistical tests are used to assess the forecast accuracy, predictability, and rationality of survey measures. I begin with a graphical review of the recent behavior of these survey measures and then describe the tests themselves.

Figures 1 through 4 chart the Livingston, Michigan-mean, Michigan-median, and SPF inflation forecasts, along with the subsequently realized CPI inflation rates for the pertinent sample periods.[9] Panel B in each figure charts the forecast error, defined as the subsequently realized CPI inflation

[9] The Livingston survey is semiannual and published in June and December of each year. The Livingston survey forecasts actually cover a 14-month period, because respondents who are asked to forecast the level of CPI expected to prevail the following June and December have information about the actual level of CPI for April and October. In contrast, the Michigan survey has been conducted quarterly from 1959 through 1977 and monthly since then. Hence, observations in the Livingston survey are semiannual and cover a 14-month-ahead period, whereas in the Michigan

Figure 2

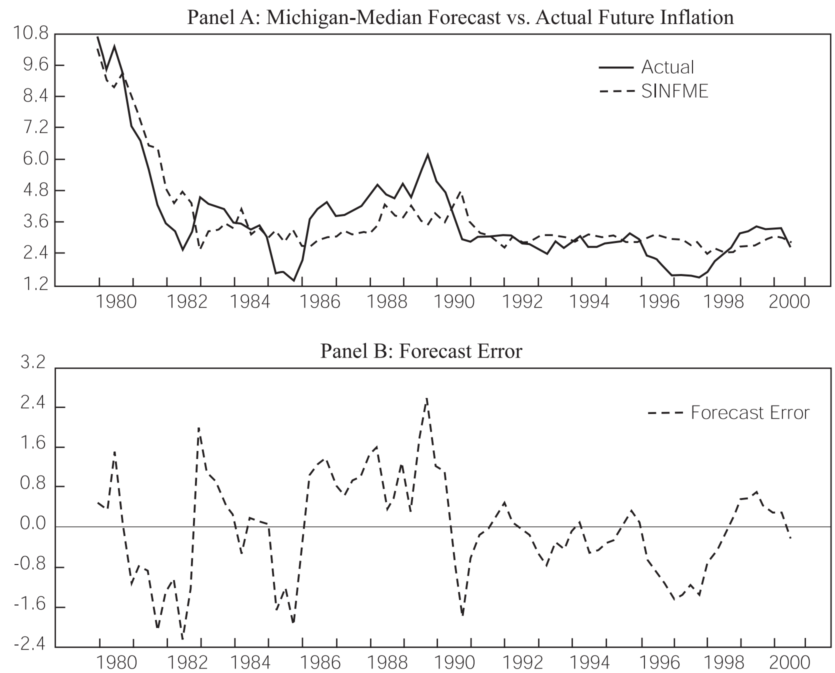


minus its survey forecast. Several observations stand out. First, if we focus on the Livingston and Michigan-mean forecasts that are available over the full period, we see that the turning points in expected inflation appear to lag behind the turning points in actual inflation, suggesting the presence of a backward-looking component in inflation expectations. Furthermore, both Livingston and Michigan respondents appear to underestimate inflation in the early period, when inflation is trending upward, and overestimate inflation in the later period, when it is trending downward (see Figures 1 and 2).

Second, if we focus on the Michigan-median and SPF forecasts available only for the 1980s and the 1990s (see Figures 3 and 4), the SPF respondents also overestimate inflation in periods when inflation is falling. In particular, the SPF respondents seriously underestimated the decline in inflation that occurred in the early 1980s (see Panel B of Figure 4). The Michigan-median inflation forecasts look good in comparison, the extent of overprediction being relatively mild.

survey they are quarterly and cover a one-year-ahead period. See Thomas (1999) for a recent overview of other details.

Figure 3



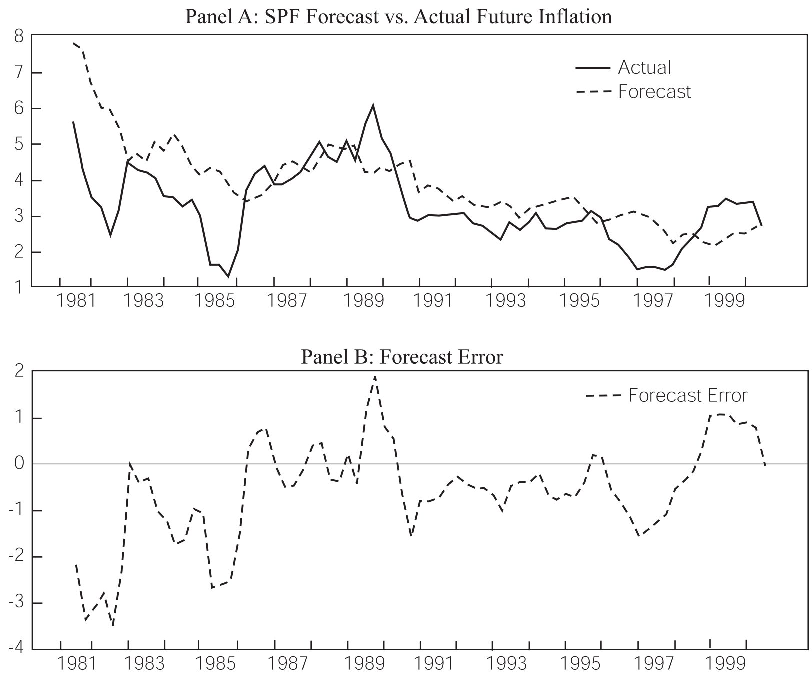
Although Figures 1 through 4 indicate that survey inflation forecasts move together with the subsequently realized inflation rates, it is not clear whether this comovement results from survey respondents adjusting their forecasts in response to past inflation rates or anticipating actual future inflation rates. From a policy perspective, survey measures of expected inflation are useful if they help predict actual future inflation rates. Hence, I examine their predictive content using the test of Granger-causality, allowing for the possibility that survey inflation forecasts and actual realizations of inflation may be cointegrated, as in Engle and Granger (1987). In particular, consider the following regressions:

$$\Delta A_t = g_0 + \lambda_a (A_{t-1} - S_{t-1}) + \sum_{k=1}^{n} g_{1k} \Delta A_{t-s} + \sum_{k=1}^{n} g_{2k} \Delta S_{t-s} + \varepsilon_{1t} \quad (1)$$

and

$$\Delta S_t = g_0 + \lambda_s (A_{t-1} - S_{t-1}) + \sum_{k=1}^{n} g_{3k} \Delta A_{t-s} + \sum_{k=1}^{n} g_{4k} \Delta S_{t-s} + \varepsilon_{2t}, \quad (2)$$

Figure 4



where A is the actual future inflation rate, S is the survey inflation forecast, and ε_s are disturbance terms. Survey measures Granger-cause inflation if $\lambda_a \neq g_{2k} \neq 0$. In that case, survey inflation forecasts provide information about the subsequently realized inflation rates beyond what is already contained in the past history of actual inflation. Similarly, inflation Granger-causes survey measures if $\lambda_s \neq g_{3k} \neq 0$. In that case, inflation has information about future survey measures beyond what is already contained in the past history of survey measures. In the context of these regressions, survey measures are completely backward looking in expectation formation if $\lambda_a = g_{2k} = 0$, but $\lambda_s \neq g_{3k} \neq 0$.

Regressions (1) and (2) include a variable that measures deviations of the actual future inflation rates from their survey forecasts. The hypothesis that actual future inflation rates and survey forecasts may be cointegrated in the long run implies that these two series will move together in the long run.[10] In the short run, they may drift apart, but ultimately they will revert toward

[10] The results here (not reported) are consistent with the evidence in Grant and Thomas (1999) that Livingston and Michigan forecasts are cointegrated with actual inflation.

one another if they are cointegrated. This comovement may, however, occur when survey forecasts revert to actual realization of inflation ($\lambda_s \neq 0$ in (2)), actual future inflation reverts to survey forecasts ($\lambda_a \neq 0$ in (1)), or both adjust in response to such deviations ($\lambda_a \neq 0, \lambda_s \neq 0$). The variable that measures deviations is usually referred to as the error-correction variable, and the coefficients (λ_a, λ_s) are referred to as the error-correction coefficients. From a policy perspective, the most interesting case is the one in which the adjustment occurs mostly through actual realizations of inflation reverting to survey forecasts, so that $\lambda_a \neq 0$ but $\lambda_s = 0$. In that case, survey forecasts have predictive content for future inflation.

Tests of rationality of survey measures have emphasized two key properties of rational expectations. One, they should be unbiased in the sense survey respondents forecast inflation correctly on average. Two, forecasts should be efficient in that survey respondents should consider all information pertinent to the future behavior of inflation. The test for bias is usually implemented by running the following regression:

$$A_t = a_0 + a_1 S_t + \nu_t \tag{3}$$

where A is actual future inflation rate, S is the survey forecast, and ν is the disturbance term. Survey forecasts are unbiased if $a_0 = 0, a_1 = 1$.[11] Similarly, if survey forecasts are efficient, then the forecast error should not be correlated with known, pertinent information. The test for efficiency is often implemented by running the following regression:

$$e_t = b_0 + b_1 I_{t-1} + \eta_t, \tag{4}$$

where e_t is the forecast error ($A_t - S_t$), I is the information set containing variables pertinent to the behavior of inflation, and η is the disturbance term. Survey forecasts are said to be efficient if the forecast error is uncorrelated with the variables in the information set I, either individually or jointly.[12] This statement implies that the coefficients vector $b_1 = 0$.

The efficiency test brings up two other issues. In previous work, the test for efficiency has generally been performed including the economic variables in (4), one at a time, as in Thomas (1999). But, as noted in Maddala (1990), inferences on efficiency based on the individual consideration of economic

[11] The test for unbiasedness is generally conducted including the constant term, implicitly allowing for the possibility that actual inflation may not at all be correlated with the survey forecasts. Hence, the specification (3) nests this hypothesis.

[12] For rational agents, the question of what variables should be included in the information set depends on costs and benefits. Since past values of a variable being forecast (inflation) are readily available, that variable should be in the information set. But this cannot be said of other variables. The agents will set the marginal cost equal to the marginal benefit of acquiring information. This analysis leads to the distinguishing of weak-form efficiency, where the information set includes only past values of the variable being forecast, from strong-form efficiency, where the information set also includes past values of other variables. A good review appears in Maddala (1990).

variables may change when variables are considered jointly.[13] The empirical work here therefore considers economic variables both individually and jointly. The other issue in the test for efficiency concerns the use of revised as opposed to real-time data. In most previous work, the tests were performed using the revised data on the past values of the economic variables in the information set. But many analysts, including Keane and Runkle (1989) and Maddala (1991), correctly point out that such revised data would not have been known to the survey respondents at the time they made their predictions. It is suggested that real-time data on the past values of the economic variables should be used in the test for efficiency.

In addition to the tests for predictive content and rationality, I also present summary error statistics that measure the overall predictive accuracy of survey forecasts. The summary statistics considered here are the mean error (ME), the mean absolute error (MAE), and the root mean squared error ($RMSE$). The mean error is a simple measure of forecasting bias; a positive mean error implies that survey respondents on average underestimated inflation. The mean absolute error and the root mean squared error are measures of forecasting accuracy. If a string of positive forecast errors is accompanied by a string of negative forecast errors, the survey respondents may issue forecasts with a zero mean error, but large mean absolute errors. The root mean squared error is the other measure of forecast accuracy. Since the root mean squared error is the square root of the mean value of the squares of the forecast errors, large forecast errors have a greater effect on the $RMSE$ than the MAE.

2. EMPIRICAL RESULTS

Table 1 presents the summary error statistics for the full sample period 1961:1 to 2001:3, as well as for two subperiods, denoted as before the early period (1961:1 to 1980:2) and the later period (1980:3 to 2001:3). It also contains results for the Greenspan period and presents the relevant error statistics for the naïve inflation forecasts. The forecasting accuracy of a survey measure relative to the benchmark naïve forecast is assessed by computing the ratio, defined as the $RMSE$ of the survey forecast divided by the $RMSE$ of the naïve forecast. If this ratio is less than unity for a survey forecast, then it means the survey forecast is more accurate than the benchmark forecast.

The results on forecast accuracy reported in Table 1 suggest the following observations. First, the three survey forecasts considered here are more accurate than the naïve forecast, indicating that survey measures contain information about future inflation rates beyond what is already contained in the most recent past inflation rate. Second, the mean error is positive in the early period

[13] Tests for efficiency based on including variables one at a time would be subject to the biases generated by the omission of other relevant variables.

Table 1 Forecasting Accuracy of Survey Measures of Expected Inflation Ahead CPI

Survey	Mean Error	Mean Absolute Error	Root Mean Squared Error	Ratio
	(1)	(2)	(3)	(4)
	Panel A: Full Period 1961:1–2000:3			
Livingston	−0.22	1.17	1.57	0.73
Michigan-Mean	−0.43	1.21	1.55	0.73
Naïve	0.06	1.53	2.14	
	Panel B: Early Period 1961:1–1980:2			
	(1)	(2)	(3)	(4)
Livingston	0.66	1.11	1.59	0.66
Michigan-Mean	0.17	1.23	1.63	0.67
Naïve	0.75	1.76	2.42	
	Panel C: Later Period 1980:3–2000:3			
	(1)	(2)	(3)	(4)
Livingston	−1.14	1.25	1.55	0.51
Michigan-Mean	−1.00	1.19	1.48	0.81
Michigan-Median	−0.03	0.78	0.98	0.53
Professional Forecasters*	−0.60	0.95	1.24	0.68
Naïve	−0.51	1.29	1.83	
	Panel D: Greenspan Period 1987:4–2000:3			
	(1)	(2)	(3)	(4)
Livingston	−0.65	0.82	0.94	0.94
Michigan-Mean	−0.89	1.00	1.25	1.25
Michigan-Median	0.01	0.66	0.86	0.86
Professional Forecasters*	−0.24	0.69	0.80	0.80
Naïve	−0.08	0.74	1.00	

*For Professional Forecasters, the sample period is 1981:3–2000:3.

Notes: The naïve forecast is simply a backward-looking forecast, measured here by the recent one-year CPI inflation known to the survey respondent at the time the forecast is made. Ratio is the root mean squared error of the survey forecasts divided by the root mean squared error of the naïve forecasts; a value below unity indicates that the survey forecasts outperform the naïve forecasts. The forecast horizon for the Livingston forecasts is the 14-month period.

and negative in the later period for both the Livingston and Michigan-mean forecasts. The SPF forecasts that are available only for the later period have a negative mean error. Those results suggest that survey respondents underestimated inflation in the early period, when inflation was trending upward, and overestimated inflation in the later period, when inflation was trending downward. The exception is the Michigan-median forecasts, which are available only for the later period and have a mean error that is negligible. These results are in line with those in Thomas (1999).

As Table 1 shows, for the later period the forecast bias is generally negative, implying that survey respondents overestimated inflation. There is a substantial reduction in the size of the bias if the Volcker period is excluded, implying that survey respondents probably did not believe in the deflationary nature of Fed policy when it was first put in place in 1979 (see Panels C and D, Table 1).[14] One key aspect of these results is that the negative bias appears in the Michigan-mean forecasts, but not in the Michigan-median forecasts. This difference occurs because a small percentage of the households constituting the Michigan respondents overestimated inflation by a large amount over the period. This feature of Michigan household forecasts has the effect of inflating the mean value of the forecasts but not the median, so the negative bias persists in the Michigan-mean forecasts (Thomas 1999).

The survey forecasts are somewhat more accurate than a benchmark naïve forecast. This result implies that survey forecasts have some information about future inflation beyond that already contained in the most recent past inflation rate. I now consider the results of the test for Granger-causality reported in Table 2, a more rigorous test of predictive content. As the table shows, (see χ_1^2 statistics), with the exception of the Livingston forecasts, survey forecasts considered here Granger-cause inflation, implying that survey forecasts have information about the subsequently realized inflation rates beyond what is already contained in the past history of actual inflation rates. The results for the Livingston forecasts are mixed: the Livingston forecasts do not Granger-cause inflation in the full and later periods. In contrast, inflation Granger-causes all three survey forecasts, implying the presence of a backward-looking component in the formation of inflationary expectations (see χ_2^2 statistics in Table 2).

The error-correction variable is usually significant in equation (1) for explaining changes in the realizations of future inflation rates when the Michigan-mean, Michigan-median, and SPF forecasts are used (see Table 2). This result implies that in the short run a persistent deviation of the survey forecast from inflation is corrected in part through adjustment of actual future inflation rates.

[14] This is consistent with the evidence in Dotsey and DeVaro (1995), indicating the deflation of the early 1980s was not anticipated by economic agents.

Table 2 Test for Predictive Content

Panel A: Full Period 1961:1–2000:3

Survey	λ_a	$S1$	χ_1^2	λ_s	$S2$	χ_2^2
Livingston	−0.02 (0.2)	−0.15 (0.5)	03.4	0.24 (5.5)	0.18 (1.2)	122.9*
Michigan-Mean	−0.10 (1.8)	00.11 (0.9)	16.9*	0.23 (3.5)	0.68 (2.2)	36.6*

Panel B: Early Period 1961:1–1980:2

	λ_a	$S1$	χ_1^2	λ_s	$S2$	χ_2^2
Livingston	−0.70 (2.6)	−0.26 (0.5)	67.2*	0.14 (1.1)	0.58 (1.7)	19.9*
Michigan-Mean	−0.10 (2.2)	−0.15 (1.2)	19.3*	0.26 (2.4)	0.86 (1.5)	53.1*

Panel C: Later Period 1980:3–2000:3

	λ_a	$S1$	χ_1^2	λ_s	$S2$	χ_2^2
Livingston	−0.21 (1.0)	0.55 (0.9)	07.8	0.14 (2.2)	0.25 (1.4)	71.6*
Michigan-Mean	−0.23 (2.5)	0.37 (1.9)	14.6*	0.17 (2.6)	0.40 (2.3)	101.4*
Michigan-Median	−0.20 (2.8)	0.62 (3.0)	37.8*	0.06 (1.0)	0.60 (4.1)	77.1*
Professional Forecasters	−0.19 (2.3)	0.57 (2.1)	35.3*	0.05 (1.3)	0.37 (2.6)	58.0*

Panel D: Greenspan Period 1987:4–2000:3

	λ_a	$S1$	χ_1^2	λ_s	$S2$	χ_2^2
Livingston	−0.52 (2.5)	0.43 (0.5)	94.1*	0.02 (0.1)	0.44 (1.1)	13.4*
Michigan-Mean	−0.13 (1.6)	0.25 (1.2)	20.6*	0.15 (2.0)	0.39 (1.8)	65.1*
Michigan-Median	−0.14 (2.0)	0.59 (2.1)	28.6*	0.05 (1.9)	0.50 (3.1)	35.4*
Professional Forecasters	−0.22 (2.9)	0.52 (1.8)	24.4*	0.02 (0.7)	0.29 (1.5)	65.2*

*Significant at the 5 percent level.

Notes: The coefficients reported above are from regressions of the form

$$\Delta A_t = a_0 + \lambda_a (A_{t-1} - S_{t-1}) + \sum_{s=1}^{k} a_{1s} \Delta A_{t-s} + \sum_{s=1}^{k} a_{2s} \Delta S_{t-s} + \varepsilon_1$$

$$\Delta S_t = a_0 + \lambda_s (A_{t-1} - S_{t-1}) + \sum_{s=1}^{k} a_{3s} \Delta A_{t-s} + \sum_{s=1}^{k} a_{4s} \Delta S_{t-s} + \varepsilon_2,$$

where A is actual future inflation, and S is the survey inflation forecast. Parentheses contain t-values. $S1$ is $\sum_{s=1}^{k} a_{2s}$ and $S2$ is $\sum_{s=1}^{k} a_{3s}$. χ_1^2 tests ($\lambda_a = 0; a_{2s} = 0$) and χ_2^2 tests ($\lambda_s = 0; a_{3s} = 0$). The regressions above are estimated by ordinary least squares, the standard errors being corrected for the presence of serial correlation. The parameter k measures the lag length, which is set at 4. The sample period is 1981:3–2000:3 for Professional Forecasters.

Therefore, these survey forecasts have predictive content for actual future inflation.

Table 2 also presents the sum of coefficients that appear on lagged values of realized inflation in forecasting equations of the form (2) (see $S2$ in Table 2). We may interpret this sum coefficient as a measure of the degree of backward-looking behavior in expectation formation of survey respondents. In the later period, this coefficient is usually larger for Michigan-median households than for Livingston or SPF respondents, indicating that Michigan-median households paid more attention to past realized inflation rates when making inflation predictions than did the Livingston or SPF respondents. Since inflation has trended downward in the later period, in part due to change in monetary policy regime, Michigan-median households predict actual inflation well compared to professional economists and forecasters. It appears that Livingston economists and SPF forecasters did not believe the deflation of the early 1980s was there to stay, so they continued to give less weight to lower realized inflation rates.

Tables 3 and 4 present tests for rationality. Table 3 contains test results for unbiasedness and Table 4 for efficiency with respect to past information on economic variables pertinent to the behavior of inflation. If we focus on the results for unbiasedness in Table 3, three observations stand out. First, test results for the Livingston and Michigan-mean forecasts are sensitive to the sample period. The Livingston forecasts are unbiased over the full period, but biased within each period. The Michigan-mean forecasts are biased over the full period and the later period, but unbiased over the early period. Second, for the later period of downward trending inflation, all survey forecasts considered here are biased except the Michigan-median forecasts. Excluding observations pertaining to the Volcker period does not alter results on the biasedness of the Livingston and Michigan-mean forecasts (see Panel D in Table 3).

As I discussed earlier, tests for efficiency in previous research have generally been reported using revised data on the past values of the economic variables. The economic variables that have usually been employed are actual inflation, money growth, increase in oil prices, and the level of the output gap. The empirical work reported in Thomas (1999) indicates that the forecast error in the Livingston and Michigan-mean forecasts is correlated with the level of the output gap but none other of the economic variables. This result implies that survey respondents considered past values of actual inflation, money growth, and energy price inflation, but ignored the behavior of the output gap.

The forecast error may be correlated with the past values of the output gap because of the use of the revised data on the output gap. The recent work in Orphanides and van Norden (2002) shows that real-time estimates of the level of the output gap are generally subject to significant revisions. If this is true, then the revised data on the output gap used in tests for efficiency would not have been available to the survey respondents. This result can be

Table 3 Test for Unbiasedness

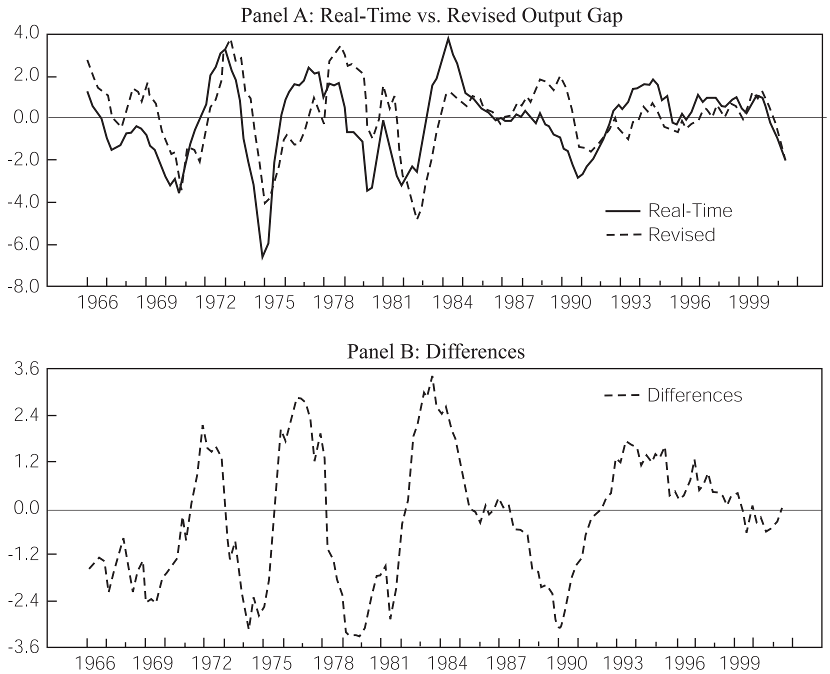
Survey	a	b	R-Squared	χ^2
Panel A: Full Period 1961:1–2000:3				
	(1)	(2)	(3)	(4)
Livingston	0.26 (0.6)	0.88 (05.6)	0.59	0.91
Michigan-Mean	−0.80 (2.0)	1.00 (11.4)	0.73	9.33*
Naïve	1.40 (3.3)	0.70 (05.8)	0.51	11.20*
Panel B: Early Period 1961:1–1980:2				
	(1)	(2)	(3)	(4)
Livingston	0.55 (1.5)	1.02 (10.6)	0.76	5.5**
Michigan-Mean	−0.36 (0.8)	1.10 (11.3)	0.79	1.1
Naïve	1.59 (2.4)	0.80 (5.2)	0.56	6.6*
Panel C: Later Period 1980:3–2000:3				
	(1)	(2)	(3)	(4)
Livingston	0.56 (1.5)	0.58 (7.2)	0.56	59.3*
Michigan-Mean	−0.20 (0.3)	0.81 (5.5)	0.56	30.2*
Michigan-Median	0.37 (0.8)	0.89 (7.5)	0.60	0.9
Professional Forecasters	1.42 (2.7)	0.48 (3.4)	0.26	21.1*
Naïve	1.73 (4.0)	0.44 (4.0)	0.50	26.9*
Panel D: Greenspan Period 1987:4–2000:3				
	(1)	(2)	(3)	(4)
Livingston	−0.16 (0.3)	0.85 (4.5)	0.45	19.1*
Michigan-Mean	−0.14 (0.1)	0.81 (5.5)	0.30	16.8*
Michigan-Median	−0.90 (0.7)	1.30 (3.1)	0.35	0.5
Professional Forecasters	−0.00 (0.1)	0.96 (3.9)	0.48	1.9

*Significant at the 5 percent level.

**Significant at the 10 percent level.

Notes: The coefficients reported above are from regressions of the form $A_t = a + b\, P_t + e_t$, where A is the actual future inflation rate and P is its survey forecast. Inflation forecasts are unbiased if $a = 0, b = 1$. χ^2 is the Chi-square statistic that tests the null hypothesis $a = 0, b = 1$. Ordinary least squares are used, and the standard errors are corrected for the presence of serial correlation. Parentheses contain t-values.

Figure 5



seen in Figure 5, which charts real-time and final estimates of the output gap, generated using the historical real-time data in Croushore and Stark (1999).[15] Figure 6 presents real-time and revised data on money growth. It shows that the level of the output gap has been subject to far more significant revisions than has the measure of money growth (compare the revisions charted in Panel B of Figure 5 with that in Figure 6).

Table 4 presents test results for efficiency using both revised and real-time estimates of the output gap. I also use real-time estimates of money growth in tests for efficiency.[16] The forecast error in the Livingston and Michigan-mean forecasts is correlated with the output gap variable when revised data are used, but this correlation weakens or disappears when real-time data are used (compare t-values on the gap variable in Panels A and B of Table 4). Also, the forecast error in the Livingston and Michigan-mean forecasts is correlated

[15] The measure of the output gap used in Thomas (1999) is the Hodrick-Prescott filtered estimate of the output gap. I use the same filter, but employ the real-time historical data available on output to generate estimates of the output gap series.

[16] Since real-time data available in Croushore and Stark (1999) begins in 1966, the sample period covering the tests for efficiency starts in 1966:1.

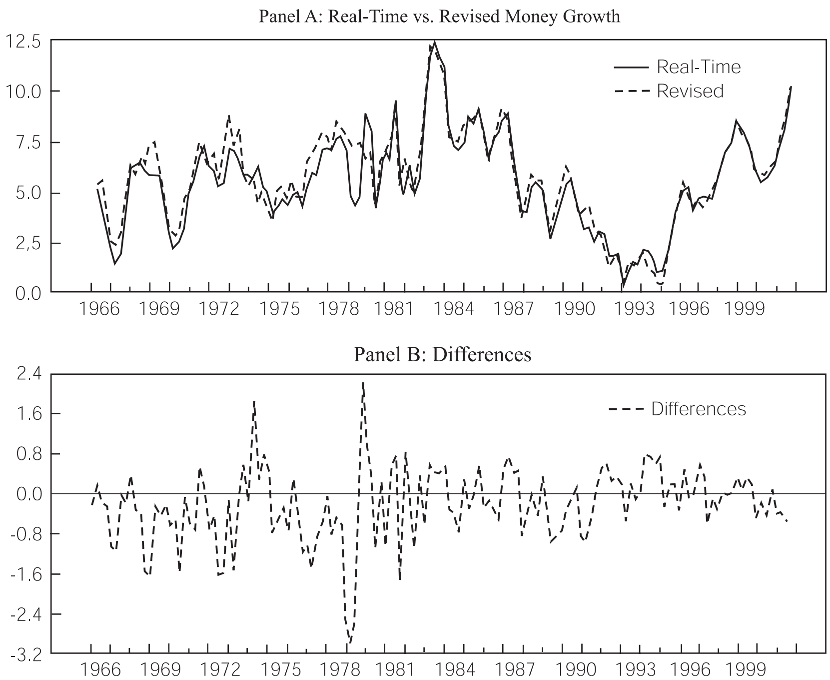
Table 4 Test for Efficiency

	Panel A: Livingston, 1961:1–2000:3					
	Revised Data			Real-Time Data		
Independent Variable (X)	c_o	c_1	χ^2_1	c_0	c_1	χ^2_1
Inflation	−0.07 (0.2)	−0.03 (0.5)		−0.26 (0.5)	0.01 (0.1)	
Gap	−0.24 (0.9)	0.45 (3.4)		−0.26 (0.7)	0.27 (1.4)	
Money Growth	−0.33 (0.7)	0.02 (0.1)		0.17 (0.3)	−0.09 (0.7)	
Oil Prices	−0.23 (0.7)	0.00 (0.1)		−0.31 (0.9)	0.00 (0.1)	
Jointly			21.6*			4.9
	Panel B: Michigan-Mean, 1961:1–2000:3					
	c_o	c_1	χ^2_1	c_0	c_1	χ^2_1
Inflation	0.58 (2.0)	0.03 (0.5)		00.54 (1.4)	0.03 (0.4)	
Gap	−0.42 (2.3)	0.32 (2.4)		−0.36 (1.5)	0.23 (1.5)	
Money Growth	−1.2 (3.9)	0.14 (1.8)		−0.87 (2.4)	0.08 (1.1)	
Oil Prices	−0.42 (1.9)	0.00 (0.7)		−0.40 (1.6)	0.00 (0.7)	
Jointly			8.8*			5.2
	Panel C: Michigan-Median, 1980:1–2000:3					
	c_o	c_1	χ^2_1	c_0	c_1	χ^2_1
Inflation	0.51 (1.4)	−0.15 (1.7)		0.51 (1.4)	−0.15 (1.7)	
Gap	−0.02 (0.1)	0.05 (0.3)		−0.03 (0.2)	0.02 (0.8)	
Money Growth	−0.33 (1.2)	0.05 (1.3)		−0.28 (0.9)	0.04 (0.9)	
Oil Prices	0.03 (0.2)	0.00 (0.9)		−0.03 (0.2)	0.00 (0.9)	
Jointly			5.3			7.2
	Panel D: Professional Forecasters, 1981:3–2000:3					
	c_o	c_1	χ^2_1	c_0	c_1	χ^2_1
Inflation	00.62 (2.0)	−0.34 (4.0)		0.62 (2.0)	0.34 (4.0)	
Gap	−0.58 (2.8)	0.20 (1.7)		−0.61 (2.7)	0.09 (0.5)	
Money Growth	−0.44 (1.9)	0.03 (0.6)		−0.40 (1.5)	0.03 (0.7)	
Oil Prices	−0.60 (2.7)	0.00 (0.3)		−0.60 (2.7)	0.00 (0.3)	
Jointly			37.4*			41.8*

*Significant at the 5 percent level.

Notes: The coefficients reported above are from regression of the form $e_t = c_0 + c_1 X_{t-1}$, where e is the forecast error and X_{t-1} is the lagged yearly growth rate of prices or money or oil prices, or the level of the output gap. Gap is the Hodrick-Prescott filtered estimate of the output gap. The regressions are estimated including one variable at a time as well as all of them together (jointly). Parentheses contain t-values. χ^2_1 tests all variables that when included jointly are not significant in explaining the forecast error.

Figure 6



with lagged inflation, money growth, energy price inflation, and output gap variables when they are jointly included in the pertinent regression estimated using revised data. But this correlation again disappears when real-time data are used (compare χ_1^2 statistics in Panels A and B of Table 4). These results indicate caution is merited when interpreting the results on efficiency derived using revised data.[17]

Another notable result is that for the later period of downward trending inflation, the SPF forecasts are correlated with the past values of inflation, suggesting that professional forecasters ignored the past information in actual inflation rates. In contrast, the forecast errors in Michigan-median forecasts are not correlated at all with any of the economic variables in the information set used here. These results hold even when real-time data are used (see Panels C and D in Table 4).

[17] This result may not be surprising given the results of some recent research. Orphanides and van Norden (2002) present evidence indicating real-time estimates of the output gap do not do as well in predicting inflation, as do the estimates based on the revised data. Amato and Swanson (2001) also report considerable reduction in the predictive content of money for output when real-time data on money growth is used.

3. CONCLUDING REMARKS

I have examined the forecasting accuracy, predictive content, and rationality of three survey measures of one-year-ahead CPI expected inflation: the Livingston forecasts of professional economists, the mean and median forecasts of Michigan households, and the consensus forecasts of the professional forecasters. Three interesting findings emerge from this analysis. First, the median inflation forecasts of Michigan households outperform those of professional economists and forecasters in the period covering the 1980s and 1990s. They are more accurate, unbiased, have predictive content for future inflation, and are efficient with respect to economic variables generally considered pertinent to the behavior of inflation. Second, in the full period the Livingston inflation forecasts appear unbiased and efficient, but those properties do not carry over to the subperiods studied here. Third, the inflation forecasts of professional forecasters are biased and inefficient. The results in the article indicate that Livingston and SPF survey respondents overestimated inflation in the deflationary period of the early 1980s and the 1990s and that they were slow in adjusting their inflation expectations in response to lower realized inflation rates, generated in part by change in the monetary policy regime. The fact that the survey respondents overestimated may explain in part why inflation forecasts of professional economists and forecasters do not perform well relative to those of Michigan households.

REFERENCES

Amato, Jeffery D., and Norman R. Swanson. 2001. "The Real-Time Predictive Content of Money for Output." *Journal of Monetary Economics* 48 (August): 3–24.

Croushore, Dean. 1997. "The Livingston Survey: Still Useful After All These Years." Federal Reserve Bank of Philadelphia *Business Review* (March-April): 15–26.

______________, and Tom Stark. 1999. "Real-time Data Set for Macroeconomists: Does the Data Vintage Matter?" Federal Reserve Bank of Philadelphia Working Paper 21.

Dotsey, Michael, and Jed L. Devaro. 1995. "Was the Disinflation of the Early 1980s Anticipated?" Federal Reserve Bank of Richmond *Economic Quarterly* (Fall): 41–59.

Engle, Robert F., and C. W. J. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55 (March): 251–76.

Grant, Alan P., and Lloyd B. Thomas. 1999. “Inflationary Expectations and Rationality Revisited.” *Economics Letters* 62 (March): 331–38.

Keane, Michael P., and David E. Runkle. 1989. “Are Economic Forecasts Rational?” Federal Reserve Bank of Minneapolis *Quarterly Review* (Spring): 26–33.

Maddala, G. S. 1991. “Survey Data on Expectations: What Have We Learnt?” In *Issues in Contemporary Economics, Macroeconomics, and Econometrics: Proceedings of the Ninth World Congress of the International Economic Association,* vol. 2. ed. Marc Nerlove. New York: New York University Press: 319–44.

Orphanides, Athanasios. 2001. “Monetary Policy Rules Based on Real-time Data.” *The American Economic Review* 91 (September): 964–85.

__________, and Simon van Norden. 2001. “The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time.” Mimeo, Board of Governors of the Federal Reserve System (September).

__________. “The Unreliability of Output Gap Estimates in Real Time.” *Review of Economics and Statistics*. Forthcoming.

Thomas, Lloyd B. 1999. “Survey Measures of Expected U.S. Inflation.” *Journal of Economic Perspectives* 13 (Fall): 125–44.

Zarnowitz, Victor. 1985. “Rational Expectations and Macroeconomic Forecasts.” *Journal of Business and Economic Statistics* 3 (October): 293–311.

Private Money and Counterfeiting

Stephen D. Williamson

Perhaps the most fundamental question in monetary economics pertains to the role of the government in providing money. A widely held view among economists is that the supply of media of exchange is an activity that should not be left to the private sector. Indeed, even Milton Friedman, who in most respects has viewed the economic role of the government quite narrowly, argues in Friedman (1960) that the provision of money is fraught with peculiar market failures and that the government should have a monopoly in the supply and control of the stock of circulating currency.

Monetary systems that include the private provision of circulating media of exchange were not uncommon in the past. In the United States, most of the stock of currency in circulation prior to the Civil War consisted of notes issued by state-chartered banks. The U.S. pre–Civil War monetary system has been judged by some, but not all, as chaotic (Rolnick et al.1997; Rolnick and Weber 1983, 1984), since it included thousands of note-issuing banks and the quality of these notes was difficult to distinguish. Counterfeiting was a problem, and there was sometimes poor information on a particular bank's chances of defaulting. However, the Suffolk Banking System in pre–Civil War New England is thought to have functioned quite efficiently (see Smith and Weber [1998]). In addition, the monetary system in place in Canada prior to 1935 featured private note issue by a small number of chartered banks, and this system also appears to have worked quite well (see Williamson [1999] and Champ, Smith, and Williamson [1996]).

Private money systems are not just of historical interest. In the United States, the government monopoly on the issue of circulating media of exchange

Chester A. Phillips Professor of Financial Economics, University of Iowa, and Visiting Scholar, Federal Reserve Bank of Richmond. Thanks go to Huberto Ennis, Tom Humphrey, John Weinberg, and Alex Wolman for their helpful comments and suggestions. The views expressed in this article are those of the author and are not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

resulted from the federal taxation after the Civil War of the notes issued by state-chartered banks, and from the elimination of the supply of government bonds qualifying as backing for notes issued by national banks. As argued by Schuler (2001), all serious federal impediments to private bank note issue in the United States were removed in 1976 and 1994 (also see Lacker [1996]). Thus, it would seem that private banks in the United States are currently free to issue circulating pieces of paper, though how U.S. regulators would respond to private note issue is uncertain. New transactions technologies also give financial institutions the capability to issue private electronic monies, such as stored-value cards, and several banks have conducted market trials of such products.

The purpose of this article is to study some of the benefits and costs of private money issue. The key benefit of privately issued money is that these private liabilities can intermediate productive assets, much as the deposit liabilities of private banks do. Economic efficiency is enhanced if private money is permitted, as this private money is backed by productive investment, which ultimately enhances production and welfare. If private money is banned, then circulating currency takes the form of barren, unbacked fiat money. However, one cost of having circulating private money is that it can be more easy to counterfeit than fiat money. If counterfeiting is not very difficult, then it can have negative ramifications for social welfare.

I explore these issues here using a search model of money. Early versions of these monetary search models were developed by Kiyotaki and Wright (1989, 1993), with later developments by Trejos and Wright (1995) and Shi (1995). Some of the ideas in this article are closely related to those in Williamson (1999) and Temzelides and Williamson (2001b). In a monetary search model, economic agents typically find it difficult to get together to trade and make transactions, and there are limits on the flows of information. These are frictions which the use of money can help to overcome in the model—and in reality.

My first step will be to investigate the properties of the model when counterfeiting is not possible. I will show that government-supplied fiat money in this context is always detrimental to social welfare. Fiat money displaces private money, resulting in less investment and production and in lower welfare. If the counterfeiting of private money is possible, then with the cost of counterfeiting sufficiently low, monetary exchange may be supported only if private money is prohibited.

In Section 1 I construct the basic model, and first assume that there is no opportunity for the issue of counterfeits. I then study the nature of equilibrium in this model and show that, in the absence of counterfeiting, it is inefficient for fiat money to circulate. In the second section, I permit the costly issue of counterfeit private money and show that the potential for counterfeiting can lead to a classic type of market failure, much as in the "lemons" model of Akerlof (1971).

1. A MONETARY SEARCH MODEL WITH PRIVATE MONEY AND FIAT MONEY

The first step will be to study how an economy works when it is possible for banks to issue private monies that can circulate as media of exchange, and where economic agents can also use fiat money in exchange. In this basic model, there are no private information frictions, and there is no potential for counterfeiting.

In basic search models of money, three key assumptions are usually made: (i) people have difficulty meeting to carry on economic exchange, (ii) there is randomness involved in how people make contact, and (iii) people have limited information about what others are doing. These three assumptions capture important elements of real-world economic activity and economic exchange that help explain the existence and use of money in developed economies.

Assumption (i) is realistic, as it is clearly costly in terms of time and resources for people to get together to trade goods and assets. Shopping takes time, and it is impossible to be in two places at once. While internet shopping has reduced shopping costs dramatically, physical goods are costly to ship and cannot be delivered immediately. Assets appear to be less costly to trade than goods, as trade in assets typically involves only a change in an electronic record of ownership. For example, to trade shares on the New York Stock Exchange, one needs only to communicate with a broker. However, asset exchanges are still costly, and people do not have access to the communication technology required to make some kinds of asset trades at all times and places.

Assumption (ii) is probably the least realistic of the above three assumptions, for most individuals exercise much thought and planning in determining when they will trade goods and assets. For example, an individual's food purchases might involve no randomness at all. He or she plans to visit the supermarket regularly on a particular day of the week, draws up a shopping list, fills it at the supermarket, and returns home. However, people often find themselves in circumstances where they need to make unplanned purchases or sales of goods and assets. A car might break down, requiring one to hire a tow truck and rent a car; an unexpected illness might require that a person sell some stocks and bonds from his or her asset portfolio; a person might find that the supermarket has stocked some unusual food that he or she has a strong preference for and make an unanticipated purchase. Assumption (iii) is clearly realistic, for it is impossible to know all the intricate details of the economic interactions of all the people living in one's own city or town, let alone of all the people living in the world.

I will now describe the model environment, which will consist of a description of the population of economic agents, the preferences of these agents, the available technology, the endowments that are available to produce goods satisfying agents' preferences, and how economic agents can interact. There is a continuum of economic agents, having unit mass, and each agent has preferences given by

$$E_0 \sum_{t=0}^{\infty} \left(\frac{1}{1+r} \right)^t [\theta_t u(c_t) - x_t],$$

where E_0 is the expectation operator, conditional on information available to the agent at time 0, r is the subjective discount rate with $r > 0$, θ_t is an independent and identically distributed preference shock with $\Pr[\theta_t = 1] = \Pr[\theta_t = 0] = \frac{1}{2}$, c_t is consumption in period t, $u(\cdot)$ is a utility function that is strictly concave and strictly increasing with $u'(0) = \infty$, and $u(0) = 0$. For convenience I will assume that there exists some $\hat{q}$ such that $u(\hat{q}) - \hat{q} = 0$. Also, x_t denotes production of goods, so that an agent suffers disutility from producing. Thus, a given economic agent will wish to consume in some periods and not in others, with the desire to consume determined at random. An economic agent can in any period produce goods, at a cost in terms of disutility. These goods can be consumed by someone else or can be used as an input to an investment technology. These goods are otherwise perishable, and an economic agent cannot consume his or her own output.

There are two sectors in the economy, the *search sector*, where economic agents are randomly matched and can trade with each other, and the *banking sector*, where an economic agent engages in banking transactions. An agent has no choice about which sector to visit during a given period, in that with probability π he or she visits the search sector, and with probability $1 - \pi$ he or she visits the banking sector. We have $0 < \pi < 1$.

If an agent is in the banking sector at the beginning of the period, that agent has the opportunity to fund an investment project. That is, an investment project is indivisible and requires γ units of goods to initiate, where $\gamma > 0$. We will call the agent's claim to the investment project a bank note, or private money. This claim is indivisible and portable at no cost, but a given agent can carry at most one bank note in inventory. At any time in the future, the bank note can be redeemed. An agent who returns to the banking sector with the bank note can interrupt the investment project and receive a return of R units of consumption goods, which must then be consumed. The payoff R is independent of when the project is interrupted, and once the project is interrupted it will yield no more payoffs. Assume that $u(R) - \gamma > 0$. One can think of the "bank" here as a machine that yields an invisible bank note if γ units of goods are inserted in it. At any period in the future, the bank note can be inserted in the bank machine, in which case the bank machine will yield R units of consumption goods.

Bank notes in the model are intended to capture some important features of the bank liabilities that circulated in the United States before the Civil War or in Canada prior to 1935. In these historical monetary regimes, bank notes circulated hand-to-hand, and they were ultimately redeemable (typically in gold or silver) at the bank of issue. The redemption value of circulating bank notes did not depend on the length of time between the issue of the note and its redemption. To keep things simple, in the model some of the features of historical bank note issue are assumed as part of the technology, but this assumption is not important for my argument.

In period 0, a fraction M of the population is endowed with one unit each of fiat money. Fiat money is assumed to be an intrinsically worthless and indivisible object. Agents can hold at most one unit of some object, so in equilibrium a given agent will be holding either one bank note, one unit of fiat money, or nothing. The assumption that assets are indivisible is common in search models of money, and this assumption is made for tractability. If assets were divisible and it were feasible for a given agent to hold any nonnegative quantity of a particular asset, then we would have to track the entire distribution of assets across the population over time. In general, this would make the model difficult to work with.[1] Given the assumption of indivisibility of assets and the constraint that any given agent can hold only one unit of some asset, we need only keep track of the fraction of agents in the population holding each asset at each point in time.

If an agent is in the search sector at the beginning of the period, he or she is matched with one other agent for the period. These matches are random except that, for analytical convenience, every match is between an agent who wishes to consume ($\theta_t = 1$) and one who does not wish to consume ($\theta_t = 0$). Thus, I have ruled out matches where there is a double coincidence of wants and both agents in the match wish to consume during the period, and where neither agent wishes to consume. In order for exchange to take place in any of the single-coincidence-of-wants matches, it must be the case that the agent who wishes to consume has an asset (either a bank note or money) and the agent who does not wish to consume has no asset. Clearly, if the agent who wishes to consume has no asset, he or she has nothing to offer in exchange for the other agent's output, and if the agent who does not wish to consume already has an asset, he or she will not accept more assets as he or she is not able to carry them into the future.

One important assumption is that an agent cannot make contact with other agents visiting the banking sector at the same time. That is, suppose that agents arrive at the banking machine sequentially during the period. This prevents agents from making trades while in the banking sector, which simplifies the model. A second important assumption is that agents meeting in the search sector know nothing about others' trading histories. With knowledge of trading histories, it would be possible to support certain types of credit arrangements, which we can think of as being similar to centralized credit card networks. Such credit arrangements are studied in Aiyagari and Williamson (1999, 2000),

---

[1] See, however, Green and Zhou (1998), Lagos and Wright (2000), and Shi (1997), where search models with divisibile money are constructed.

Williamson (1999), Temzelides and Williamson (2001a), and Kocherlakota and Wallace (1998). Thus, the model environment rules out credit, which makes it simpler to focus attention on the monetary arrangements of interest here.

Equilibrium

I will confine my attention here to steady state equilibria, where prices and the distribution of assets across the population are constant over time. There may exist other equilibria in this model, such as deterministic equilibria with cycles and stochastic sunspot equilibria. However, these other equilibria are more difficult to analyze, and our points can be made in a more straightforward way by studying only steady states. The distribution of assets across agents in a steady state is described by (ρ_0, ρ_p, ρ_m), where ρ_0 denotes the fraction of agents in the population holding no asset, ρ_p is the fraction of agents holding bank notes, and ρ_m is the fraction holding fiat money. We have $\rho_0+\rho_p+\rho_m = 1$. Next, q_p is the price of a bank note in terms of consumption goods. That is, q_p is the quantity of consumption goods that an agent gives up in equilibrium for a bank note. Similarly, q_m denotes the price of fiat money. The other variables we will need to determine are V_0, the expected utility at the end of the period associated with holding no asset, or the *value* to holding nothing, and V_p and V_m, the values to holding a bank note and fiat money, respectively.

Dynamic optimization by the economic agents in the model implies a set of *Bellman equations*. First, V_0 is determined by

$$V_0 = \left(\frac{1}{1+r}\right)\left(\pi \left\{ \begin{array}{c} \left(\rho_0 + \frac{\rho_p}{2} + \frac{\rho_m}{2}\right) V_0 + \frac{\rho_p}{2} \max\left[V_p - q_p, V_0\right] \\ + \frac{\rho_m}{2} \max\left[V_m - q_m, V_0\right] \end{array} \right\} \begin{array}{c} \\ \\ +(1-\pi)\max\left[V_0, V_p - \gamma\right] \end{array} \right) \tag{1}$$

In equation (1), the value of holding no asset at the end of the current period is determined by the opportunities this represents for trading in the following period. These opportunities need to be discounted to the present using the discount rate r. In the next period, with probability π the agent will be in the search sector, and will meet an agent with nothing, with a bank note, or with fiat money, with probabilities ρ_0, ρ_p, and ρ_m, respectively. If the agent is in the search sector and meets another agent with nothing, clearly they cannot trade and the agent's value will then be V_0 at the end of the next period. However, if the agent meets someone with a bank note, trade can only take place if the other agent wishes to consume, which occurs with probability $1/2$. If the other agent wishes to consume, then the agent decides whether to trade or not based on what gives him or her the greatest utility. If he or she trades, then q_p goods must be produced at a utility cost of q_p, and the agent receives the bank note, with an associated value V_p and a net expected utility gain of

$V_p - q_p$. However, should the agent not trade, value will remain the same at V_0. Similarly, if the agent meets someone with money, trade will occur only if the other agent wishes to consume, and the agent will trade if $V_m - q_m > V_0$, and will not trade if $V_m - q_m < V_0$. Now, if the agent is in the banking sector, which occurs with probability $1 - \pi$, then he or she can choose to do nothing, which yields a value at the end of the next period of V_0, or a bank note could be purchased, yielding expected utility $V_p - \gamma$.

It is somewhat simpler to rewrite the Bellman equation (1) by multiplying both sides by $1 + r$, and then subtracting V_0 from the left and right sides to obtain

$$\begin{aligned} rV_0 &= \frac{\pi\rho_p}{2}\max\left[V_p - q_p - V_0, 0\right] + \frac{\pi\rho_m}{2}\max\left[V_m - q_m - V_0, 0\right] \\ &\quad +(1-\pi)\max\left[0, V_p - \gamma - V_0\right]. \end{aligned} \tag{2}$$

In equation (2), the right-hand side is the net expected flow return that can be obtained when no asset is held at the beginning of the period. In a manner similar to equation (2), for agents holding bank notes and fiat money, respectively, we have

$$\begin{aligned} rV_p &= \frac{\pi\rho_0}{2}\max\left[u(q_p) + V_0 - V_p, 0\right] \\ &\quad +\frac{(1-\pi)}{2}[u(R) + \max\left(-\gamma, V_0 - V_p\right)], \end{aligned} \tag{3}$$

and

$$rV_m = \frac{\pi\rho_0}{2}\max\left[u(q_m) + V_0 - V_m, 0\right]. \tag{4}$$

Note in equation (3) that in the second term on the right-hand side, the holder of the bank note redeems it in the banking sector only in the case where he or she wishes to consume. Otherwise, it is preferable to continue to hold the note so that it can be traded away or redeemed in the future. In equation (4), the holder of fiat money obtains a return only in the search sector when he or she meets an agent who holds no asset and wishes to consume.

Next, we need to describe how prices are determined in trades between asset holders and those not holding assets. In general, two agents who can potentially trade have a bargaining problem to solve, and the literature has approached bargaining problems of this nature in a variety of ways including using a Nash bargaining solution or a Rubinstein bargaining game (see Trejos and Wright [1995]). Here, I will follow the simplest possible approach, which is to assume that the asset holder has all of the bargaining power and makes a take-it-or-leave-it offer to the agent who holds no asset. That is, the asset holder sets the price for the exchange in such a way that the other agent is just indifferent between accepting the offer and declining. This gives

$$V_p - q_p - V_0 = 0, \tag{5}$$

and

$$V_m - q_m - V_0 = 0. \tag{6}$$

Equilibrium Where Bank Notes and Fiat Money Circulate

We will first examine an equilibrium where bank notes and fiat money are exchanged for goods in the search sector. Since some agents hold bank notes and some agents hold no assets in such a steady state equilibrium, then when the holder of a bank note redeems that note in the banking sector, he or she must be indifferent between acquiring another bank note and holding no asset. If this were not the case, then either the steady state supply of bank notes would be zero, or there would be no agents in the search sector with no assets, so bank notes could not be used in exchange. We then have

$$V_p - \gamma = V_0. \tag{7}$$

Equations (2), (5), (6), and (7), then, imply that $V_0 = 0$. That is, the value of holding no asset is zero, since an agent with no asset then receives no surplus from trading in the search sector with asset holders, and his or her value will not change when visiting the banking sector. Given this, equation (7) implies that $V_p = \gamma$, and (5) and (6) imply, respectively, that $V_p = q_p = \gamma$ and $V_m = q_m$.

Now, we will assume that $u(\gamma) - \gamma > 0$, or $\gamma < \hat{q}$, which implies from (3) that the holder of a bank note is willing to trade with an agent holding no asset. Since the equilibrium price of a bank note is γ, the constraint $u(\gamma) - \gamma > 0$ states simply that there is a positive surplus associated with the exchange of a bank note. Then, (3) implies that

$$r\gamma = \frac{\pi\rho_0}{2}[u(\gamma) - \gamma] + \frac{(1-\pi)}{2}[u(R) - \gamma], \tag{8}$$

and (8) is then an equation that solves for ρ_0, that is,

$$\rho_0 = \frac{2r\gamma - (1-\pi)[u(R) - \gamma]}{\pi[u(\gamma) - \gamma]}. \tag{9}$$

Since $V_m = q_m$ in equilibrium, we can substitute for V_m in equation (4), and for now we can conjecture that it will always be in the interest of a holder of fiat money to trade for goods at the price q_m. From (4), these steps then give us

$$rq_m = \frac{\pi\rho_0}{2}\left[u(q_m) - q_m\right]. \tag{10}$$

Equation (10) then solves for q_m given the solution for ρ_0 from (9). There are two solutions to (10), one where $q_m = 0$, and one where $q_m > 0$. The equilibrium where $q_m = 0$ is uninteresting since the value of holding fiat money is zero, and nothing can ever be purchased with fiat money. However, an agent holding no asset is willing to accept fiat money as that would not make him or her any worse off. We will confine our attention to the equilibrium where $q_m > 0$. Given this condition, from (10) we have $u(q_m) - q_m > 0$, and our conjecture that the holder of fiat money is always willing to trade in equilibrium is correct. An important result is that, from (8) and (10),

$$q_m < q_p = \gamma, \tag{11}$$

that is, private bank notes exchange for goods in the search sector at a premium over fiat money. This result follows because bank notes have a redemption value in the banking sector, while fiat money does not. Therefore, agents are willing to pay more for the possibility of this higher future payoff.

Now, in the equilibrium we are examining where $q_m = V_m > 0$, when a holder of fiat money goes to the banking sector, he or she will not want to acquire a bank note. Holding fiat money has strictly positive value, while acquiring a bank note implies net expected utility $V_p - \gamma = \gamma - \gamma = 0$. Thus, in a steady state, no one would want to dispose of fiat money balances. This need not necessarily imply that it would never be in anyone's interest to dispose of money along the path the economy takes from the first date to the steady state. However, suppose that there is no fiat money in existence, that the economy converges to a steady state, and that fiat money enters the economy when the central bank chooses holders of bank notes at random in the search sector and replaces each of their bank notes with one unit of fiat money. This action will have no effect on the equilibrium, other than to replace bank notes with fiat money one-for-one, and from the date when money was injected, no one would dispose of fiat money. Therefore, in this sense we can take $\rho_m = M$ in equilibrium, so that the fraction of money holders in the steady state is equal to the quantity of money injected by the central bank. Since $\rho_0 + \rho_p + \rho_m = 1$ in equilibrium, from our solution for ρ_0 in equation (9) we require that $0 < \rho_0 < 1 - M$, which implies

$$0 < 2r\gamma - (1-\pi)[u(R) - \gamma] < (1-M)\pi[u(\gamma) - \gamma] \tag{12}$$

From (12), to support an equilibrium where private bank notes circulate, the return on investment, R, cannot be too large or too small. If R is too small, then investment will not be worthwhile, and no one would be willing to hold bank notes. However, if R is too large, then the redemption value of a bank note will be sufficiently attractive that no one will want to trade away a bank note for goods in the search sector.

Note that if $M = 1$, then (12) does not hold for any values of γ, r, and π, since the upper and lower bounds on $2r\gamma - (1-\pi)[u(R) - \gamma]$ in (12) are then identical. Therefore, there is always some value for M that is sufficiently large that (12) is not satisfied (note that the upper bound decreases with M), and an equilibrium with circulating bank notes and valued fiat money does not exist.

Equilibrium Where Only Fiat Money Circulates

If fiat money circulates and there are no bank notes, then we have $\rho_m = M$, $\rho_0 = 1 - M$, and $\rho_p = 0$. From (6) and (2) we have $V_0 = 0$ and $V_m = q_m$, so equation (4) then implies that q_m is the solution to

$$rq_m = \frac{\pi(1-M)}{2}\left[u(q_m) - q_m\right]. \tag{13}$$

Just as in the previous subsection, I will ignore the equilibrium where $q_m = 0$ and focus on the solution to (13) where $q_m > 0$. Then, from equation (13), $u(q_m) - q_m > 0$, so it will always be in the interest of an agent with fiat money to trade it for goods, as I have implicitly conjectured. In an equilibrium where only fiat money circulates, it cannot be in the interest of any agent to hold a bank note. Were an agent to have a bank note, we would have $V_p = q_p$, from (5), and q_p would be determined, from (3), by

$$rq_p = \frac{\pi(1-M)}{2}[u(q_p) - q_p] + \frac{(1-\pi)}{2}[u(R) - q_p]. \tag{14}$$

Then, for it not to be in the interest of an agent to acquire a bank note, it must be the case that $V_p - \gamma \leq 0$, so that an agent prefers to hold no asset rather than acquiring a bank note in the banking sector. This inequality then implies that $q_p \leq \gamma$ or, from (14),

$$2r\gamma - (1-\pi)[u(R) - \gamma] \geq (1-M)\pi[u(\gamma) - \gamma] \tag{15}$$

Now, defining

$$\phi \equiv 2r\gamma - (1-\pi)[u(R) - \gamma], \tag{16}$$

we can conclude from (12) and (15) that an equilibrium exists where bank notes and fiat money circulate if

$$0 < \phi < (1-M)\pi[u(\gamma) - \gamma], \tag{17}$$

and that an equilibrium where only fiat money circulates as a medium of exchange exists if

$$\phi \geq (1-M)\pi[u(\gamma) - \gamma]. \tag{18}$$

Inequalities (17) and (18) imply that we are more likely to see bank notes in circulation as the redemption value of a bank note, R, increases (though recall that this redemption value cannot be too large, as we require $\phi > 0$), and that bank notes are less likely to circulate the larger is M, the quantity of fiat money in circulation. With an increase in the redemption value of a bank note, agents are much more willing to acquire notes to be used in exchange. Fiat money displaces private money in circulation, so if the quantity of fiat money is sufficiently large, then private money is driven out of the system.

Note that, if $\phi \leq 0$, then a steady state equilibrium will exist where agents acquire private money, but this private money is not exchanged in the search sector. Private money is then just held until redemption occurs. As I am primarily interested here in the medium of exchange role of private money, I will assume throughout that $\phi > 0$.

Is It Efficient for Fiat Money to Circulate?

In the equilibrium studied above where private bank notes and fiat money both circulate, an increase in M, the stock of money in circulation, will displace bank notes one-for-one. That is, since ρ_0, the fraction of agents in the population holding no assets, is determined by (9), which does not depend on M, a change in M can only affect the fractions of agents holding fiat money and bank notes. My interest in this section is in determining the welfare effects of changes in M. Is it a good thing for government-supplied fiat money to replace circulating bank notes?

To evaluate changes in welfare for this economy, I will use a welfare criterion of average expected utility across the population in the steady state. Letting W denote aggregate welfare, we have

$$W = \rho_0 V_0 + \rho_p V_p + \rho_m V_m.$$

That is, aggregate welfare is just the weighted average of values (expected utilities) across agents in the steady state. From above, in a steady state equilibrium where bank notes and fiat money circulate, we have $V_0 = 0$, $\rho_p = 1 - \rho_0 - M$, $V_p = \gamma$, $\rho_m = M$, and $V_m = q_m$, where $q_m < \gamma$. These conditions give

$$W = (1 - \rho_0 - M)\gamma + Mq_m.$$

But since $\gamma - q_m > 0$, an increase in M causes W to decrease, so welfare falls as more fiat money is introduced. Ultimately, if M becomes sufficiently large, then bank notes are driven out of the economy altogether. I can show (with some work) that, no matter what M is, welfare cannot be higher when only fiat money circulates than in an equilibrium where bank notes circulate.

The key result here is that fiat money always reduces welfare, which is true because circulating bank notes serve two roles. First, bank notes serve as a medium of exchange and therefore enhance welfare by allowing for production and consumption in the search sector that would otherwise not take place. Second, bank notes support productive investment. The value of bank notes as a medium of exchange encourages agents to hold these assets, and as a result there is productive intermediation activity. Fiat money serves only the medium of exchange function and does nothing to promote private investment. Thus, if fiat money replaces circulating bank notes, then an asset that performs only a medium of exchange function is replacing another asset; this other asset serves as a medium of exchange but also performs an important secondary role in promoting productive investment.

Prior to 1935, the assets used to back the circulating notes that Canadian banks issued were essentially unrestricted. In fact, the issue of circulating notes largely financed bank loans in Canada. This was certainly not true in the United States prior to the Civil War, where notes were issued by state-chartered banks and were typically required to be backed by state bonds. Thus, we can think of private bank notes as financing public investment in the United States and private investment in Canada. In either case, our model captures some elements of the historical role of private money.

The view of private money from the model as I have laid it out thus far is perhaps too sanguine. In practice some private money systems appear to have worked poorly, while others have done quite well. In particular, the monetary system in place prior to the Civil War in the United States certainly appears to have worked poorly, though this is the subject of some debate (Rolnick et al. 1997; Rolnick and Weber 1983, 1984). Indeed, the introduction of the National Banking System in the United States in 1863 and the contemporaneous introduction of a prohibitive tax on state bank notes appears to have been motivated in good part by the view that the existing system of private issue of bank notes by state-chartered banks was inefficient. However, the monetary system in Canada prior to 1935 seems to have been successful in that the notes issued by chartered banks were essentially universally accepted at par and there were only a few unusual circumstances of banks defaulting on their notes (in the United States prior to the Civil War, there was widespread discounting of private bank notes and there were many instances of default on private bank notes).

Three incentive problems are the primary causes of potential inefficiencies in private money systems. First, there might be an "overissue" problem, as discussed by Friedman (1960). That is, if there are many issuers of private money behaving competitively, they will tend to issue notes to the point where they collectively drive the value of private money to zero. Cavalcanti, Erosa, and Temzelides (1999) show, however, that each bank in a private money system (such as the Suffolk Banking System in pre–Civil War New England or the Canadian banking system prior to 1935) could have sufficient incentives to prevent overissue. A key element in these systems was that each private money issuer accepted the notes of other private money issuers for redemption.

The second type of incentive problem arises because private money producers might sell lemons (see Akerlof [1971]). Williamson (1991) shows that if banks differ according to the quality of their asset portfolios, and the holders of bank notes have difficulty distinguishing quality, then the market could be dominated by low-quality private bank notes that bear a low rate of return on redemption. In these circumstances, it is possible that a prohibition on private bank notes, with fiat money circulating as the sole medium of exchange, would be the most efficient monetary arrangement. While lemons problems probably created serious inefficiencies in the United States prior to the Civil

War, these problems appear to have been largely solved in the Suffolk Banking System and in Canada prior to 1935. These two private money systems had key self-regulatory mechanisms that helped prevent lemons problems; furthermore, the Canadian system had an advantageously small number of private note issuers.

The third type of incentive problem in private money systems is the potential issue of counterfeits. In terms of its function as a medium of exchange, a counterfeit is much like a lemon of extremely poor quality. If sufficient care is put into its production, the counterfeit will pass undetected in many circumstances as a medium of exchange, but in contrast to a genuine bank note it has no redemption value. While we might view Williamson (1991) as applying to counterfeiting as well as to poor-quality banking, no one has analyzed the counterfeiting problem in the context of a monetary search model. Thus, exploring the implications of counterfeiting in our model in the next section will prove useful.

2. A MODEL WITH COUNTERFEITING

One potential problem with a private money system is that this money may be counterfeited. Indeed, the counterfeiting of private bank notes appears to have been common in the United States prior to the Civil War. Clearly, government-issued fiat currency is also subject to counterfeiting, but there may exist economies of scale in counterfeit-prevention technologies and in the enforcement of counterfeiting laws. Thus, the modifications I make in the model will include the assumptions that private bank notes can be counterfeited at some cost and that (for simplicity) the cost of counterfeiting fiat money is infinite. There will be a tradeoff, then, between the benefits from the circulation of private money—the promotion of productive investment—and the costs of private money—the promotion of inefficient counterfeiting. I will show that there are circumstances in which the possibility of counterfeiting fundamentally changes the nature of the equilibria that can exist. Indeed, a ban on private money may be necessary to support a stationary equilibrium with monetary exchange.

I will assume that a counterfeit bank note can be created when an agent is in the banking sector, at a cost δ in units of goods, where $0 < \delta < \gamma$, so that it is more costly to produce a genuine bank note than a counterfeit. This counterfeit note can potentially be exchanged for goods in the search sector, but there is no investment project backing the note, and so it has no redemption value. In meetings with other agents in the search sector, a counterfeit note can be detected with probability η, if it is offered in exchange, where $0 < \eta < 1$, but otherwise the counterfeit goes undetected and is indistinguishable from a private bank note. If a counterfeit is detected, then it is confiscated and destroyed. We will let ρ_f denote the fraction of agents holding counterfeit notes in equilibrium, with $\rho_0 + \rho_p + \rho_m + \rho_f = 1$.

We determine the value of holding a counterfeit, V_f, in a manner similar to (2), (3), and (4), as

$$rV_f = \frac{\pi \rho_0 (1-\eta)}{2} \max\left[u(q_u) + V_0 - V_f, 0\right] + \frac{\pi \rho_0 \eta}{2}(V_0 - V_f). \quad (19)$$

In equation (19), note in the first term on the right-hand side that if the counterfeit goes undetected, it sells at the same price as a private bank note which cannot be identified—that is, at the price q_u, which is the price of a bank note of unidentified quality. Also note, in the second term, that if detection takes place in the search sector, the note is confiscated and the agent will have value V_0 at the end of the period. I assume for convenience that counterfeits can always be recognized in the banking sector. Thus, an agent with a counterfeit bank note arriving in the banking sector will hide the counterfeit and not attempt to redeem it.

We also need to modify equation (2), since agents with no asset can encounter an agent with a counterfeit note with whom they might trade. We have

$$\begin{aligned} rV_0 \quad = \quad & \frac{\pi(1-\eta)(\rho_p + \rho_f)}{2} \max\left[\frac{\rho_p V_p + \rho_f V_f}{\rho_p + \rho_f} - q_u - V_0, 0\right] \\ & + \frac{\pi \eta \rho_p}{2} \max\left[V_p - q_p - V_0, 0\right] \\ & + \frac{\pi \rho_m}{2} \max\left[V_m - q_m - V_0, 0\right] \\ & + (1-\pi) \max\left[0, V_p - \gamma - V_0, V_f - \delta - V_0\right]. \end{aligned} \quad (20)$$

In the first term on the right-hand side of equation (20), the agent sometimes cannot distinguish between a genuine bank note offered in exchange and a counterfeit. In this circumstance, I assume that if the agent accepts the note, he or she learns before the end of the period whether or not it is a counterfeit. Whether the note is accepted depends on the expected value of the note to the agent. In the second term on the right-hand side, the agent has encountered an agent with a bank note and has been able to verify that it is not a counterfeit. The third term on the right-hand side of equation (20) takes account of the agent's opportunity to produce a counterfeit bank note when in the banking sector. The analogue of equation (3) is

$$\begin{aligned} rV_p \quad = \quad & \frac{\pi \rho_0 \eta}{2} \max\left[u(q_p) + V_0 - V_p, 0\right] \\ & + \frac{\pi \rho_0 (1-\eta)}{2} \max\left[u(q_u) + V_0 - V_p, 0\right] \\ & + \frac{(1-\pi)}{2}\left[u(R) + \max\left(-\gamma, V_0 - V_p, V_f - \delta - V_p\right)\right]. \end{aligned} \quad (21)$$

Equation (21) takes account of the fact that an agent with a bank note can potentially encounter agents with assets who both recognize and do not recognize his or her bank note as not being a counterfeit; it also accounts for the fact that the agent can create a counterfeit in the banking sector when a bank note is redeemed. Equation (4) remains the same.

In trades where the holder of a bank note or counterfeit makes a take-it-or-leave-it offer to an agent holding no asset who does not recognize the quality of the asset, we obtain

$$\frac{\rho_p V_p + \rho_f V_f}{\rho_p + \rho_f} - q_u - V_0 = 0, \tag{22}$$

while equations (5) and (6) continue to hold.

Equilibrium

Given the possibility that counterfeit bank notes will be issued, an equilibrium can be of three types. First, there could be an equilibrium where fiat money and bank notes circulate, but where it is in no one's interest to issue a counterfeit. Second, it could be that only fiat money circulates as a medium of exchange, with no bank notes in circulation, and therefore with no opportunities for circulating counterfeits. Third, fiat money, bank notes, and counterfeits could all circulate in equilibrium. We will consider each of these possibilities in turn.

Bank Notes and Fiat Money Circulate, with No Counterfeiting

This equilibrium is similar in most respects to that considered in the previous section, where bank notes and fiat money circulate but there are no opportunities to issue counterfeits. That is, equations (7), (8), (9), (10), and (12) all hold. Here, given that there are opportunities to counterfeit, it cannot be in the economic interest of anyone to issue a counterfeit in equilibrium.

If a counterfeit were issued, from (19) and (22) its value V_f would be determined by

$$2rV_f = \frac{\phi}{[u(\gamma) - \gamma]}[(1-\eta)u(\gamma) - V_f], \tag{23}$$

where ϕ is defined as in (16). That is, if a counterfeit were issued, it would be negligible relative to the quantity of notes in circulation, and it would trade at the price $q_u = \gamma$ so long as it went undetected when offered in exchange. For it not to be in the interest of an agent to issue a counterfeit when in the banking sector, we must have $V_f \leq \delta$, which gives, from (23),

$$\phi[(1-\eta)u(\gamma) - \delta] \leq 2r\delta[u(\gamma) - \gamma]. \tag{24}$$

Thus, since $\phi > 0$, (24) essentially states that the cost of counterfeiting, δ, must be sufficiently large and the probability of detection η must also be sufficiently large for this equilibrium to exist. However, note that even if $\eta = 0$ and no counterfeits can be detected in use, (24) will hold if δ is sufficiently large. It is also true that, for any $\delta > 0$ there is some sufficiently large value for η such that (24) will hold. That is, a sufficiently high detection probability will discourage counterfeits no matter how cheap they are to produce.

Another interpretation of condition (24) is the following. Suppose that the economy is in a steady state equilibrium with circulating private money and fiat money and an infinite cost of producing counterfeits. Then suppose that there was an unanticipated innovation to the counterfeiting technology that reduced δ so that condition (24) did not hold. It would then be in the interest of agents to issue counterfeits, which would upset the steady state equilibrium.

Only Fiat Money Circulates

When counterfeiting is possible, a potential outcome is that private money is not issued in equilibrium, and only fiat money circulates as a medium of exchange. Counterfeits would always be identifiable in such an equilibrium, since there would be no private money in circulation, and it would then be an equilibrium for no one to accept counterfeits. This equilibrium will be identical in all respects to the one considered in the previous section, where counterfeiting was not possible. Thus, for this equilibrium to exist, condition (18) must hold.

Bank Notes, Fiat Money, and Counterfeits Circulate

In the final case I consider, bank notes, fiat money, and counterfeits are all exchanged for goods in the search sector. This is the most complicated of the three cases to analyze.

Here, since an agent who holds no asset never receives any surplus in trading, we will have $V_0 = 0$. Also, when an agent is in the banking sector and is not holding an asset, then he or she must be indifferent among the following choices: continue to hold no asset; acquire a bank note; acquire a counterfeit. That is, since bank notes are continually retired through redemption, and some counterfeits are detected each period and removed from circulation, there must be a flow of new bank notes and counterfeits each period. Further, some agents must wish not to hold assets in the steady state, otherwise there will be no goods offered for sale in the search sector. Thus, it must be the case that agents are indifferent among the above three options in the steady state. Thus, we must have $V_0 = V_p - \gamma = V_f - \delta$. Then, given that $V_0 = 0$, we have $V_p = \gamma$ and $V_f = \delta$. This in turn implies, from (5), (6), and (22), that $q_p = \gamma$, $q_m = V_m$, and

$$q_u = \frac{\rho_p \gamma + \rho_f \delta}{\rho_p + \rho_f}. \tag{25}$$

From (19) and (21) we obtain, respectively,

$$2r\delta = \pi \rho_0 \left[(1 - \eta) u(q_u) - \delta\right], \tag{26}$$

and

$$2r\gamma = \pi \rho_0 \left[\eta u(\gamma) + (1 - \eta) u(q_u) - \gamma\right] + (1 - \pi)[u(R) - \gamma]. \tag{27}$$

Then, equations (26) and (27) solve for ρ_0 and q_u, and since $\rho_p + \rho_f + \rho_m = 1 - \rho_0$, and $\rho_m = M$, then given a solution for q_u, we can use (25) to solve for ρ_p and ρ_f. Solving (26) and (27) for ρ_0 and $u(q_u)$, we obtain

$$\rho_0 = \frac{\phi - 2r\delta}{\pi[\eta u(\gamma) - \gamma + \delta]}, \tag{28}$$

$$u(q_u) = \frac{\delta\phi - 2r\delta(1 - \eta) u(\gamma)}{(\phi - 2r\delta)(1 - \eta)}. \tag{29}$$

Now, we require that $0 < \rho_0 < 1 - M$, or, from (28),

$$2r\delta < \phi < \pi(1 - M)[\eta u(\gamma) - \gamma + \delta] + 2r\delta. \tag{30}$$

Also, (25) implies that $u(q_u) < u(\gamma)$ and $u(q_u) > u(\delta)$ so, respectively, from (29), we must have

$$\phi[(1 - \eta) u(\gamma) - \delta] > 0, \tag{31}$$

and

$$\phi < \frac{2r\delta[u(\delta) - u(\gamma)]}{u(\delta) - \delta}. \tag{32}$$

But (30) then implies that $\phi > 0$, since $\delta > 0$, and (32) implies $\phi < 0$, since $\delta < \gamma$ and $u(\gamma) - \gamma > 0$. This resulting contradiction tells us that this type of equilibrium cannot exist.

A Prohibition on Private Bank Notes

Suppose now that the government can prohibit the issue of private bank notes. That is, assume that the government has the ability to monitor the production of bank notes, but is not able to monitor the production of counterfeits. Then, if there is a public prohibition on the production of private bank notes, everyone knows that a note offered in exchange is a counterfeit. There is then a steady state equilibrium in which counterfeits are never accepted in exchange, but fiat money is.

In this equilibrium, the price obtained for fiat money in exchange, q_m, is determined by equation (13). In contrast to the equilibrium where only fiat money circulates but private bank note issue is permitted, we do not require that condition (18) holds here for an equilibrium to exist. That is, an equilibrium where fiat money circulates under the prohibition of private money exists for all parameter values.

Existence of a Stationary Equilibrium

I have now determined that, if an equilibrium exists, it must either be one where bank notes and fiat money circulate and counterfeits do not, or where only fiat money circulates. Further, there are restrictive conditions under which fiat money will circulate when private money issue is permitted, and an equilibrium where fiat money circulates always exists when private money issue is prohibited. In this section I want to explore the possibilities for existence of stationary equilibria under counterfeiting and what they mean for the design of monetary systems.

For an equilibrium with the coexistence of fiat money and bank notes, we know from above that conditions (17) and (24) must hold. Alternatively, for an equilibrium with fiat money only, when private money issue is permitted, condition (18) must hold. Now, if

$$\delta \geq \frac{(1-M)\pi(1-\eta)u(\gamma)}{2r+(1-M)\pi}, \tag{33}$$

then if (17) holds, so does (24); under these circumstances the potential issue of counterfeits is irrelevant. That is, if counterfeiting is sufficiently costly, as defined by (33), then so long as $\phi > 0$, a stationary equilibrium with monetary exchange exists where either private money and fiat money circulate (if (17) holds) or where only fiat money circulates (if (18) holds).

Now, if

$$\delta < \frac{(1-M)\pi(1-\eta)u(\gamma)}{2r+(1-M)\pi}, \tag{34}$$

and

$$\phi \in \left(\frac{2r\delta[u(\gamma)-\gamma]}{(1-\eta)u(\gamma)-\delta}, (1-M)\pi[u(\gamma)-\gamma] \right), \tag{35}$$

then no equilibrium exists where private money issue is permitted. Under these circumstances, where the cost of counterfeiting is sufficiently small, a stationary equilibrium with monetary exchange can only be supported if there is a prohibition on private money issue.

Thus, the potential for counterfeiting makes a key difference here for the effects of government intervention in the issue of media of exchange. Without the possibility of counterfeiting, the circulation of private money is unambiguously good for economic welfare. If private money were banned under these circumstances, welfare would decrease, and even the introduction of more government-supplied fiat money into the economy would be detrimental as it inefficiently displaces private money. However, if counterfeiting is possible at a sufficiently low cost, then there can be a classic market failure of the type that can occur in the lemons model of Akerlof (1971). That is, there cannot be an equilibrium where private money and counterfeits coexist, but the issue of private money would induce a flood of counterfeits, so an equilibrium can only exist if there is a prohibition on the issue of private money. Monetary exchange in this case is supported with government supplied fiat money and a ban on private money.

3. CONCLUSION

I have shown, with the aid of a search model of money, some of the benefits and costs of a monetary system where private money can be issued. The issue of private money yields a social benefit in that it leads to productive financial intermediation, which can increase welfare. However, the potential for counterfeiting in this system can also lead to the possibility that monetary exchange can be supported only if private money issue is prohibited.

Though my analysis yields some interesting insights, there are important qualifications to what I have done here. First, I made a very simple assumption in the model: that fiat money could not be counterfeited, while private money could be counterfeited at a cost. While it may be the case that there exist economies of scale in monitoring for counterfeit money, which could imply the optimality of a government monopoly in currency provision, it seems unlikely that fiat money would in general be more difficult to counterfeit than private money. The cost of counterfeiting depends in part on the technology used to produce the money that the counterfeiter is trying to replicate. For example, the new Federal Reserve $20 note is much harder to counterfeit than the old one. In a world with many private money issuers, each private money issuer may invest too little in foiling counterfeiters relative to the social optimum, and it could be that some form of government intervention would correct this market failure. However, to address this issue properly would require a more complicated model with alternative private money production technologies.

Second, a key feature of the monetary search model that lends it tractability is that money is indivisible. Of course, money is certainly indivisible in practice, but the fact that we cannot divide money into denominations smaller than one cent cannot matter much. In our model, agents can carry at most one unit of money, and as a result money is not neutral, which is an undesirable property of the model. Changing the number of units of money in existence will change real variables in the model in the long run, and this was important for some of our results. In particular, we should be skeptical that the result in Section 1 that fiat money displaces private money one-for-one would hold if money were perfectly divisible. Some authors, in particular Lagos and Wright (2000) and Shi (1997), have studied tractable search models of divisible money. However, it remains to be seen whether these models have much to contribute above and beyond, for example, standard cash-in-advance models.

The model in this article can be extended to examine issues related to the clearing and settlement of private monies, as in Temzelides and Williamson (2001a,b). A more complete model of banking can be embedded in this framework, too, where the banks in the model share some of the features of banks in practice, such as diversification and the transformation of assets (see Williamson [1999]).

REFERENCES

Aiyagari, S. Rao, and Stephen Williamson. 1999. "Credit in a Random Matching Model with Private Information." *Review of Economic Dynamics* 2 (January): 36–64.

______________. 2000. "Money and Dynamic Credit Arrangements with Private Information." *Journal of Economic Theory* 91 (April): 248–79.

Akerlof, George. 1971. "The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84 (August): 488–500.

Cavalcanti, Ricardo, Andrés Erosa, and Ted Temzelides. 1999. "Private Money and Reserve Management in a Random Matching Model." *Journal of Political Economy* 107: 929–45.

Champ, Bruce, Bruce Smith, and Stephen Williamson. 1996. "Currency Elasticity and Banking Panics: Theory and Evidence." *Canadian Journal of Economics* 29: 828–64.

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

Green, Edward, and Ruilin Zhao. 1998. "A Rudimentary Model of Search with Divisible Money and Prices." *Journal of Economic Theory* 81 (August): 252–71.

Kiyotaki, Nobuhiro, and Randall Wright. 1989. "On Money as a Medium of Exchange." *Journal of Political Economy* 97: 927–54.

______________. 1993. "A Search-Theoretic Approach to Monetary Economics." *American Economic Review* 83: 63–77.

Kocherlakota, Narayana, and Neil Wallace. 1998. "Incomplete Record-Keeping and Optimal Payment Arrangements." *Journal of Economic Theory* 81 (August): 272–89.

Lacker, Jeffrey M. 1996. "Stored Value Cards: Costly Private Substitutes for Government Currency." Federal Reserve Bank of Richmond *Economic Quarterly* 82 (Summer): 1–25.

Lagos, Ricardo, and Randall Wright. 2000. "A Unified Framework for Monetary Theory and Policy Analysis." Manuscript, New York University and University of Pennsylvania (October).

Rolnick, Arthur, Bruce Smith, and Warren Weber. 1997. "The United States' Experience with State Bank Notes: Lessons for Regulating E-Cash." Manuscript, Federal Reserve Bank of Minneapolis and University of Texas-Austin.

______________, and Warren Weber. 1983. "New Evidence on the Free Banking Era." *American Economic Review* 73 (December): 1080–91.

______________. 1984. "The Causes of Free Bank Failures: A Detailed Examination." *Journal of Monetary Economics* 14 (November): 267–92.

Schuler, Kurt. 2001. "Note Issue by Banks: A Step Toward Free Banking in the United States?" *Cato Journal* 20 (Winter): 453–65.

Shi, Shouyong. 1995. "Money and Prices: a Model of Search and Bargaining." *Journal of Economic Theory* 67: 467–98.

______________. 1997. "A Divisible Search Model of Fiat Money." *Econometrica* 65 (January): 75–102.

Smith, Bruce, and Warren Weber. 1998. "Private Money Creation and the Suffolk Banking System." Federal Reserve Bank of Cleveland Working Paper 9821 (December).

Temzelides, Ted, and Stephen Williamson. 2001a. "Payments Systems in Deterministic and Private Information Environments." *Journal of Economic Theory* 99 (July): 297–326.

______________. 2001b. "Private Money, Settlement, and Discounts." *Carnegie-Rochester Conference on Public Policy* 54 (Spring): 85–108.

Trejos, Alberto, and Randall Wright. 1995. "Search, Bargaining, Money, and Prices." *Journal of Political Economy* 103: 118–41.

Williamson, Stephen. 1991. "Laissez Faire Banking and Circulating Media of Exchange." *Journal of Financial Intermediation* 2 (June): 134–67.

______________. 1999. "Private Money." *Journal of Money, Credit, and Banking* 31 (August): 469–91.

Knut Wicksell and Gustav Cassel on the Cumulative Process and the Price-Stabilizing Policy Rule

Thomas M. Humphrey

In economics as in anthropology, old artifacts spur continuing debates. A case in point is Knut Wicksell's celebrated 1898 analysis of the cumulative process of price inflation in pure credit, cashless economies. Some economists view Wicksell's model as a milestone in the evolution of quantity-theoretic monetary analysis inasmuch as it constitutes the seminal rigorous explanation of how loan-created stocks of bank money translate interest rate differentials into price level changes. Others, however, dispute this point and instead argue that money plays no role in determining price level changes in Wicksell's model.

Unfortunately, Wicksell's own writings do little to resolve the debate. Ambiguous in the extreme as to whether the cashless society version of the cumulative process has quantity-theoretic roots, his writings support quantity- and anti-quantity theory interpretations alike.

One person who could have resolved the debate was Wicksell's countryman and contemporary, the Swedish economist Gustav Cassel. In a 1928 journal article Cassel provided an extremely clear, compelling articulation of the quantity-theoretic foundations of the cumulative process. He then demonstrated the far-reaching significance of that articulation by extending it to more generalized considerations, including an analysis of the business cycle and

For valuable comments I am indebted to Beth Anderson, Kartik Athreya, Yash Mehra, Alan Rabin, and especially to John Weinberg for alerting me to the distinction between general equilibrium and dynamic disequilibrium analyses. This article expands on my paper presented to the Wicksell Chair Centennial Symposium at Trolleholm Castle, Sweden, 1–2 November 2001. The views expressed in this article are not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

alternative monetary policy rules. While Wicksell conducted monetary policy analysis using a model without money, Cassel showed that money plays a crucial, behind-the-scenes role even when excluded as a variable from the constituent equations of policy models and policy rules.

Cassel's demonstration should have made it clear that the quantity theory interpretation of the cumulative process and the operation of policy rules was correct and the anti-quantity theory interpretation was suspect. But that did not happen. Instead, Wicksell's Swedish followers largely overlooked Cassel's demonstration, perhaps because it was confined to a single published article in a foreign journal they did not ordinarily read. For whatever reason, Cassel's explanation exerted little influence and did nothing to prevent the flourishing of anti-quantity theoretic interpretations of Wicksell's work from the 1920s through the 1980s. The situation is different now. Cassel's rediscovered insights locate Wicksell's pure credit analysis of the cumulative process squarely in the quantity theory tradition. And, by stating that schema in its most precise, transparent form—not to mention extending its range of application—they spotlight the prescience, originality, and inventiveness of its creator, confirming Wicksell's place in the front rank of monetary theorists.

Wicksell's Three Contributions

Knut Wicksell's claim to fame as a monetary theorist rests on three contributions presented in his 1898 *Interest and Prices* and volume two of his 1906 *Lectures on Political Economy*. First is his concept of the hypothetical *pure credit economy,* or cashless society. In this regime all hard, or outside, money (gold coin and convertible paper currency) ceases to exist, the banking system consists of a single central bank that holds no reserves, and the medium of exchange is composed entirely of inside money, that is, checking deposits created by the central bank when it makes loans. With no reserve constraint to anchor nominal variables in the pure credit regime, deposit supply possesses potentially unlimited elasticity and the price level theoretically can rise (or fall) forever. It is the job of the central bank to prevent this outcome by means of its rate-setting policy. Such policy replaces the missing reserve constraint in imposing determinacy on an otherwise indeterminate money stock and price level.

Second is Wicksell's famous analysis of the *cumulative process* according to which price level movements stem from the differential between natural (equilibrium) and market (loan) rates of interest and continue as long as the differential persists. The rate differential is of key importance to Wicksell. It generates a gap between new capital investment and household saving, a gap that manifests itself in the form of an excess aggregate demand for goods that bids up prices cumulatively until the differential vanishes.

Wicksell's third contribution is his celebrated *feedback policy rule,* under which the central bank stabilizes the price level by adjusting its interest rate in response to price level deviations from target, stopping only when prices converge to target. A precursor of the modern Taylor rule, Wicksell's rule is the prototype of all feedback policy rules discussed in the monetary literature today.

Area of Disagreement

Wicksell scholars are in agreement on the originality, fecundity, and usefulness of these pioneering constructs. Agreement ends, however, on the role that Wicksell intended for changes in the quantity of deposit money to play in these constructs. Do bank money stock changes play an active, causal role in the transmission mechanism connecting rate differentials to price level changes? Or do they occur passively as a consequence of price level changes produced by nonmonetary means? In short, is bank money a price-determining or a price-determined variable in the workings of the cumulative process and the policy rule? Does causation run from deposits to prices as the quantity theory of money predicts? Or does it run from prices to deposit money, contrary to the quantity theory?

Active Money View

One group of scholars, including Arthur Marget (1938), Johan Myhrman (1991), Don Patinkin (1965), and Hans-Michael Trautwein (1996), contend that Wicksell saw endogenous (that is, responding to other variables in the model) changes in the stock of bank money as playing a crucial causal role. They argue that for him changes in the quantity of deposits constitute the necessary connecting link between the natural rate–market rate differential and the resulting rise in the price level. In their view, Wicksell understood that such money stock changes transform the interest differential and its associated investment-saving gap into the excess aggregate demand that bids up prices. They claim that without this monetary expansion to mediate and finance the excess demand, there would be no inflationary pressure and the rate differential would be abortive in influencing prices.

Patinkin explains how an excess of the natural over the market rate in Wicksell's pure credit economy engenders profit opportunities for investors and leads them to "increase their bank borrowings. The new demand deposits. . . placed at their disposal will enable them to increase their 'demand for goods and services as well as for raw materials already in the market for future production'" thereby raising prices (1965, 589–90). "By increasing the quantity of money in this way, the banks can bring about any specified price level by maintaining a discrepancy between the market and real [natural] rates until the desired price level is reached, and then equalizing the rates at that point" (594). Rate differential determines deposit growth, which in turn determines price level change.

Marget repeatedly makes exactly the same point (1938, 179–86, especially 184–85), arguing, for example, that the level "of general prices depends upon the total amount of bank money issued," which, "in turn, depends upon the relation of bank rate to natural rate" (263). He likewise voices the related point that Wicksell saw adjustments in the central bank's loan rate of interest as working through money stock changes to stabilize prices in the feedback policy rule. Loan rate changes lead to corresponding changes in the demand for and supply of bank loans. More importantly, such rate changes lead to changes in the stock of deposit money created as a byproduct of the loans. This monetary change in turn moves prices. Here, then, we find the quantity theory proposition that although the interest rate differential determines changes in the stock of bank money, those money stock changes must precede and cause the resulting price level movements. Myhrman's summary of the quantity theory interpretation is apt: Wicksell "explained the role of... inside money and the rate of interest in the transmission of monetary impulses to the price level [showing that] causation runs from the monetary system to the price level" (1991, 272).

Passive Money View

In contrast to Marget, Myhrman, Patinkin, and Trautwein, however, other prominent Wicksell scholars, notably Trygve Haavelmo (1978), Jürg Niehans (1990), and Axel Leijonhufvud (1981), deny that changes in the stock of bank money play a crucial, price-determining role in Wicksell's cumulative process. In their interpretation, Wicksell held that interest rate differentials and the resulting excess aggregate demand drive up prices *directly* without the necessary intervention of bank money creation. Instead, bank money expansion comes at the end of each stage of the cumulative process and only then to accommodate, or validate, price increases already produced by nonmonetary forces. As Haavelmo puts it, rate-created "excess [aggregate] demand is the primary force, which inflates the value of PX [nominal output]" (1978, 214). Afterwards, the stock of loan-created bank money moves "along passively in order to cover the public's [monetary] requirements, which in turn depend on PX" (214).

In short, according to Haavelmo the behavior of bank money in Wicksell's cumulative process is best described by the Banking Principle, according to which

> the quantity of [bank] money plays a. . . passive role; it adjusts in accordance with the [monetary] requirements created by changes in the value of transactions when the price level is forced up or down by *other* factors. (1978, 210)

Niehans explicitly endorses Haavelmo's passive-money interpretation. He asserts that in "Wicksell's approach" the supply of bank money, far from playing an active, price-determining role, instead "adjusts passively to whatever households and firms demand" at given prices (1990, 275). Leijonhufvud agrees. He writes that "the excess demand for commodities" rather than "acceleration in the growth rate of 'money'" is what "drive[s] the price-level up" (1981, 159–60). It follows that "watching '*M*'. . . would not be of much help in forming rational expectations. In a world like Wicksell's, the money stock would be a lagging indicator. The growth rate of *M* is not driving the cumulative process" (159–60).

Leijonhufvud, Niehans, and Haavelmo are far from the first to claim that Wicksell's cumulative process consists of a transmission mechanism with links running unidirectionally from aggregate demand to prices and thence to money demand and supply. Earlier interpreters claimed to find this same mechanism in which bank money appears at the tail end of the causal queue. Thus William P. Yohe quotes a 1908 statement by one S. F. Altman alleging that Wicksell "believes that the [money] holding follows the price movement, which takes place through stronger purchase or sale of goods" (1959, 144, n. 67). Small wonder that Hugo Hegeland observed that "Knut Wicksell has provoked more discussion as to whether he was a opponent or adherent of the quantity theory than perhaps any other economist" (1951, 133).

Five Contentions

In an effort to resolve the controversy over the active money versus passive money interpretations of Wicksell, this article argues five points. First, proponents of the quantity theory interpretation may perhaps possess the correct analysis of the cumulative process and the operation of the feedback rule, namely that changes in the stock of bank money must precede and induce price level changes. Second, those proponents, their claims to the contrary notwithstanding, cannot consistently and unambiguously find that interpretation in Wicksell, who at times seems to side with the passive money view. That Wicksell's formulation could spawn two polar opposite views—one monetarist, the other antimonetarist—is not surprising given its ambiguities, inconsistencies, and peculiarities of phrasing and definition.

Third, for the quantity-theoretic version of the cumulative process and the policy rule, one must go not to Wicksell but rather to Cassel, his rival for the professorship at Lund, who presented that version in a remarkable but underrated article entitled "The Rate of Interest, the Bank Rate, and the Stabilization of Prices" and published in the August 1928 issue of the *Quarterly Journal of Economics*. The article is especially noteworthy because it challenges the widespread view that Cassel adhered to a simple monetary model that excluded interest rates and had the path of the price level determined solely by the differential growth rates of the nominal supply of and real demand for monetary gold (see Jonung [1979]). True, Cassel used that simple model in much of his empirical work as reported in his famous textbook *The Theory of Social Economy.* But his *QJE* piece shows that, in at least one key theoretical essay, he employed a Wicksellian framework that incorporated natural and market rates of interest as well as an endogenous stock of inside, loan-created money to determine the price level.[1]

The fourth contention of this article is that Cassel's active-money exposition of the cumulative process contains innovations that advance it beyond Wicksell's exposition. Cassel, like Wicksell, uses the cumulative process model to derive a stabilizing policy rule, but unlike Wicksell, he extends it to the analysis of the business cycle and alternative proposed monetary norms as well. With respect to the business cycle, he applies the cumulative process to show that monetary factors amplify real fluctuations. In other words, he broadens the scope of application of the cumulative process analysis beyond the confines imposed by Wicksell. In so doing, he demonstrates the versatility and explanatory power of the quantity theory.

Fifth, on one matter at least, namely the analysis of the operation of the price-stabilizing feedback policy rule, Cassel's discussion lacks the precision of Wicksell's. Wicksell not only specified the exact indicator to which the central bank responds but also described the behavior of the time path of the price level when it is constrained or influenced by the policy rule. Nevertheless, Cassel more than Wicksell saw that quantity-theoretic logic underlay their policy rules.

1. QUANTITY THEORY INTERPRETATION

Proponents of the quantity theory interpretation of Wicksell's work generally attribute to him a version of the cumulative process model describable by five relationships shown below. These relationships are meant to depict the case of Wicksell's pure credit economy in which (i) all saving S is deposited in banks,

[1] Cassel's article is noteworthy also because it runs counter to Bo Gustafsson's contention that Cassel's "expositions are not seldom marred by contradictions and a vagueness in expression, only scantily veiled by his mastery of round and polished sentences" (1987, 375). Contrary to that verdict, Cassel's exposition of the cumulative process in his *QJE* article is among the clearest and most succinct to be found in the literature on Wicksell. The mystery is why the Swedish successors of Wicksell and Cassel ignored this article. Had they read and cited it, the subsequent anti-quantity theory interpretation of the cumulative process might never have appeared, or at least would have been rendered more suspect than it was.

(ii) all investment I is bank-financed, (iii) the economy is closed such that all saving and investment are of domestic origin, (iv) banks lend solely to finance investment, (v) full employment prevails such that shifts in aggregate demand affect prices but not real output, which remains at its capacity level, and (vi) agents, always expecting current prices to prevail in the future, anticipate none of the price changes that occur.

Embodying the foregoing assumptions, the five relationships are capable of depicting steady state equilibrium as well as the dynamic disequilibrium adjustment process triggered by disturbances to equilibrium. The steady state solution obtains when the relationships are set equal to zero, resulting in the celebrated conditions of Wicksellian monetary equilibrium. These conditions are market rate of interest equals natural rate, saving equals investment, aggregate demand equals aggregate supply both in real and nominal terms, and the stock of bank money and the price level are stable and unchanging.

Now, a modern general equilibrium theorist, schooled in the notion that self-corrective forces operate with sufficient swiftness to maintain model economies in equilibrium, would solve the equations for their above-mentioned steady state values. He or she would further treat dynamics not as disequilibrium processes, but rather as equilibrium paths driven by moving state variables. Not so Wicksell. Believing that persistent departures from equilibrium were commonplace, he had more ambitious plans for his model. To him and many of his interpreters, the baseline conditions of monetary equilibrium merely set the stage for the cumulative disequilibrium process, which begins when the natural rate diverges from the market rate (see Trautwein [1996], 31–32). Wicksell attributed such divergences to a multitude of real shocks that disturb the natural rate while the inertial forces of habit, routine, and absence of base-money reserve constraints in the pure credit economy introduce sluggishness into bankers' adjustment of the market rate. In the pure credit economy, central bankers theoretically could hold the market rate—which in pure cash and mixed cash-credit economies tends to converge to the natural rate—below or above that latter rate forever.

Let the resulting natural-market rate divergence activate the cumulative process. Immediately the relationships shed their zero equilibrium steady state solutions to depict dynamic disequilibrium responses and adaptations. Shown below, the relationships in their dynamic setting treat causality as running unidirectionally from the independent variables on the right side of each equation to the dependent variables on the left. True, the modern theorist versed in formal equilibrium analysis may question this mode of reasoning. Accustomed to thinking in terms of a system of equations simultaneously satisfied by a set of variables, he or she would argue that it makes no sense to think of one variable adjusting first and thus causing another to adjust, and so on. Nevertheless, it is just this sort of chain of causation that lies at the heart of Wicksell's inflation mechanism and of the active versus passive money debate.

And it is just this sort of chain that the following relationships describe:

$$I - S = a(r - i), \tag{1}$$

$$dM/dt = I - S, \tag{2}$$

$$X = dM/dt, \tag{3}$$

$$E = X, \tag{4}$$

and

$$dP/dt = bE. \tag{5}$$

Equation (1) says that because lower market interest rates i encourage capital formation and discourage thrift, the planned investment expenditure I of business firms exceeds the planned voluntary saving S of households when the natural rate of interest r (the rate that equilibrates saving and investment) exceeds the lagging market rate i set by the banking system.[2] Here the coefficient a is the parameter that relates the investment-saving gap to the interest rate differential that generates it.

Equation (2) states that the gap, or excess of desired investment over desired saving, equals the additional money dM/dt newly created as a byproduct of the loans made to finance the gap. In other words, since the central bank (the only bank in the pure credit economy) creates new check-deposit money by way of loan, monetary expansion occurs when it lends more funds to business investors than it receives on deposit from savers (who Wicksell assumes lodge all their savings with the bank). Equation (2) admits of a simple derivation. Denote business demands for bank loans L_D as $L_D = I(i)$, where $I(i)$ is the schedule relating desired investment spending (assumed to be entirely financed by bank loans) with the loan rate of interest, or cost of borrowing, i. Similarly, denote bank loan supply L_S as the sum of household saving S deposited with banks plus new money dM/dt created by banks in accommodating loan demands. In short, $L_S = S(i) + dM/dt$. Equating loan supply and loan demand $L_S = L_D$ (where the causal arrow runs from right to left

[2] A lower market rate stimulates planned investment by raising the present discounted value of the stream of expected future returns to capital. The rise in this discounted revenue stream raises the price of capital goods above their replacement cost and makes it profitable to produce more of them. Furthermore, since the market rate is the intertemporal relative price of consumption today in terms of consumption sacrificed tomorrow, a fall in that price induces people to take more of consumption today. Consumption rises and saving falls, hence the shortfall of saving below investment at lower than natural interest rates.

since loan supply passively accommodates itself to loan demand) and solving for the gap between investment and saving yields equation (2).

Equation (3) is absolutely essential to the quantity theory interpretation. It recognizes that while the quantity of bank *loans* in an accommodative banking system is passively demand determined and can never be in excess supply, the same cannot be said for the stock of *money* created as a byproduct of the loans. On the contrary, such a loan-created money stock can, as long as nominal transactions and thus the public's demand for transaction balances remains momentarily unchanged, indeed be redundant, or overissued. Thus equation (3) says that because at prevailing prices P and real output Q the public's demand for money M_D as expressed by the equation $M_D = kPQ$ has not yet changed, the new money dM/dt created by loan constitutes an excess supply of money X.[3] This undesired excess money supply is essential to the operation of the cumulative process because without it moneyholders would have no incentive to spend the additional money away. And with no incentive to spend it away, there would be no force to propel prices upward. Instead, the new money would be willingly held and absorbed into transaction balances and thus could never spur spending and prices.

Accordingly, equation (4) says that cashholders attempt to rid themselves of the excess money X by spending it on goods and services. The result is that the surplus money spills over into the commodity market to underwrite and mediate the excess aggregate demand for goods E implied by the gap between investment and saving. Indeed, the expenditure of the excess money is what transforms the excess *desired,* intended demand implicit in the investment-saving gap into excess *effective,* actual demand. In sum, equation (4) embodies Walras's Law according to which an excess demand for goods must be matched by a corresponding excess supply of something else, which quantity theorists take to be money.

According to equation (5), because Wicksell assumed that output is always at its full capacity level and so cannot expand, the excess effective demand E must exhaust its force in bidding up prices, which rise by an amount dP/dt proportional to the excess demand, with the coefficient b denoting the factor of proportionality. Substituting equations (1)–(4) into (5) and (1) into (2) yields the two equations

$$dP/dt = ab(r - i) \tag{6}$$

and

$$dM/dt = a(r - i), \tag{7}$$

[3] The money demand function $M_D = kPQ$ is the famous Marshall-Pigou (or Cambridge) *cash-balance* equation, in which the parameter k denotes the fraction of nominal income PQ that people desire to hold in the form of money balances M. Continental European quantity theorists already were beginning to employ this function, often in verbal rather than symbolic form, in Wicksell's time (see Ellis [1937, 154–75]).

which together state that price inflation and the money growth that underlies and permits it stem from discrepancies between the natural and market rates of interest. Further substitution of equation (7) into equation (6) yields the expression

$$dP/dt = b(dM/dt), \tag{8}$$

with causation running as always from right to left. Per this quantity theory interpretation, bank monetary expansion dM/dt is the necessary link that translates interest differentials into price level changes dP/dt in Wicksell's cumulative process.

2. ANTI-QUANTITY THEORY INTERPRETATION

By contrast, proponents of the passive-money interpretation who claim Wicksell as an adherent drop equations (2), (3), and (4) and have excess aggregate demand E itself (which they define as identical to the investment-saving gap) directly determine the price level change according to the three-equation system in which money is conspicuously absent:

$$I - S = a(r - i), \tag{9}$$

$$E = I - S, \tag{10}$$

and

$$dP/dt = bE. \tag{11}$$

In the passive-money interpretation, money stock changes, far from being the active intervening element that transforms interest differentials into price level changes, adapt passively to support the price changes already produced by excess aggregate demand. That is, assuming (i) that purchasers demand loans L_D from banks in order to be able to buy the same real quantity of goods Q at the raised prices dP/dt, (ii) that banks accommodate these borrowers by supplying new loans L_S in the form of bank money creation dM/dt, and (iii) that money circulates against goods with a given turnover velocity V, one obtains

$$L_D = (Q/V)dP/dt, \tag{12}$$

$$L_S = dM/dt, \tag{13}$$

and

$$L_S = L_D, \tag{14}$$

which upon substitution yields

$$dM/dt = (Q/V)dP/dt, \tag{15}$$

with causation running from price level changes dP/dt to money stock changes dM/dt.

In short, with the money stock adjusting passively to changes in the price level component of the money demand function, there can be no excess money supply. And without an excess supply of money, there is nothing to induce moneyholders to attempt to rid themselves of it by spending it away. No redundant money exists to spill over into the commodity market in the form of an excess demand for goods to bid up prices. On the contrary, far from over- or underissue forcing a change in prices, money supply conforms to money demand with neither excess nor deficiency and causality runs from prices to money in the passive money view. Here is an interpretation stemming from Wicksell's own analysis that is antithetical to what quantity theorists claim he sought to accomplish.

3. WICKSELL'S OWN VIEW

Quantity theorists may be right in contending that Wicksell, in Hans-Michael Trautwein's words, "wanted to demonstrate that the quantity theory of money is valid even in the extreme [pure credit economy] case of money supply endogeneity" (1996, 31). Still, it is difficult if not impossible to prove Trautwein's proposition conclusively from a representative sample of Wicksell's own writings. It is no wonder that quantity and anti-quantity theorists alike can claim Wicksell as an ally. In some passages, he indeed sides with the quantity theory, holding that bank money expansion is the crucial link connecting rate differentials to price level changes and transforming ex ante investment-saving gaps into ex post excess aggregate demand. In his 1898 article "Influence of the Rate of Interest on Commodity Prices," Wicksell speaks of prices adapting "themselves to the increase in the amount of money," implying that monetary expansion occurs before prices can change (80). Again, in volume two of his *Lectures on Political Economy* he implies money-to-price causality when he writes "of the influence of credit [demand deposits] on prices" (1906, 164).

Passive Money and Reverse Causality

In other passages, however, Wicksell unambiguously sides with the passive-money view. Asserting reverse causality, he writes in his 1925 piece "The Monetary Problem of the Scandinavian Countries" that monetary expansion may occur after rather than before prices have increased. Specifically, he argues that spenders themselves can directly raise nominal national income simply by bidding up all prices (accomplished through a temporary rise in velocity) and subsequently borrowing from the banking system to cover the increased monetary requirement. Describing a pure credit economy in which "all payments were made on a cheque basis," he says that whereas deposit checking accounts

> would constantly increase in amount as prices rose, *at first... there would be no increase in the average amount or in the aggregate of these accounts.* In the course of time they would become inconveniently small in proportion to the increased volume of monetary payments [required to buy the national product valued at the higher prices]. *They would consequently need to be adjusted upwards.* In the final analysis this presupposes an increase in bank credit. [In this manner] as prices rose, bank deposits and bank loans would swell more or less automatically. (1925, 202, emphasis added)

The causal and temporal sequence here runs from prices P to loans L to money M with M adapting itself passively to prior changes in P.

Again, in still another passage asserting reverse causality, Wicksell writes that "a general rise in prices will cause banks of issue to increase their issue of notes" and that even if the "banks flatly refuse to expand their circulation" they cannot "prevent the rise [of prices] or force prices down"—those prices obeying nonmonetary imperatives (202). It is on the basis of these passages that Bertil Ohlin, in his introduction to the English translation of *Interest and Prices,* claims that Wicksell believed that "a general rise in prices may well come about because consumers increase their demand... for consumption goods. This... need not have anything to do with too large credits to producers. The conclusion to be drawn... is that... prices may rise or fall *ad libitum*" (1936, xx–xxi). In short, Wicksell provides ammunition for quantity and anti-quantity theory forces alike.

Application of Real Balance Mechanics to Outside Money

Wicksell's inconsistency is most apparent in his contradictory treatment of an excess supply of outside versus inside money. In the case of outside money—gold coin and convertible currency—he recognized that such an excess money supply indeed could occur in pure cash and mixed cash-credit regimes and then spill over into the commodity market in the form of an excess demand for goods that drives up prices. In perhaps the best description of the operation of a real balance effect in the neoclassical monetary literature, he explained ([1898] 1965, 39–40) how a rise in M (or a random fall in P) would cause actual cash balances to become greater than desired. He then described how cashholders, in an effort to work off these undesired balances, would spend the excess money on goods until prices rose sufficiently to render actual balances equal to desired ones.

Failure to Apply Real Balance Mechanics to Inside Money

When it came to inside, bank-created money, however, Wicksell abandoned the notion of an excess supply of money. The impossibility of a redundant stock of deposit money is already implicit in his tendency to define deposits and loans indiscriminately as *credit*. With this definition, he conflated a non-demand-determined variable (deposits) with a demand-determined one (loans). Treating both identically, he failed to see that deposits could be in excess supply even if loans—passively provided upon demand by a pliant, accommodative banking system—were not. As far as deposits were concerned, he argued that their quantity (like that of loans) is always demand determined. Further, he contended that deposit supply and demand are identical at all prices, such that both the price level and the nominal quantity of deposit money are indeterminate in the pure credit economy. In his words,

> We have seen that in our ideal state [the pure credit economy] every payment, and consequently every loan, is accomplished by means of cheques or *giro* facilities. It is then no longer possible to refer to the supply of money as an independent magnitude, differing from the demand for money. No matter what amount of money may be demanded from the banks, that is the amount which they are in a position to lend. . . . The banks have merely to enter a figure in the borrower's account to represent a credit granted or a deposit created. When a cheque is then drawn and subsequently presented to the banks, they credit the account of the owner of the cheque with a deposit of the appropriate amount (or reduce his debit by that amount). *The "supply of money" is thus furnished by the demand itself.* (1898, 110, emphasis added)

If Wicksell's conclusion is correct, it follows that bank money can never be in excess supply. And if it can never be in excess supply, it cannot induce holders to attempt to rid themselves of it by spending it away. And if it is not spent away, it cannot be the force that generates an excess demand for goods and bids up the price level. One has to question, therefore, quantity theorists' wisdom in attributing equations (3) and (4) to Wicksell.

In short, with bank money completely demand determined, there can be no real balance effects of the kind that Wicksell applied to coin and currency in his treatment of pure cash and mixed cash-credit economies. Bank money, that is, cannot be the source of price level changes. It is hard to dispute David

Laidler's summary judgment: "There is no logical reason why Wicksell could not have" acknowledged that the public's demand for exchange media "would tend to give bank deposits the same role in the credit economy as currency in the cash economy: and then to note that deposits generated as a byproduct of credit creation would have, by way of cash balance mechanics, their own influence on the economy," that is, on the price level (1991, 148). "He did not take these steps, however" (148). His failure to do so would deprive his pure-credit-economy version of the cumulative process of quantity theory foundations and render it susceptible to anti-quantity theory interpretations.[4]

4. CASSEL'S QUANTITY-THEORETIC VERSION OF THE CUMULATIVE PROCESS

For a straightforward, consistent account of the quantity theory version of the cumulative process and feedback policy rule one must look not to Wicksell but rather to the work of his compatriot and sometime rival Gustav Cassel. In his 1928 *Quarterly Journal of Economics* article, Cassel, without once mentioning Wicksell's name,[5] developed the cumulative process analysis for the case of a loan-created inconvertible banknote money administered by a central bank, which Cassel treats as the only bank in the economy.[6] The monetary regime he

[4] In contrast to the position taken above, a modern equilibrium theorist might find Wicksell's ambiguity commendable. He or she would argue as follows: First, one cannot rely on a full, or complete, general equilibrium analysis of Wicksell's model economy since the maintained assumption is that the bank rate is exogenously fixed for a time below its natural equilibrium level. Second, given this assumption, the proper method of analysis is to ask what the consistency conditions arising from market clearing and individual optimization imply about the remaining variables, money and prices in particular. These conditions imply that the stock of money and the price level both must rise. But nothing in the pure logic of Wicksell's abstract economy requires that the rising of one variable must be causally or temporally prior to the rising of the other. Therefore, Wicksell was right to leave the matter ambiguous. If so, then the whole active-versus-passive-money debate reduces to much ado about nothing. The fact remains, however, that the debate, pointless or not, has raged for almost one hundred years.

[5] Despite Cassel's failure to cite Wicksell, he was clearly polishing and perfecting the latter's model.

[6] Cassel's article exemplifies the tendency of scientific integrity to prevail over personal animosity in rigorous disciplines such as economics. It is no secret that Wicksell and Cassel disliked each other and frequently disagreed on issues other than the goal of price stability (Seligman 1962, 562; Blaug 1986, 43). Enmity between the two surfaced during their competition for the professorship at Lund when Wicksell advised Cassel to withdraw his application and disparaged his capital theory as the work of a rank amateur (Gardlund 1958, 321–22). Later, in correspondence, Wicksell complained of Cassel's arrogance, his overweening self-esteem, his pretensions to originality, and his notorious failure to cite predecessors and contemporaries whose ideas he used (Gardlund 1958, 322). Wicksell was, in his own words, put off by Cassel's habit of "incessantly singing his own praises, appointing himself generalissimus over the rest of us poor creatures" (322). Mutual antagonism intensified in 1919 when Wicksell published a devastating critique of Cassel's *Theory of Social Economy*, a critique that Cassel's favorite pupil, Gunnar Myrdal (1945, 10, quoted in Carlson 1994, 31, n. 4), called "bitter and uncomprehending" and that Cassel's secretary, Ingrid Giöbel-Lilja (1948, 231, quoted in Carlson 1994, 31, n. 4) described as revealing "a deep lack of understanding, almost bordering on hatred, of Cassel's whole personality." Following the publication of Wicksell's critique, Cassel ceased attending meetings of the Political Economy Club in

considers is therefore virtually the same as Wicksell's pure credit regime, the only difference being that inside money in the form of inconvertible banknotes replaces checking deposits as the sole medium of exchange.

Quantity Theory Components

Cassel provides a verbal account of all the components of the quantity theory version of the cumulative process. Of the equations $I - S = a(r - i)$ and $dM/dt = I - S$, he says, "there exists a definite equilibrium rate of interest [r]. If the bank rate [i] is lower than this equilibrium rate, people will go to the bank for covering their needs for capital [$I - S$], and the bank will have to issue notes [dM/dt] to meet such needs" (1928, 517).

He likewise makes it clear that the initial effect of the interest differential is to generate a loan-created monetary expansion that occurs prior to the rise in prices. "If the bank rate is kept too low [$r - i$]," he writes, "people will find it advantageous to borrow at the bank [$L_S = L_D$] and thus the supply of the means of payment [dM/dt] will swell" (516). In other words, a monetary overissue occurs as "the market borrows unduly much from the bank and becomes too abundantly supplied with means of payment" (517). The result is "an unnecessarily large issue of notes" (517) or "excessive supply of means of payment [X]" (527)—excessive, that is, in relation to the real demand for it, which, "without any more goods having been produced," (517) remains unchanged. Here is Cassel's recognition of the excess money supply condition $X = dM/dt$. Here, too, is his recognition of the corresponding excess aggregate demand and price-rise relationships $E = X$ and $dP/dt = bE$. These conditions hold, he says, when the excess money supply spills over into the commodity market in the form of an excess demand for goods that, in the fully employed economy, "is bound to force up prices" (517).

Application to Deflationary Case

Cassel applied the cumulative process analysis to the symmetrical case of price deflation. "If the bank rate [i] is raised above the equilibrium rate of interest [r], the demand for loans is affected" (1928, 525). Loan demand shrinks and with it loan supply and the nominal stock of money. The fall in the money stock means that "the nominal purchasing power of the market is reduced"

Stockholm, where Wicksell regularly aired his views. The antipathy culminated in Cassel's (1926; see Ohlin [1972, 107]) declining to write an obituary article on the recently deceased Wicksell on the grounds that "too much separated us" and that he could not in good faith give an unbiased appraisal of a man whose "extraordinarily dogmatic" character prevented him from appreciating Cassel's own work and that of others. Yet this antipathy did not prevent Cassel from inadvertently doing Wicksell—and monetary science—the supreme favor, two years after his death, of shearing Wicksell's cumulative process analysis of ambiguities and inconsistencies and securing it with solid quantity-theoretic foundations. Though delayed, the drive for scientific integrity triumphed after all.

below the unchanged real demand for it. In an effort to restore money balances to their desired level, people cut back their spending for goods “with the result that prices in general must fall” (525). Through the creation of an excess demand for money matched by an excess supply of goods, “the raising of the bank rate above the equilibrium rate... brings about a fall in the general level of prices,” just as “a reduction of the bank rate below the equilibrium rate,” by generating an excess money supply, is “bound to raise the general level of prices” (525–26).

To summarize, the foregoing constitute Cassel’s statements of the equations $X = dM/dt$, $E = X$, $dP/dt = bE$, and, via substitution, $dP/dt = b(dM/dt)$. This last equation encapsulates his acknowledgement of “the rise in prices that must follow upon the excessive supply of means of payment” just as the quantity theory’s postulate of money-to-price causality contends (527).

5. CASSEL ON THE CONDITIONS NECESSARY FOR PRICE STABILITY

Cassel’s credentials as a quantity-theory interpreter of the cumulative process manifest themselves most strongly in his discussion of the conditions required for price stability. Like Wicksell, Cassel stressed that “stability of prices is possible only when the bank rate is kept equal to the equilibrium rate of interest,” that is, when the rate differential is zero (1928, 517). Far more emphatically than Wicksell, however, he argued that not the two-rate equality per se but rather the resulting monetary limitation is the fundamental condition for price stability. Said he, “the purchasing power of the monetary unit is... determined by the scarcity that the central bank chooses to give to its note circulation” (516). Without such scarcity, “any price could be paid and prices would continue to rise indefinitely” (515). It therefore follows that an “indispensable condition of [price] stability is... that the supply of means of payment should be limited and thus that a certain scarcity in this supply should exist” (515). So when the central bank brings its bank rate to equality with the natural rate in order to “restrict its issue of notes,” it is the latter restriction itself and not the rate adjustment that stabilizes prices (516). The rate adjustment, because it limits loan demands and the quantity of bank money created as a byproduct of their accommodation, is merely the means by which the end of price-stabilizing monetary restriction is achieved.

6. CASSEL’S REJECTION OF INTEREST COST-PUSH THEORIES

The preceding has argued that Cassel, more so than Wicksell, established the quantity-theoretic foundations of the pure-credit-economy version of the cumulative process. Further evidence confirming Cassel's strong adherence to the quantity theory comes from his critique of cost-push, or more precisely interest cost-push, theories of inflation.

Cost-push theories, of course, are the very antithesis of the quantity theory. They attribute price inflation not to excess money growth, but rather to underlying rises in factor-input prices (wages, rents, interest) that enter into unit costs of production. These costs are then passed on to consumers in the form of higher product prices. As a species of this genus, interest cost-push theories identify rises in the price of capital services as the inflationary culprit. As Cassel put it, they hold that "since the rate of interest is the price for a [capital] service" that "enters into the cost of production just as the price of any other service required in the process of production" (1928, 525), it therefore follows "that an increase in the rate of interest is bound to increase the cost of all products and therefore to enhance prices" (524). Cassel attributes this theory to the "practical business man" who, believing that rate hikes raise prices, "finds it very confusing when he hears a scientific economist or a representative of a central bank proclaim that the rate is increased in order to force prices down" (524–25).

Fallacies of the Interest Cost-Push View

Cassel rejected the interest cost-push view on two grounds. First, it confuses relative prices with the general (absolute) level of prices. Cost changes indeed influence the former set of prices, but money supply and demand determine the latter. It follows that if the central bank keeps the nominal supply of money equal to the real demand for it, relative prices will move with changes in the cost of production while aggregate prices remain unchanged. The structure and composition of relative prices will change, but not their general average.

Second, the theory erroneously assumes wages and rents do not fall when interest rates rise. In fact, economic logic strongly suggests that the opposite is true. Confronted with rising interest rates, cost-minimizing producers are likely to respond by cutting production and laying off labor and land. Owners of those factor inputs, in a successful effort to keep them fully employed, reduce their asking prices. Wages and rents fall. With capital inputs rising in price and labor and land inputs falling in price, the upshot is clear. The relative cost (and price) of capital-intensive goods—goods using capital relatively intensively in their production—rise when interest rates rise whereas the relative costs (and prices) of labor and land-intensive goods tend to fall.

Interest Cost-Push Affects Relative Prices, Not Absolute Prices

These considerations led Cassel to argue that, provided the central bank holds constant the stock of money per unit of real output by maintaining equality between market and natural rates, rises in interest rates can raise the relative prices of capital-intensive goods but not the aggregate of all prices. With the central bank limiting the money stock, "every rise in some prices must necessarily be counterbalanced by a fall in others" (Cassel 1928, 525). Why? Because the higher-cost and hence dearer-priced capital-intensive goods will require more money to be spent on their purchase leaving less for spending on labor- and land-intensive goods whose prices will accordingly fall. In the final analysis, upon a matched rise in the level of market and natural interest rates such that the money stock and aggregate spending remain unchanged, "only those goods will rise in price for the production of which a particularly large amount of disposal of capital has been required, whereas other prices must sink so low that the average level of all prices remains unaltered" (525).

Here is Cassel's contention that the aggregate price level is a monetary phenomenon immune to matching (equilibrium) changes in the natural and market rates of interest. Here is his claim that such rate changes, being real phenomena, affect only relative real prices. Here too is his recognition that if the average of all prices is kept unchanged, it follows as a matter of arithmetic that a rise in some relative prices must be offset by a compensating fall in others.

7. EXTENSIONS OF THE CUMULATIVE PROCESS ANALYSIS

Wicksell applied the cumulative process analysis to explain price level movements alone. Cassel's active-money view of the cumulative process, however, led him naturally to extend the analysis to examine cyclical fluctuations in real activity, something Wicksell was loath to do. Wicksell attributed business cycles to fluctuations in the natural rate and its underlying real determinants (technological progress, wars, and the like) rather than to discrepancies between that rate and the market rate. Hence, to him the cumulative process model with its two-rate differential was irrelevant to the analysis of the cycle.

Monetary Misbehavior Amplifies Real Cycles

Cassel disagreed. He held that rate differentials and the attendant surpluses and shortages of bank money magnify the amplitude and duration of cycles caused by real shocks. They "very much increase the strength of the cyclical movement of trade, with all its pernicious effects" (Cassel 1928, 528). In upswings, when cyclical improvements in capital productivity raise the natural rate above the sluggishly adjusting market rate, the resulting rate differential and the excess money it creates produce too much investment compared to the amount savers are willing to supply. The result is an unsustainable overinvestment boom that inevitably gives way to an underinvestment slump when cyclical falls in capital productivity lower the natural rate below the market rate. Clearly the monetary surpluses and shortages spawned by rate differentials accentuate real cycles. If they could be removed by central bank policy that keeps the market rate in continuous alignment with the natural rate, then, according to Cassel, "the whole cyclical movement of trade must become very much attenuated. For it [the cycle] will then be deprived of the great stimulus derived from the continual falsification of the capital market that is the consequence of an alternatively too abundant and too scarce supply of means of payment" (528).

Here was a key difference between Wicksell and Cassel. Both believed that cycles were essentially real phenomena generated by movements in the natural rate. But Cassel, wedded as he was to the active money view, further believed, as Wicksell did not, that monetary factors augmented real cycles and rendered them more damaging than they otherwise would be. Here then was Cassel's justification for using the cumulative process analysis to study trade fluctuations: it revealed how money stock surpluses and shortages emanating from two-rate differentials exacerbated real cycles. In so doing, it revealed still another rationale for the active pursuit of monetary and price level stability: such stability could help constrain the business cycle and keep it within the limits dictated by real shocks and real propagation mechanisms alone.

Rejection of Non-Price-Stabilizing Policy Norms

It was on these grounds that Cassel (1928, 519–20) rejected alternative policy norms calling for (i) gently rising prices or creeping inflation, (ii) price deflation at a rate equal to the rate of productivity growth, and (iii) cyclically fluctuating prices. By departing from absolute monetary and price level stability, such norms implied corresponding deviations between market and natural rates of interest with all the cyclical dislocations attendant thereto.

Critique of the Gold Standard

It was also on these same grounds that Cassel (1928, 520–22) criticized the gold standard as a monetary regime. Under the gold standard, the nation's price level was determined by the following relationship: dollar price of goods (the price level) equals fixed dollar price of gold times worldwide gold price of goods. By permitting movements in the worldwide gold price of goods—movements virtually guaranteed by dissimilar fluctuations in the respective growth rates of gold and goods—to pass through to corresponding movements in national general price levels, the gold standard institutionalized price instability and the disruptions it would bring. Little wonder that Cassel, unconvinced as he was that foolproof ways could be found to prevent fluctuations in the world gold price of goods from affecting national price levels, recommended abolishing the gold standard for a rational paper standard administered by the central bank.

8. CASSEL AND WICKSELL ON THE FEEDBACK POLICY RULE

A rational money standard works only as well as the rule or norm the central bank employs in conducting policy. Both Wicksell and Cassel thought that the theoretically ideal policy rule was for the central bank to maintain its bank rate in continuous equality with the natural rate. But both also believed that such a rule was infeasible because it required knowledge of the natural rate, seen by them as an unobservable variable that is impossible to target.

Still, both men contended that the bank could target the price level even though it could never directly target the unobservable natural rate. It could determine from movements in the price level whether the bank rate was too low or too high relative to the natural rate and thus needed adjustment. As Cassel put it, since "it is impossible for the central bank to know exactly what this 'natural rate' is" (1928, 528), the "only practical way of ascertaining what is the correct bank rate is, therefore, by observing the results. If prices are seen to rise continuously, the bank may be sure that the rate is too low. Vice versa, when prices fall, the bank may conclude that the rate is too high" (518).

Cassel's Statement of the Rule

From these considerations Cassel derived his version of the Wicksellian policy rule: "The bank has to adjust its rate so that no general tendency either to a rise or to a fall in prices arises. The practical rule is, therefore, that the bank rate should be so adjusted as to keep the general level of prices as constant as possible" (1928, 512).

Cassel's rule, however, lacks the precision of Wicksell's. In the latter rule, the bank rate adjusts in response to price deviations from target, and the response continues until prices roll back to their pre-inflation or pre-deflation levels. By contrast, Cassel's rule is hardly that specific. It says only that the rate must be manipulated to hold prices constant. It fails to specify the indicator variable—namely price deviations from target $P - P_T$—to which the central bank responds. And it fails to note that the response must be sustained until prices return to target.

Cassel's imprecise formulation of the policy rule prevented him from seeing what Wicksell understood implicitly, namely that the rule can at best only stabilize prices on average over time. It cannot stabilize them at every point in time. It can constrain their fluctuations within a narrow band about target, but it cannot continually keep them at target.

Dynamic Stability-of-Equilibrium Analysis Applied to Wicksell's Rule

Wicksell's conclusion—that a feedback policy rule linking bank rate adjustments to price level deviations from target can at best deliver price stability on average—emerges from a stability-of-equilibrium analysis performed on the model presented earlier in the article. Although there is no evidence that Wicksell himself performed this analysis, it is useful to do so here. First, reduce the rule-constrained cumulative process model to two differential equations. One, $dP/dt = a(r - i)$, states that prices adjust linearly to the natural rate–bank rate differential. The other, $di/dt = g(P - P_T)$, states that the central bank adjusts its rate di/dt in a fixed proportion g to price level deviations from target $P - P_T$. Here, of course, the natural rate r and price target P_T are treated as given, fixed constants, the natural rate having attained its predetermined level from a prior real shock.

Second, form the Jacobian matrix of the partial derivatives of the differential equations. This two-by-two matrix has as elements 0 and $-a$ in the first row and g and 0 in the second.

Third, observe that the matrix possesses a zero trace and a positive determinant ag. These two conditions, well known from stability analysis, indicate that the price level oscillates ceaselessly about target at an amplitude that depends upon the magnitudes of the adjustment parameters a and g.

Wicksell, of course, intuitively understood this result. He maintained that his feedback rule, if implemented, could deliver approximate stability in the sense of constraining price level fluctuations within a narrow band of plus and minus 3 percent about target (Uhr 1991, 94). Evidently such modest perpetual overshooting of the price level target bothered him not in the least. Had it bothered him, he might have modified his rule slightly to prevent such ceaseless overshooting and to ensure that prices eventually converge to target either monotonically or via damped oscillatory paths.

Wicksell's Rule Augmented

The modification in question calls for the central bank to adjust its interest rate in response both to price level deviations from target and to the rate of change (time derivative) of the price level according to the augmented rule $di/dt = g(P - P_T) + h(dP/dt)$. Adding this last term to the reduced-form model's rate-adjustment equation yields a Jacobian with a negative trace $-ha$ and a positive determinant ag. Both are required to ensure price convergence to target.

This modified rule seems eminently reasonable. Certainly central bankers, if charged with the duty of stabilizing prices, would respond to price level changes dP/dt as well as to price level gaps $P - P_T$. For just as a pilot landing a jumbo jet must heed his plane's vertical distance from the runway and its speed of approach lest it descend too rapidly and crash, so too must the central bank watch the gap between actual and target prices and the rate of price change lest it overshoot its target. Aside from this oversight, however, Wicksell's understanding of the feedback rule must be judged superior to Cassel's.

Bank Rate Affects Money Stock, Which Affects Price Level

Still, on one point at least Cassel outshone Wicksell. Cassel made it clear that the bank rate operates to stabilize prices not directly but indirectly, through the money stock. Bank rate adjustment affects the demand for and supply of loans and the quantity of money created as an offshoot of the loans. Changes in the money stock then restore prices to target. The rate is the central bank's instrument variable, the money stock its intermediate variable, and the price level its goal variable. In Cassel's own words, "the purchasing power of the monetary unit" is "determined by the scarcity the central bank chooses to give to its note circulation" (1928, 516). And "the ultimate and essential means" whereby "it is able to restrict its issue of notes" is "the bank rate" (516). Causation runs from bank rate to money supply to price level.

The significance of Cassel's contribution is this: it implies that money may be crucial to the workings of monetary policy even in models that exclude money from their equations. Wicksell, of course, had constructed just such a model. The cumulative process and policy rule equations of his model omit money and instead have interest rate adjustments alone moving prices. They give the impression that the behavior of the quantity of money is essentially a sideshow, irrelevant to the operation of the policy rule. Cassel's work implies that this impression is wrong. Although he failed to write down a formal, price-stabilizing policy rule, Cassel instinctively understood that the quantity theory underlies such a rule just as it does the cumulative process. If so, then money plays a role even in Wicksell's moneyless model. Money is crucial to the workings of the model because it translates rate changes and differentials into price level changes.

9. CONCLUSION

Anti-quantity and quantity theory interpretations vie in modern readings of Wicksell's cashless-economy model of the cumulative process. And for good reason: Wicksell wrote passages that support both interpretations. Some passages allude to quantity-theoretic money-to-price causality, others to anti-quantity theoretic reverse, price-to-money causality. Moreover, he explicitly states the anti-quantity theory notion of a passive, demand-determined stock of inside money. Believing that such money can never be in excess supply, he fails to apply real balance mechanics to it to explain why people attempt to rid themselves of it by spending it away and so force a rise in prices. His pure credit economy case differs from his pure cash and mixed cash-credit economy cases in which the quantity theory always plays a dominant role.

His inconsistency is easily explained. As a pioneering monetary theorist, he was engaged in pathbreaking work of the highest order. Operating in new and unfamiliar territory, he was forging a complex analysis that combined elements of capital theory, price theory, production and distribution theory, and monetary theory. Involved as he was in this ambitious and far-reaching project, he could hardly be expected to state every nuance with the precision, clarity, and consistency that later scholars well acquainted with his analysis could give it. In any case, he failed to convey his intentions as clearly as one might have wished. In so doing, he left the door open for some of his successors to give his cumulative process analysis anti-quantity theory interpretations.

It remained for Gustav Cassel, writing 30 years after the publication of Wicksell's *Interest and Prices,* and fully cognizant of what Wicksell had sought to accomplish, to express matters clearly and to articulate the active money view. In so doing, he established for all time the quantity-theoretic foundations of the Wicksellian triumvirate: pure credit economy, cumulative process, and stabilizing policy rule. He showed that an endogenous, loan-created stock of bank money was essential to translate interest rate differentials into price level changes in the pure credit economy. Likewise, he established that bank rate adjustments work through money stock changes to stabilize prices in the operation of the feedback policy rule. In short, he completed the work Wicksell had started 30 years before.

Unfortunately, Cassel's contributions to cumulative process analysis and to the theory of stabilizing policy rules have gone largely unnoticed. Few cite his 1928 *QJE* article featuring those contributions. Citations instead are made to his *Theory of Social Economy* in which the contributions are missing. He is remembered today for (i) his purchasing power parity theory of exchange rates, (ii) his simplified version of the Walrasian system of general equilibrium, a version stripped of Walras, mathematics, marginal utility, and marginal productivity, (iii) his empirical claim that the differential growth rates of the gold stock and real output determine the path of the price level, and (iv) his theory that the limited life span (interest earning period) of savers sets a floor to interest rates. It is clear that he also deserves equal credit for establishing quantity theory foundations for policy rules and the cumulative process. Had Cassel's successors been more fully aware of his work in this area, subsequent interpretations of Wicksell's monetary constructs might have taken a different turn. In any case, Cassel's rediscovered insights, highlighting as they do the originality and explanatory power of Wicksell's analytical model, confirm and underscore Wicksell's place in the pantheon of monetary theorists.

REFERENCES

Blaug, Mark. 1986. *Great Economists before Keynes: An Introduction to the Lives & Works of One Hundred Great Economists of the Past.* Cambridge: Cambridge University Press.

Carlson, Benny. 1994. *The State as a Monster: Gustav Cassel and Eli Heckscher on the Role and Growth of the State.* Lanham, MD: University Press of America.

Cassel, Gustav. Letter of 10 May 1926 to Bertil Ohlin.

______________. 1928. "The Rate of Interest, the Bank Rate, and the Stabilization of Prices." *Quarterly Journal of Economics* 42 (August): 511–29.

______________. 1932. *Theory of Social Economy,* 2nd ed. Revised. New York: Harcourt, Brace & Co.

Ellis, Howard S. 1937. *German Monetary Theory 1905–1933.* Cambridge, Mass.: Harvard University Press.

Gardlund, Torsten. 1958. *The Life of Knut Wicksell.* Stockholm: Almqvist & Wiksell.

Giöbel-Lilja, Ingrid. 1948. *Gustav Cassel. En livsskildring. (Gustav Cassel. A Biography).* Stockholm: Natur och Kultur.

Gustafsson, Bo. 1987. "Cassel, Gustav." In *The New Palgrave: A Dictionary of Economics,* vol. 1, ed. John Eatwell, Murray Milgate, and Peter Newman. London: Macmillan: 375–77.

Haavelmo, Trygve. 1978. "Wicksell on the Currency Theory vs. The Banking Principle." *The Scandinavian Journal of Economics* 80: 209–15.

Hegeland, Hugo. 1951. *The Quantity Theory of Money.* Goteborg: Elanders Boktryckeri Aktiebolag.

Jonung, Lars. 1979. "Knut Wicksell and Gustav Cassel on Secular Movements in Prices." *Journal of Money, Credit and Banking* 11 (May): 165–81.

Laidler, David. 1991. *The Golden Age of the Quantity Theory: The Development of Neoclassical Monetary Economics, 1870–1914.* Princeton: Princeton University Press.

Leijonhufvud, Axel. 1981. "The Wicksell Connection—Variations on a Theme." In *Information and Coordination: Essays in Macroeconomic Theory.* Oxford: Oxford University Press: 131–202.

Marget, Arthur. 1938. *The Theory of Prices: A Re-examination of the Central Problems of Monetary Theory,* vol. 1. New York: Prentice-Hall.

Myhrman, Johan. 1991. "The Monetary Economics of the Stockholm School." In *The Stockholm School of Economics Revisited*, ed. Lars Jonung. Cambridge: Cambridge University Press: 267–93.

Myrdal, Gustav. 1945. "Gustav Cassel in memoriam." *Ekonomisk Revy.*

Niehans, Jürg. 1990. *A History of Economic Theory: Classic Contributions, 1720–1980.* Baltimore: Johns Hopkins University Press.

Ohlin, Bertil. 1936. "Introduction" to the English translation of *Wicksell's Interest and Prices.* London: Macmillan: vii–xxi.

______________. 1972. *Ung man blir politiker (Young Man Becomes Politician).* Stockholm: Bonniers.

Patinkin, Don. 1965. *Money, Interest, and Prices,* 2nd ed. New York: Harper & Row.

Seligman, Ben. 1962. *Main Currents of Modern Economics: Economic Thought Since 1870.* New York: Free Press.

Trautwein, Hans-Michael. 1996. " Money, Equilibrium, and the Business Cycle: Hayek's Wicksellian Dichotomy." *History of Political Economy* 28 (1): 27–55.

Uhr, Carl G. 1991. "Knut Wicksell, Neoclassicist and Iconoclast." In *The History of Swedish Economic Thought,* ed. Bo Sandelin. London and New York: Routledge.

Wicksell, Knut. [1898] 1965. *Interest and Prices.* Trans. R. F. Kahn. London: Macmillan; reprinted New York: Augustus M. Kelley.

______________. [1906] 1978. "Money." In *Lectures on Political Economy,* vol. 2. Trans. E. Classen, ed. Lionel Robbins. London: Routledge and Kegan Paul; reprinted New York: Augustus M. Kelley.

______________. [1925] 1936. "The Monetary Problem of the Scandinavian Countries." *Ekonomisk Tidskrift*. Reprinted in *Interest and Prices*. London: Macmillan: 197–219.

______________. 1969. "The Influence of the Rate of Interest on Commodity Prices." Reprinted in Knut Wicksell, *Selected Papers on Economic Theory,* ed. Erik Lindahl. New York: 67–89.

Yohe, William P. 1959. The Wicksellian Tradition in Swedish Macroeconomic Theory. Ph.D. diss., University of Michigan.