Research Summaries

Introduction

Bank of Canada staff undertake research designed to improve overall knowledge and understanding of the Canadian and international financial systems. This work is often pursued from a broad system-wide perspective that emphasizes linkages across the different parts of the financial system (institutions, markets, and clearing and settlement systems) linkages between the Canadian financial system and the rest of the economy, and linkages to the international environment, including the international financial system. This section summarizes some of the Bank's recent work.

Financial institutions and clearing houses face a number of financial risks, including the credit and market risks that arise from their participation in financial infrastructures, such as the securities clearing and settlement system. Collateral in the form of equities and fixed-income instruments is commonly used to manage these risks. But collateral itself can change in value over time. Thus, it is important to require a pledge of collateral large enough to cover any losses should a risk materialize. In **Collateral Valuation for Extreme Market Events**, Alejandro García and Ramazan Gençay propose a framework that can be used to compare different methods of measuring the risk surrounding the future value of collateral. This analysis is useful in determining the amount of collateral required to cover risks.

An efficient and productive financial system is important for the development and longer-run growth of the economy. To better understand the factors that might contribute to improved economic performance, policy-makers are often interested in cross-country comparisons. In this regard, comparisons of Canada-U.S. productivity have become topical, with suggestions of a "productivity gap" in some Canadian industries, including financial services, where Canadian banks play a very prominent role. Jason Allen, Walter Engert, and Ying Liu have recently studied the efficiency of major Canadian banks, measuring it against that of comparable U.S. banks. That work is summarized in **Are Canadian Banks Efficient? A Canada-U.S. Comparison**.

Payments systems are typically characterized by some degree of tiering, with upstream firms (clearing agents) providing settlement accounts to downstream institutions that wish to clear and settle payments indirectly (indirect clearers). Clearing agents provide their indirect clearers with an essential input (clearing and settlement services), while also competing directly with them in the retail market for payment services. In the article, **Credit in a Tiered Payments System**, Alexandra Lai, Nikil Chande, and Sean O'Connor construct a model of a clearing agent with an indirect clearer to examine the clearing agent's incentives to lever its upstream position to gain a competitive advantage in the market for retail payment services. The model demonstrates that a clearing agent can achieve this competitive advantage by raising the indirect clearer's costs; however, the incentive to raise these costs is mitigated by credit risk to the clearing agent arising from the provision of uncollateralized overdrafts to its indirect clearer. The results suggest that tiered payments systems, which require clearing agents to provide overdraft facilities to their indirect clearers, may result in a more competitive retail payment services market.

In the article, **Using No-Arbitrage Models to Predict Exchange Rates**, Antonio Diez de los Rios proposes a model of the joint behaviour of interest rates and the exchange rate in two countries. In the model, movements in these variables are related in such a way as to preclude the existence of arbitrage opportunities. The term structure and the expected rate of depreciation of the exchange rate are functions of both domestic and foreign short-term interest rates.

The author finds that imposing the no-arbitrage restrictions in the estimation of the model produces exchange rate forecasts that are superior to those produced by time-series methods such as a random-walk model or a vector autoregression. This is a notable result, given that the random-walk model has proved very difficult to beat in forecasting exchange rates.

Collateral Valuation for Extreme Market Events

*Alejandro García and Ramazan Gençay**

Clearing and settlement systems are critical to the infrastructure of financial markets because of the large values of funds and securities that settle through them. For instance, in 2005, $49.9 trillion was settled through the Canadian securities clearing and settlement system (CDSX). Given the large values flowing through these systems, regulators and banking professionals have taken initiatives to make them safer.

A common factor in many of these initiatives is the use of collateral to manage financial risks. For example, participants in a clearing and settlement system may have to pledge collateral equivalent in value to the amount they owe. If a participant fails and is unable to pay the amount owing, the collateral can be sold to generate the needed funds. But collateral itself may consist of risky assets and thus can change in value over time. It is therefore necessary to require a pledge of collateral large enough to adequately cover all losses in the event of a failure.

To manage the risk created by the uncertainty surrounding the future value of collateral, the initial value of the collateral is discounted. In other words, participants must pledge a greater amount of collateral than the amount owing. This discount is often referred to as the "haircut."[1] The larger the haircut, the lower the risk, but the higher the costs incurred by participants using the system.

In this article, we propose a framework that can be used to compare different methods for calculating haircuts. Particular attention is paid to selecting an appropriate method for low-probability events (e.g., large, unexpected declines in asset prices) that might affect the stability of the financial system, and one that also takes into account the cost of pledging collateral.

Methods for Estimating Haircuts

Two components are needed to calculate a haircut for collateral. The first is a model of the distribution of losses (i.e., frequency with which the asset declines in value), since the distribution of returns is unknown. The second is a risk measure, which can be thought of as a way of mapping the loss distribution into a single number (the haircut).

There are several ways to model the loss distribution for collateral based on historical data for returns. These include:

- **Parametric approaches** that use historical data to obtain the parameters necessary to characterize a given distribution (e.g., Normal, *t*, etc.). These parameters are then used to approximate the return distribution, and the haircut is obtained from the resulting quantile, given a particular distribution and a confidence level.[2]
- **Non-parametric approaches**, such as historical-simulation techniques, that do not model the return distribution under some explicit parametric model, but instead use the empirical distribution of the data to estimate the quantiles, for a given confidence level.

1. The haircut represents the amount by which the security could decline in value subject to a confidence level and a holding period.

* This article summarizes García and Gençay (2006).

2. Quantiles are points taken at regular intervals from the cumulative distribution function. Dividing the ordered data into *q* equal-sized data subsets is the motivation for *q*-quantiles. The quantiles are the corresponding data values marking the boundaries between consecutive subsets.

Along with choosing one of the above approaches, the estimation of haircuts requires a means of quantifying risk: a risk measure. Various risk measures can be used. One of the most common is the Value at Risk (VaR). We also use an alternative risk measure called Expected Shortfall (ES).[3]

The method for calculating a haircut can most easily be explained with an example. Consider an exposure of $100 in a system for clearing and settling securities. This exposure is collateralized by an asset that has a market price of $100. To estimate the haircut for such an asset, we use a parametric approach (e.g., a normal return distribution) and select a risk measure (e.g., VaR). Knowing that the asset has a daily percentange change in price with a mean of zero and a standard deviation of 3 per cent, we estimate the corresponding normal distribution. Next, we choose a confidence level for the haircut (e.g., 0.5 per cent)[4] and then select a holding period (e.g., 1 day). Finally, we calculate the corresponding VaR obtained from a normal distribution with the mean and standard deviation of the data and assign this value as the haircut.[5] This parametric approach, combined with VaR, yields a haircut of 7.72 per cent (quantile of the distribution), which is associated with a tail risk of 0.5 per cent (confidence level). With this haircut, the amount of collateral required to cover the exposure of $100, given the characteristics of the asset pledged, would be $108.36 which is (100/[1–haircut]).

Using Extreme Value Theory to Characterize the Distribution of Returns

A number of empirical observations generally hold for a wide range of financial time series.[6] One of these is that return series have fat tails. This means that compared with a normal distribution, there are fewer observations around the mean, and more in the tails or extremes of the distribution. This is true for many equities and certain fixed-income instruments that may be pledged as collateral. For such assets, it is not appropriate to use a normal distribution to estimate the distribution of market returns. This is because the normal distribution cannot capture values at very low or high tails of the distribution. Extreme value theory (EVT) methods are more appropriate for modelling the tail behaviour of the distribution of returns for securities.[7]

The intuition of EVT is as follows. While the normal distribution is the important limiting distribution for sample averages (central limit theorem), *the family of extreme value distributions is used as the limiting distribution of the sample extremes*. Thus, it is more relevant when we are interested in the extremes of the distribution. This family can be presented under a single parametrization known as the generalized extreme value distribution.[8]

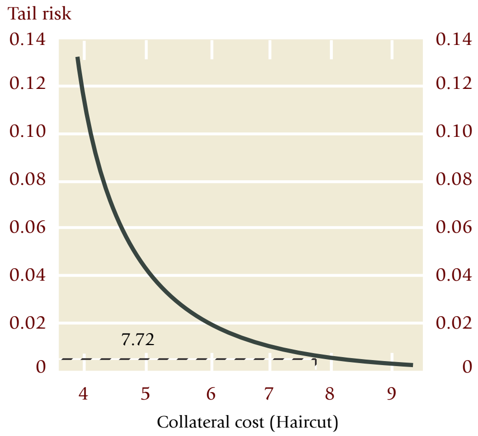
The power of EVT methods to capture extreme events is illustrated in Gençay and Selçuk (2006), where the authors use data for Turkey's overnight interest rate prior to the crisis when the rate reached a level of 873 per cent on 1 December 2000 and 4,000 per cent on 21 February 2001. The authors find that estimation results from the pre-crisis data indicate that a day with overnight interest rates over 1,000 per cent (simple annual) could be expected every 4 years. In other words, the extraordinary levels observed during the crisis were in the nature of the economy before they actually materialized.

The Risk-Cost Frontier

Having suggested some alternative methods for estimating collateral haircuts, we now need a framework for comparing the methods. We propose the "risk-cost frontier" as such a framework. The frontier is a way of summarizing the risk-cost trade-off implied by each method. Each method has its own trade-off between the risk that price fluctuations in collateral value are not covered by a haircut (tail risk), and the cost of pledging collateral, measured by the excess collateral above the exposure that corresponds

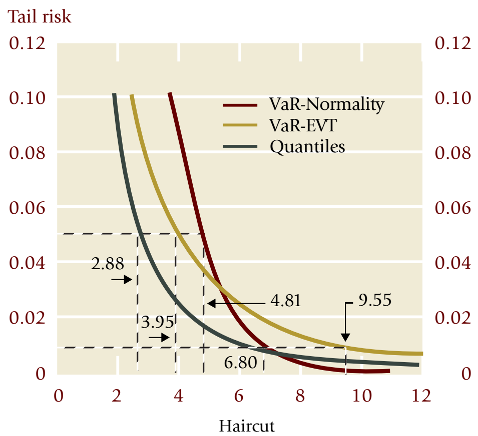
3. ES is a *coherent* alternative to VaR, where *coherence* is defined as axioms that capture the desired properties of a risk measure. This term is from Artzner et al. (1997, 1999).
4. This means that 1 day out of 200, the haircut would not be sufficient to cover the daily price fluctuations.
5. VaR is simply a quantile of the loss distribution of returns. This quantile represents the maximum loss that is not exceeded with a given high probability.
6. A good reference of the stylized facts for financial time series can be found in Mandelbrot (1963).
7. Embrechts, Klüppelberg, and Mikosch (1997) is a comprehensive source of theory and applications of extreme value theory to the finance and insurance literature.
8. This result is known as the Fisher-Tippett theorem.

Chart 1 Risk-Cost Frontier under Normality



Note: Haircuts and tail risk values obtained from a simulation of a normal distribution with a zero mean and standard deviation of 3 per cent.

Chart 2 Comparison of Methods for Calculating Haircuts



Note: Collateral returns are simulated from a *t* distribution with 2.2 degrees of freedom.

to the haircut (collateral cost). The trade-off exists because *larger haircuts* imply *lower tail risk* but *higher collateral cost*.

The risk-cost frontier can be constructed by calculating haircuts for different levels of tail risk but using the same method to model the return distribution. For example, the level of tail risk could start at 0.5 per cent and go up to 10 per cent. We can then calculate the associated haircuts. From these pairs of points, we can construct a risk-cost frontier. Chart 1 depicts the risk-cost frontier corresponding to the example given earlier (normal with mean zero and standard deviation of 3 per cent and a VaR risk measure).

Evaluating Haircut Estimation Methods

The risk-cost frontier can be used to compare different methods of calculating haircuts. Haircuts for the same levels of tail risk are calculated using different methods (i.e., combinations of (i) models for the loss distribution and (ii) risk measures).

The risk-cost frontier can then be used to determine the most appropriate method by selecting one whose frontier is closest to a benchmark frontier constructed from the data, but that does not cross it and, therefore, does not underestimate the haircuts. Consider the following example. First, the returns on a hypothetical asset are simulated using a *t*-distribution with 2.2 degrees of freedom. This specification shares similar statistical properties, such as fat tails, with those in financial time series. Two different methods are then used to estimate the haircuts. Knowing the underlying data-generating distribution allows us to determine that the best method for calculating the haircut is the one that has a risk-cost frontier closer to the risk-cost frontier calculated directly for the simulated data (using a non-parametric approach).

In this example, we compare two methods: both use a parametric approach, but one will assume a normal distribution and one an extreme value distribution. Both methods use VaR as the risk measure. Chart 2 shows the three risk-cost frontiers: the benchmark case with a green line (non-parametric approach for the empirical quantiles), the method based on the normal distribution with a red line, and the method that uses an extreme value theory distribution with a gold line.

Chart 2 illustrates the mismeasurement of risk when comparing the risk-cost frontier of the method that assumes a normal distribution, with the benchmark risk-cost frontier calculated from the simulated data (denoted by a green line). In Chart 2, we also observe that use of an extreme value distribution gives haircuts that are closer to benchmark given by the quantiles of the simulated *t* data (green line in Chart 2). Chart 2 suggests that the method that uses an extreme value distribution is the more appropriate one.

In our study, we also conduct the same analysis using real market data and find similar results. These results can be summarized as follows:

- Methods that use VaR on the assumption of normality overestimate (at high levels of tail risk) and underestimate (at low levels of tail risk) the values for the haircuts. This happens because the risk-cost frontier that uses the normality assumption crosses the benchmark frontier constructed from the empirical quantiles (green line in Chart 2). Thus, for the purpose of covering extreme risk, VaR with normality may not be adequate.
- VaR calculated with EVT methods provides a good fit in terms of slope to quantiles of the data. Nevertheless, VaR with EVT gives larger values for haircuts compared with the actual quantiles of the data. For the purpose of covering extreme risk, VaR with EVT is adequate. It should be kept in mind, however, that although they provide a cushion for extreme events, larger haircuts are costly to participants of the system.

Ultimately, the selection of the method for calculating haircuts depends on the weight placed on collateral costs versus coverage of extreme risk, and this depends on the objectives of the risk manager. Managers in critical financial infrastructures may choose to select a haircut that corresponds to a higher quantile than managers in organizations with greater tolerance for risk. No matter what the weights placed on risk and cost may be, a careful examination of the statistical properties of the return distribution is always recommended in order to select the most appropriate method for calculating haircuts.

Conclusions

We propose a framework that allows us to (i) characterize the risk-cost trade-off for a particular risk measure and method of haircut estimation, and (ii) compare different risk measures from alternative estimation methods, using the risk-cost frontier. The framework proposed is useful for understanding the risk-cost trade-off implied by the method used to calculate the collateral value (haircuts) that institutions must pledge to cover their exposures. These institutions may be clearing houses, central counterparties, payment system operators, central banks, or commercial banks determining their risk capital.

References

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. 1997. "Thinking Coherently." *Risk* 10: 68–71.

———. 1999. "Coherent Measures of Risk." *Mathematical Finance* 9: 203–28.

Embrechts, P., C. Klüppelberg, and T. Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance.* New York: Springer.

García, A. and R. Gençay. 2006. "Risk-Cost Frontier and Collateral Valuation in Securities Settlement Systems for Extreme Market Events." Bank of Canada Working Paper No. 2006-17.

Gençay, R. and F. Selçuk. 2006. "Overnight Borrowing, Interest Rates and Extreme Value Theory." *European Economic Review* 50: 547–63.

Mandelbrot, B. 1963. "New Methods in Statistical Economics." *Journal of Political Economy* 71: 421–40.

Are Canadian Banks Efficient? A Canada-U.S. Comparison

Jason Allen (Bank of Canada), Walter Engert (Bank of Canada), and Ying Liu (Université de la Méditerranée)

Box 1

Canadian and U.S. Banks

The six major Canadian banks in our sample comprise over 90 per cent of the assets of the Canadian banking sector. The banks are Royal Bank Financial Group, Bank of Montreal, Canadian Imperial Bank of Commerce, TD Bank Financial Group, Bank of Nova Scotia, and National Bank.

The comparisons reported in this article consider total U.S. banks and a sample of 12 U.S. bank holding companies (BHCs). The BHCs are selected from the top 20 U.S. banks in terms of assets as of 31 December 2004. They were selected because there are continuous data from 1986 to 2004, and because most of these banks have a business mix broadly similar to that of the Canadian banks, benchmarked in a specific manner. That is, most of these BHCs make a similar proportion of revenue from retail banking.

The BHCs are JPMorgan Chase & Co., Bank of America Corp., Wachovia Corp., Wells Fargo & Co., U.S. Bancorp, SunTrust Banks Inc., National City Corp., Citizens Financial Group Inc., BB&T Corp., Fifth Third Bancorp, Keycorp, and The PNC Financial Services Group Inc.

An efficient and productive financial system is important for the development and longer-run growth of the economy. Indeed, a recent comprehensive survey of the research literature suggests that the quality of financial service provision is a key ingredient for economic growth (Dolar and Meh 2002).

To better understand the factors that might contribute to improved economic performance, policymakers are often interested in cross-country comparisons. In this regard, Canada-U.S. productivity comparisons have become topical, with suggestions of a "productivity gap" in some Canadian industries, including financial services—where Canadian banks play a very prominent role.

Given these various considerations, we recently studied the efficiency of major Canadian banks, and compared it with the efficiency of U.S. banks (Allen, Engert, and Liu 2006). This article presents a summary of that work.

Performance Measures

We begin by considering common performance ratios, comparing the six largest Canadian banks (which account for the vast majority of Canadian banking assets) with total U.S. banks, and with a subset of U.S. bank holding companies (BHCs). (See Box 1 for more on these banks.)

The data that we use are from the balance sheets and income statements reported by these institutions to the banking supervisors in Canada and in the United States. We deflate all variables by the consumer price index, excluding food and energy prices, in the respective country. We also adjust the data for the different purchasing powers of the Canadian and U.S. currencies.[1]

1. We use the Rao, Tang, and Wang (2004) calculation of a PPP measure for value-added in financial services (1.09 in 1999).

Expense ratio

The expense ratio is often used by analysts to evaluate bank performance. It is defined as the ratio of non-interest expense to net operating revenue (net interest income plus non-interest income).[2]

Chart 1 presents the expense ratio for Canadian banks, the U.S. BHCs, and total U.S. banks. The expense ratio of Canadian banks was lower than that of U.S. banks in the late 1980s and early 1990s. But this measure has been trending up at the Canadian banks and down at the U.S. banks over the sample period, so that the expense ratio of Canadian banks currently exceeds that of U.S. banks.

Our analysis indicates that the difference in the expense ratios can be currently attributed to a higher labour cost component (wages and benefits) at Canadian banks. However, this differential does not imply disparities in productivity, which concerns how much output is produced per unit of input (typically, labour).

Labour productivity ratio

Accordingly, we also consider measures that focus on the output produced by banks, relative to labour input. Bank output is difficult to measure, on both conceptual and pragmatic grounds. Indeed, it is widely believed that official statistics (based on the system of national accounts) on output in financial-services industries are subject to large errors. (See, for example, Triplett and Bosworth 2004 or Diewert 2005.)

In our study, we do not use national accounts data. As noted above, we draw on data from balance sheets and income statements provided to bank supervisors. To measure productivity, we begin with total assets reported on balance sheets as our measure of output.

Chart 2 compares total assets per full-time equivalent employee of Canadian banks, the

Chart 1 Expense Ratio

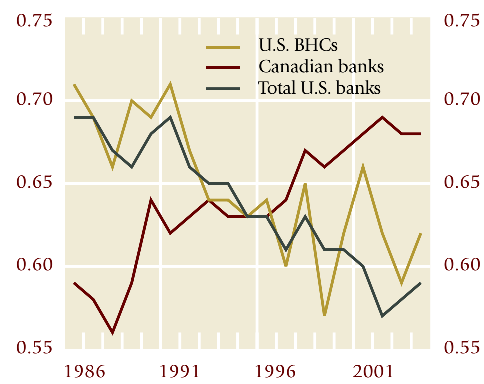
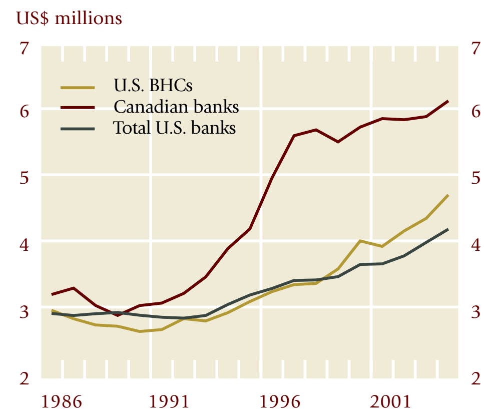
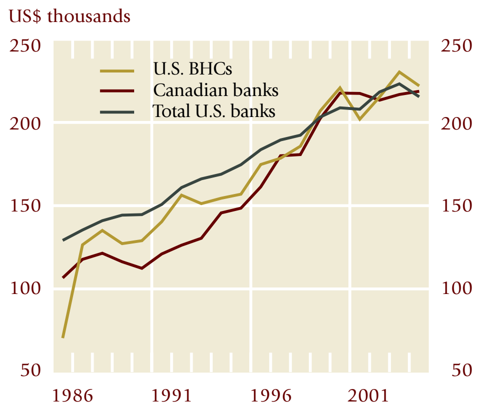


Chart 2 Assets per Employee



2. The denominator of this ratio—particularly net interest income—depends on the risk differential between assets and liabilities. Therefore, a change in the ratio can be caused by changes in risk taking and not necessarily by changed efficiency. A change in the mix of a bank's services or products (say, towards non-traditional banking services) can also affect this ratio by altering the mix of inputs and expenses. Thus, we prefer the term "expense ratio," and not "efficiency ratio," as it is sometimes called.

Chart 3 Net Operating Revenue per Employee



U.S. BHCs, and total U.S. banks, in constant 1999 U.S. dollars. This chart suggests that the productivity of Canadian banks has been considerably higher than that of U.S. banks in the past decade.[3]

Next, we consider a measure that effectively internalizes differences in asset generation and management, and focuses on overall results. Specifically, Chart 3 shows net operating revenue per full-time equivalent employee of Canadian banks, the U.S. BHCs, and total U.S. banks.

According to this measure, Canadian bank employees were less productive than their U.S. counterparts in the late 1980s, but started to catch up in the early 1990s. In fact, according to this measure, the three groups of banks have converged since the late 1990s, indicating that Canadian banks are as productive as their U.S. counterparts.[4]

Economies of Scale and Cost-Inefficiency

We also consider another means of gauging bank efficiency, based on econometric methods, using disaggregated bank data. In this case, our analytical framework is the translog cost function (as in Allen and Liu 2005), which has become a standard tool in the research literature.

Methodology

In this framework, a bank's cost-minimization problem can be written as a general cost function:

$$C = f(\mathbf{q}, \mathbf{w}) + \theta + \xi ,$$

where C is bank costs; $\mathbf{q}$ is a vector of bank outputs; $\mathbf{w}$ is a vector of input prices that a bank faces; and $f(\mathbf{q}, \mathbf{w})$ is a translog function, consisting of the individual and cross-product terms of $\mathbf{q}$ and $\mathbf{w}$. The term θ represents effects unique to each bank, and the error term ξ represents all other unexplained influences on a bank's cost structure.

3. Including a measure of non-traditional activities (such as those related to off-balance-sheet assets) in total assets does not change this conclusion.
4. It follows from these various performance ratios that the return on assets of Canadian banks is less than that of U.S. banks, which is what we see in the data. On the other hand, the return on equity of Canadian banks is comparable to (if not greater than) that of U.S. banks.

Inferences regarding economies of scale are drawn from the derivative of C with respect to $\mathbf{q}$; that is, how a bank's costs vary with its scale of output.

The error term ξ provides the basis for the measurement of "cost-inefficiency." We define the efficient frontier as the (benchmark) bank with the lowest inefficiency measure (based on its ξ), and then measure each bank's distance from that efficient frontier. An efficient banking system is represented by relatively small inefficiency measures and convergence over time towards the efficient frontier.

An additional parameter of interest is technological progress, which we approximate initially with a quadratic time trend and then with other variables in different specifications of the model. We also include variables to capture the effects of regulatory changes in Canada and the United States.[5]

Data

Three input prices are included in the model: labour, capital, and deposits. They are measured, respectively, as the average hourly wage of bank employees, the expenses on real estate and fixtures divided by the total stock of these items, and the effective interest rate paid on deposits. A bank's output is divided into five categories: consumer loans, mortgage loans, non-mortgage loans, other financial assets on the balance sheet, and an asset-equivalent measure of non-traditional activities (following the method of Boyd and Gertler 1994).

We estimate the model by panel dynamic least squares using quarterly data from 1983 through 2004 for the Canadian banks, and from 1986 through 2004 for the U.S. BHCs.[6]

5. The financial systems in Canada and the United States have been affected by a series of legislative changes over the past 20 years regarding bank powers, organization, and regulation. The specific nature and timing of these changes have been different in the two countries. But a cumulative effect has been the development of essentially universal banks in both countries over time.

6. Given the differences in the development of the institutional and regulatory environments (among other things) in Canada and in the United States, separate cost functions and efficient frontiers are estimated for the two countries. (Pooling the data across countries would make interpretation of ξ unreliable.) Also relevant in this regard is the fact that there is a larger size dispersion among the U.S. BHCs than in the Canadian bank sample.

Results

For our sample of Canadian banks, we reject the null hypothesis of constant returns to scale. Instead, we find increasing returns to scale (of about 7 per cent), suggesting that Canadian banks would gain (modestly) from being larger.

As regards the measure of cost-inefficiency for Canadian banks, we find that the gap between the efficient frontier and other banks averages less than 10 per cent, depending on the specification considered. More refined measures of technological change (capturing investment in employee training and automated banking machines, for example) lead to measures of cost-inefficiency among Canadian banks averaging about 6.5 per cent. As well, the estimates indicate that Canadian banks have tended to move closer to the efficient frontier over time.

For the U.S. case, the null hypothesis of constant returns to scale is rejected as well. Increasing returns to scale of about 2 per cent are estimated.

Estimates of cost-inefficiency for the sample of U.S. banks indicate that the gap between the efficient frontier and other banks is greater than 10 per cent, which is a typical result in the academic literature on U.S. bank efficiency (for example, Berger and Mester 1997). In our preferred specification, the average cost-inefficiency measure is about 14 per cent. As well, cost-inefficiency among the U.S. BHCs has not narrowed appreciably over the sample period.

We also find that the estimate of technological progress for Canadian banks is greater than for U.S. banks. Indeed, the results suggest that the effect of technological progress in lowering Canadian bank costs is three times greater than in the U.S. case—a result that we find surprising.[7]

Finally, we find that some of the legislative changes that have occurred in the past 20 years have reduced the cost structures of banks in both countries. For example, in Canada, the financial legislation revisions in 1987 and 1997

7. Other research, such as Tang and Wang (2004), also suggests that, in the recent past, productivity growth in Canadian financial services has been greater than in U.S. financial services, but not by a large margin. In our work, the time trend used to proxy technological progress is probably capturing the large increase in Canadian bank assets in the 1990s, when banks were expanding into a wide range of financial services.

were particularly beneficial in lowering banks' costs.

Conclusions

This work examines the efficiency and productivity of Canadian and U.S. banks in three ways. First, we compare key performance ratios and find that (i) the average Canadian bank employee produces more assets than the average U.S. bank employee, and (ii) in terms of producing net operating revenue, Canadian and U.S. bank workers are similarly productive.

Second, we investigate whether there are economies of scale in the cost functions of Canadian banks and a sample of U.S. BHCs. We find larger economies of scale for Canadian banks than for the U.S. BHCs. This suggests that Canadian banks are less efficient with regard to the scale of their operations and would have more to gain in terms of efficiency benefits from becoming larger.

Third, we measure cost-inefficiency in Canadian banks and in U.S. BHCs relative to the domestic efficient frontier in each country (the domestic best-practice institution). We find that Canadian banks are closer to the domestic efficient frontier than are the U.S. BHCs, and that they have moved closer to that efficient frontier over time.

Overall, these results do not suggest relative efficiency or productivity gaps in the Canadian banking industry. On the contrary, Canadian banks compare generally favourably.

Finally, as noted above, legislative and regulatory changes have benefited efficiency in Canadian financial services. This shows the importance of removing any remaining restrictions that inhibit competition and efficiency, but provide little (or no) benefit in terms of financial soundness.

References

Allen, J., W. Engert, and Y. Liu. 2006. "Are Canadian Banks Efficient? A Canada-U.S. Comparison." Bank of Canada Working Paper No. 2006-33.

Allen, J. and Y. Liu. 2005. "Efficiency and Economies of Scale of Large Canadian Banks." Bank of Canada Working Paper No. 2005-13. Forthcoming in *Canadian Journal of Economics*.

Berger, A.N. and L. Mester. 1997. "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?" *Journal of Banking and Finance* 21: 895–947.

Boyd, J. and M. Gertler. 1994. "Are Banks Dead? Or Are the Reports Greatly Exaggerated?" *Federal Reserve Bank of Minneapolis Quarterly Review* 18: 2–23.

Diewert, E. 2005. "Progress in Service Sector Productivity Measurement: Review Article on Productivity in the U.S. Services Sector: New Sources of Economic Growth." *International Productivity Monitor* 11: 57–69.

Dolar, V. and C. Meh. 2002. "Financial Structure and Economic Growth: A Non-Technical Survey." Bank of Canada Working Paper No. 2002-24.

Rao, S., J. Tang, and W. Wang. 2004. "Productivity Levels Between Canadian and U.S. Industries." Industry Canada Working Paper.

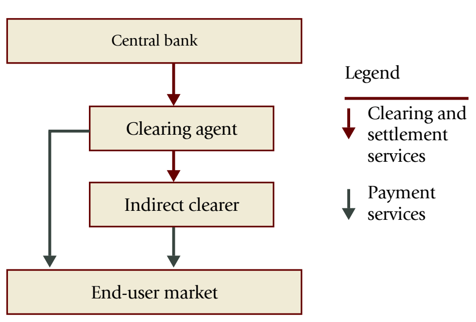
Tang, J. and W. Wang. 2004. "Sources of Aggregate Labour Productivity Growth in Canada and the United States." *Canadian Journal of Economics* 37: 421–44.

Triplett, J. and B. Bosworth. 2004. *Productivity in the U.S. Services Sector: New Sources of Economic Growth*. Washington, D.C.: Brookings Institution Press.

Credit in a Tiered Payments System

*Alexandra Lai, Nikil Chande, and Sean O'Connor**

Figure 1 A Tiered Payments System



Most payment, clearing, and settlement systems are characterized by some degree of tiering. In a tiered system, some of the financial institutions participating directly in a first-tier network for clearing and settlement (clearing agents) operate a second-tier network that provides similar services to other institutions (indirect clearers). Clearing agents not only provide wholesale clearing and settlement services to the indirect clearers participating in their second-tier networks, but also compete against these same indirect clearers in the provision of retail payment services to individuals and businesses. This arrangement is illustrated in Figure 1.

Survey evidence in Canada (Tripartite Study Group 2006) indicates that because of the high fixed costs associated with operating in the first-tier network, indirect clearers prefer to participate in a clearing agent's lower-cost, second-tier network. Clearing agents choose to operate second-tier networks to obtain scale economies and additional fee revenue. There are, however, some questions about the efficiency of pricing in service markets in tiered networks.

In a tiered payments system, a clearing agent has the ability, and may have the incentive, to raise the marginal cost for an indirect clearer in order to gain a competitive advantage in the market for retail payment services. Because of data-processing lags and distribution problems in the flow of settlement funds, the process of settling payments combines settlement services with credit services to network participants. For example, clearing agents provide overdraft credit to indirect clearers. We investigate the impact of uncollateralized overdrafts provided by a clearing agent in a second-tier system on the pricing strategy for its payment services, and on equilibria in the wholesale and retail markets.

If we abstract from inter-network competition to focus on within-network competition, the analysis indicates that the incentive for the clearing agent to raise the indirect clearer's costs is mitigated by the credit risk from the uncollateralized overdrafts that the clearing agent provides to its indirect clearer. In fact, in the model, the wholesale service fee charged by the clearing agent is always lower when credit risk is a meaningful consideration. The results also indicate that a clearing agent would then alter the price of its retail services to allow

* This article summarizes Lai, Chande, and O'Connor (2006).

its indirect clearer to acquire a greater share of the retail service market and higher profits. Furthermore, with a sufficiently high degree of competition in the retail market, customers are charged lower service fees by both the clearing agent and the indirect clearer.

Approach to the Analysis

The analytical approach builds on the existing research on the vertical integration of firms in wholesale and retail markets and on settlement credit. It links and extends these separate bodies of literature.

The literature on vertical integration establishes that where there is imperfect competition in both upstream and downstream markets (Spengler 1950), and even where competition is perfect in the downstream retail market (Salop 1998), incentives exist for a firm to vertically integrate the production of complementary services in both markets. Vertical integration eliminates double markups in the integrated firm's retail price and gives the integrated firm an opportunity to raise its rival's costs. Economides (1998), for example, demonstrates that when the price of upstream (input) services is regulated, a vertically integrated firm has an incentive to impose non-price costs on its downstream rivals. In the absence of input-price regulation, Bustos and Galetovic (2003) show that a vertically integrated firm prefers to increase a downstream rival's costs through the input price.

Similar modelling approaches have been applied to securities settlement systems. In particular, Holthausen and Tapking (2004) demonstrate that a central securities depository (CSD), vertically integrated with a custodian bank, will raise the costs of a rival custodian bank. Rochet (2005) shows that a CSD has an incentive to vertically integrate with a custodian bank and would either refuse to provide a rival bank with settlement services or, if regulation prevents exclusion, would raise the rival's costs.

None of this literature models the joint provision of settlement services and credit by the service provider, which is the case in a payments settlement system. Kahn and Roberds (1998) construct a single-network model for banks facing uncertain payment inflows and outflows through the period, with final settlement at the end on a net basis. In this system, network participants exchange intraday credit bilaterally or multilaterally to settle payments but, in doing so, also face the prospect of credit default.

Key Model Features

By combining the survey information with relevant studies on vertical integration, tiered systems, and settlement credit, we construct a model of a vertically integrated bank (the clearing agent) that competes downstream with a rival bank (the indirect clearer) in the end-user market for retail payment services. The clearing agent and the indirect clearer are Cournot competitors in the market for retail payment services,[1] but the indirect clearer purchases clearing and settlement services, and acquires overdraft credit, from the clearing agent. The clearing agent first chooses a clearing and settlement fee to charge the indirect clearer. Then, the clearing agent and indirect clearer simultaneously choose a desired volume of payment services in the end-user market and charge the corresponding retail service fee. Since each unit of service is measured by a payment transaction, and since the net value of these transactions is allowed to be random, net payment flows and settlement overdrafts from the clearing agent to the indirect clearer are uncertain at the time of their wholesale and retail pricing decisions.

Results

The results are derived from both analytical and numerical solutions to the model. The model shows that, to maximize expected net worth, the clearing agent will take advantage of its upstream position as an essential provider of clearing and settlement services to raise the indirect clearer's costs relative to its own marginal cost of clearing and settling these payments. Consequently, the indirect clearer offers its services at a higher price than those of the clearing agent, which enables the clearing agent to attract a greater share of the retail market and a relatively higher overall profit than the indirect clearer. This is the "integration" effect.

1. Cournot competitors select optimal strategies that take account of the rival's market reaction.

Credit risk to the clearing agent from the provision of overdrafts to its indirect clearer mitigates the clearing agent's incentive to raise the indirect clearer's costs. A default by the indirect clearer on its overdraft credit, resulting from insufficient profits and available assets, will lower the clearing agent's expected net worth. In selecting its pricing strategy, a forward-looking clearing agent will therefore take account of the prospect of overdraft credit to the indirect clearer, the probability of credit default by the indirect clearer, and the possible impact of higher pricing on the indirect clearer's default probability. The clearing agent must balance its potential gains in net worth from vertical integration against the potential losses it might incur by indirectly increasing its credit risk through its own pricing strategy. Therefore, recognizing that a decrease in the indirect clearer's profits implies that the indirect clearer is more likely to default, the clearing agent lowers its service fee. This is the "credit-risk" effect.

Numerical techniques help to determine whether the integration effect or the credit-risk effect dominates under different market conditions. For a broad range of parameter values, the credit-risk effect dominates. Specifically, when credit risk is meaningful to the clearing agent, it selects a wholesale service fee that is lower than the risk-free price. This allows the indirect clearer to acquire market share and earn higher expected profits, which lowers the probability of default on any overdraft credit that it may incur. There is, however, a level of retail competition below which the indirect clearer's profits are sufficiently high (with greater market power) that it can easily repay the settlement overdrafts provided by the clearing agent. Below this critical level of retail competition, credit risk is no longer a meaningful concern to the clearing agent, which allows the agent to charge a higher wholesale service fee. But the range of parameter values for which the integration effect dominates the credit-risk effect is very small. The presence of credit risk generally results in the clearing agent lowering its wholesale service fee relative to the risk-free case.

In addition to lowering its wholesale service fee when faced with sufficient credit risk, the clearing agent also selects a retail service price that lowers its own volume of retail payments. This pricing strategy allows the indirect clearer to raise the volume of its retail payments. Despite the loss of retail market share and a lower wholesale service fee, the clearing agent earns higher expected profits from combining clearing and settlement services with overdraft credit. The indirect clearer also earns higher profits, except where the degree of competition between the indirect clearer and the clearing agent is so low that the credit risk imposed on the clearing agent is insufficient to encourage the agent to lower its fee.

While the price of retail payment services charged by the indirect clearer is always lower in the presence of sufficient credit risk, the clearing agent's price is lower only when there is a high degree of competition between the two. In other words, significant competition is required for credit risk to lower the clearing agent's fee for retail payment services and, thus, make consumers unambiguously better off.

Conclusion

In a tiered payments system, a clearing agent provides its indirect clearer with an essential input (clearing and settlement services), but may also compete against the indirect clearer in the retail market for payment services. In the stylized model developed for this analysis, the clearing agent could take advantage of its position as operator of the second-tier network by strategically pricing its wholesale clearing fee so as to raise its rival's costs. But when the credit effect dominates, the clearing agent's incentive to raise the indirect clearer's costs is mitigated by the provision of overdraft settlement loans to the indirect clearer.

When clearing agents provide uncollateralized overdraft credit to indirect clearers and credit risk is significant, wholesale service fees are generally lower and the market for retail payment services can be more competitive. Furthermore, when there is a high degree of competition between clearing agents and indirect clearers, a tiered arrangement with credit is welfare-superior, from a consumer-price perspective, to one without credit and meaningful credit risk.

References

Bustos, A. and A. Galetovic. 2003. "Vertical Integration and Sabotage in Regulated Industries." University of Chile Centre for Applied Economics Working Paper No. 164.

Economides, N. 1998. "The Incentive for Non-Price Discrimination by an Input Monopolist." *International Journal of Industrial Organization* 16: 271–84.

Holthausen, C. and J. Tapking. 2004. "Raising Rival's Costs in the Securities Settlement Industry." European Central Bank Working Paper No. 376.

Kahn, C. and W. Roberds. 1998. "Payment System Settlement and Bank Incentives." *Review of Financial Studies* 11: 845–70.

Lai, A., N. Chande, and S. O'Connor. 2006. "Credit in a Tiered Payments System." Bank of Canada Working Paper No. 2006-36.

Rochet, J.-C. 2005. "The Welfare Effects of Vertical Integration in the Securities Clearing and Settlement Industry." IDEI, University of Toulouse. Unpublished manuscript.

Salop, S. 1998. "Vertical Mergers and Monopoly Leverage." In *The New Palgrave Dictionary of Economics and the Law*, P. Newman (ed.), 669–73. New York: Stockton Press.

Spengler, J. 1950. "Vertical Integration and Antitrust Policy." *Journal of Political Economy* 58: 347–52.

Tripartite Study Group. 2006. *Conditions for Direct Participation in the ACSS, Final Report*. Canadian Payments Association, June.

Using No-Arbitrage Models to Predict Exchange Rates

Antonio Diez de los Rios

Exchange rate predictions have many important applications. Risk managers use exchange rate predictions when deciding if (and when) to hedge currency movements. Portfolio managers use exchange rate predictions to obtain expected returns on foreign assets. Academics test their models of exchange rate determination on the basis of their ability to predict exchange rate movements.

Central bankers are also interested in having accurate models of exchange rate determination. For example, it is important to understand the forces that are driving currency movements, because different causes will have different implications for the economy. Ultimately, they may even require a different monetary policy response (Bailliu and King 2005; Ragan 2005). An assessment of international financial market stability and contagion also depends on the ability to understand large movements in currency markets.[1]

Predicting currency movements is, however, a difficult task. Despite the large body of research on exchange rate modelling, a key stylized fact in international finance is that the best prediction for tomorrow's exchange rate is today's rate (known as the "random-walk forecast").[2] This result was first discovered by Meese and Rogoff (1983a, b) and, even 25 years later, few models can do better than this one.[3] A related result, also found in the literature starting in the early 1980s, is that the forward rate does not provide the best prediction for tomorrow's exchange rate.[4] Thus, Clarida et al. (2003) note that "from the early 1980s onwards, exchange rate forecasting in general came increasingly to be seen as a hazardous occupation, and this remains largely the case."

This article summarizes a working paper (Diez de los Rios 2006) that proposes an arbitrage-free model of the joint behaviour of interest rates and exchange rates that provides exchange rate forecasts with improved predictive power when compared with the current set of foreign exchange rate models that do not impose these no-arbitrage restrictions.

No Arbitrage

It is hard to believe that exchange rates move independently of, for example, interest rates. The reason for such a skeptical statement is the concept of arbitrage in financial markets. If the prices of two related securities differ by a great amount, then an investor will have an incentive to buy the undervalued asset and sell the overvalued one to make a profit.[5] Thus, in an efficient market, arbitrage ensures that the prices of both assets do not move independently. For example, spot, forward, and Eurocurrency interest rates are mutually dependent through the familiar covered interest parity condition.[6]

1. See Berg, Borensztein, and Pattillo (2004) for a review on early-warning systems for currency crises.
2. Similarly, the best prediction at the one-month or one-year horizon is also today's exchange rate.
3. See Bailliu and King (2005) for a review of these successful models (including the Bank of Canada's Exchange Rate Equation).
4. Finance theory suggests that a risk-neutral investor should be indifferent between buying a one-month forward contract for a foreign currency or waiting one month and buying the currency directly in the spot market. This theory, known as "uncovered interest rate parity," implies that the best prediction for the future exchange rate is its forward counterpart (see Hansen and Hodrick 1980).
5. The technical definition of the absence of arbitrage states that it is impossible to obtain a portfolio that might provide a positive payoff (and never incur losses) without cost (see Cochrane 2001).
6. See Mark (2001) for more details on the covered interest parity condition.

A similar argument applies to domestic and foreign bonds. These assets are essentially imperfect substitutes with different levels of exchange rate risk. For instance, a Canadian investor who buys a one-year bond in the United Kingdom will know how many pounds sterling he will get in the future, but not how many Canadian dollars. Therefore, a Canadian investor will demand compensation for bearing the exchange rate risk. In other words, he will expect compensation for holding an asset that, from his point of view, is not perfectly risk free. If the rate of return (in Canadian dollars) of this British bond does not reflect this compensation, then the prices of British and Canadian bonds, as well as the bilateral exchange rate, should adjust until any arbitrage opportunities disappear. Therefore, the absence of arbitrage opportunities links the way in which interest rates and exchange rates can move over time.[7]

Overall, these so-called "no-arbitrage restrictions" provide useful information on how to model exchange rate movements and, therefore, how to improve exchange rate predictions.[8]

Model and Methodology

Motivated by the above arguments, Diez de los Rios (2006) uses a two-country affine term-structure model[9] to predict currency movements. The model leverages the no-arbitrage relationship between interest rates and exchange rates, itself a generalized version of the covered interest rate parity relation described above. In this model, the yield curve and the expected rate of depreciation of a currency are functions of the same set of state variables: domestic and foreign short-term interest rates.

The model is estimated for two different currency pairs: U.S. dollar–pound sterling and U.S. dollar–Canadian dollar. The dataset consists of monthly rates of depreciation[10] of these two currency pairs over the period January 1976 to December 2004, along with monthly observations of the corresponding U.S., British, and Canadian Eurocurrency interest rates for maturities of one, three, six, and twelve months. These Eurocurrency deposits are essentially zero-coupon bonds whose payoffs at maturity are the principal plus the interest payment.

The estimations are carried out using data over the period January 1976 to December 1997 in order to reserve the last seven years for an out-of-sample forecasting exercise. The exchange rate forecasts, in particular, are computed according to a recursive procedure: at each month *t*, the model is re-estimated using data up to and including that month, and then forecasts of the spot exchange rate, up to one year ahead, are obtained.

A "horse race" is conducted between the forecasts obtained using this no-arbitrage model and those generated by three alternative benchmarks: a random walk, a vector autoregression on the forward premiums and the rate of depreciation, and the forward-premium regression. A comparison of the author's forecasts with those produced by the random-walk model is motivated by the fact that the random-walk model is considered to be the usual metric by which to evaluate exchange rate forecasts since the original work of Meese and Rogoff (1983a, b). However, Clarida and Taylor (1997) show that if one uses a vector autoregression (VAR) on the forward premiums and the rate of depreciation, it is possible to obtain out-of-sample forecasts of spot exchange rates that beat the random-walk model. Therefore, a VAR model is also included as a second benchmark. Finally, and for completeness, the author also includes the forecasts produced by a standard ordinary least-squares regression of the rate of depreciation onto a constant and the lagged forward premium (the forward-premium regression).

The forecasts produced by the term-structure model, as well as those of the three competing models, are evaluated in terms of two widely used criteria: the root-mean-square error (RMSE) and the mean-absolute error (MAE). The smaller these criteria are, the better the performance of the model.

7. The absence of arbitrage opportunities will not only restrict the way in which interest rates and exchange rates move, but will also restrict how interest rates at different maturities move together.
8. In fact, there is empirical evidence that one can also improve interest rate predictions if such no-arbitrage restrictions are exploited (Duffee 2002; Ang and Piazzesi 2003).
9. For a review of affine term-structure models and their applications, see Piazzesi (2003).
10. Note that a negative rate of depreciation would imply an appreciation in the currency.

Results

The author finds that using no-arbitrage restrictions reduces, for example, the RMSE in forecasting the spot U.S. dollar–pound sterling rate by about 35 per cent at the one-year forecast horizon relative to the VAR approach, and by about 15 per cent for the U.S. dollar–Canadian dollar rate. The gains from using a VAR model over a random-walk model are negligible. For example, the gain at the one-year horizon for the U.S. dollar–pound sterling pair is only 2.4 per cent (versus the 40 per cent reported by Clarida and Taylor 1997). Similar results are obtained when using the MAE criteria.

Conclusions

Overall, these results support the use of no-arbitrage methods to generate more accurate exchange rate predictions. The success of this approach provides indirect support for the assumption that markets are efficient, since it is based on a generalization of covered interest rate parity. Still, more work can be done in this direction. The predictions in these models are based exclusively on the information contained in interest rates, while one would also like to use the information contained in other macroeconomic variables (such as output growth, inflation, or even commodity prices) to obtain even better predictions. Developing a no-arbitrage model of the joint behaviour of macroeconomic variables, interest rates, and exchange rates that, at the same time, is able to deliver good exchange rate forecasts is a new challenge that is left for further research.

References

Ang, A. and M. Piazzesi. 2003. "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables." *Journal of Monetary Economics* 50: 745–87.

Bailliu, J. and M.R. King. 2005. "What Drives Movements in Exchange Rates?" *Bank of Canada Review* (Autumn): 27–39.

Berg, A., E. Borensztein, and C. Pattillo. 2004. "Assessing Early Warning Systems: How Have They Worked in Practice?" IMF Working Paper No. 52.

Clarida, R. H., L. Sarno, M.P. Taylor, and G. Valente. 2003. "The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond." *Journal of International Economics* 60: 61–83.

Clarida, R.H. and M.P. Taylor. 1997. "The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors." *Review of Economics and Statistics* 79: 353–61.

Cochrane, J. 2001. *Asset Pricing*. Princeton, N.J.: Princeton University Press.

Diez de los Rios, A. 2006. "Can Affine Term Structure Models Help Us Predict Exchange Rates?" Bank of Canada Working Paper No. 2006-27.

Duffee, G.R. 2002. "Term Premia and Interest Rate Forecasts in Affine Models." *Journal of Finance* 57: 405–43.

Hansen, L.P. and R.J. Hodrick. 1980. "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis." *Journal of Political Economy* 88: 829–53.

Mark, N.C. 2001. *International Macroeconomics and Finance: Theory and Econometric Methods*. Malden, Mass.: Blackwell Publishers.

Meese, R.A. and K. Rogoff. 1983a. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics* 14: 3–24.

———. 1983b. "The Out-Of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?" In *Exchange Rates and International Macroeconomics*, J.A. Frenkel (ed.), 67–109. Chicago: University of Chicago Press, for National Bureau of Economic Research.

Piazzesi, M. 2003. "Affine Term Structure Models." Forthcoming in *Handbook of Financial Econometrics*, Y. Aït-Sahalia and L.P. Hansen (eds.).

Ragan, C. 2005. "The Exchange Rate and Canadian Inflation Targeting." *Bank of Canada Review* (Autumn): 41–50.