![Bank of Canada / Banque du Canada logo]

BANK OF CANADA
BANQUE DU CANADA

Bank of Canada Review

Autumn 2011

MEMBERS OF THE EDITORIAL BOARD

The *Bank of Canada Review* is published four times a year under the direction of an Editorial Board, which is responsible for the editorial content. The contents of the *Review* may be reproduced or quoted provided that the *Bank of Canada Review*, with its date, is specifically quoted as the source.

Notice to Subscribers

This issue of the** Bank of Canada Review **is the final print version of the publication.

Readers of the *Bank of Canada Review* can continue to access the publication electronically on the Bank of Canada website. Subscribers can register to be notified of newly published content through RSS feeds (**http://www.bankofcanada.ca/rss-feeds/**) or, alternatively, can receive an email alert of each new release (see **http://www.bankofcanada.ca/email-alerts/** to register).

Back issues of the *Review* and other publications are available on the Bank's website at **http://www.bankofcanada.ca**.

For further information contact:

Publications Distribution Services
Communications Department
Bank of Canada
Ottawa, Ontario K1A 0G9
613 782-8248
1 877 782-8248 (toll free in North America)
publications@bankofcanada.ca

Inquiries related to interest rates or exchange rates should be directed to 613 782-7506.

Bank of Canada Review

Autumn 2011

Early Chinese Coinage

David Bergeron, Curator, Currency Museum

The basic shape of coinage has remained largely unchanged for more than 2,500 years. With few exceptions, coins have been round, flat and compact, no matter the period. But China's earliest coins were not round. Instead, they imitated the commodities and tools that were part of a primitive barter system. These early forms of coinage appeared from about the sixth to the fourth century BC in what is now northeastern China during the Zhou Dynasty. The powerful Zhou Dynasty (1100 BC–256 BC) had ruled the entire area, but by 700 BC its power had waned. The state of Zhou was merely the central hub of government for numerous independent feudal states. The production and issue of coinage was very much a local affair.

Cowrie shells were China's earliest medium of exchange. Imitations made of bone, stone and bronze circulated in the state of Chu, located between the Yangtze and Yellow rivers (Jiangsu and Anhui provinces). Most likely introduced because of a shortage of real cowries, these early coins were known as "ghost-face" or "ant-nose" money because of their markings. Since the weight of these imitation cowries was inconsistent, their value probably equalled that of a real cowrie shell. Both genuine and imitation cowries used in trade are about the size of a kidney bean.

The provinces of Shaanxi, western Shanxi and Henan used hollow-handled spades as coinage. These were the first replicas of tools to be used as money, and prototypes marking their transition from barter items to money date as far back as 1200 BC. The hollowed handle and long pointed feet required complex moulds, and the resulting coins were very fragile. To increase durability and allow easier casting of large quantities, the profile was simplified and the handle flattened. These coins were called *jin*, meaning hoe, because of their familiar shape. The shape and markings on the coins varied, depending on the location of production and the issuing authority. During their use, there were many different types of hoe money in circulation, but their weights and dimensions remained consistent within a specific area and from one issue to the next.

Knives were also copied and cast as coins in ancient China. Unlike spade and hoe money, the earliest examples were cast in a way that allowed their use as real knives. For practicality and portability, however, the edges were never sharpened. Knife money (called *dao bi*, meaning "pointed" or "knife-shaped") was used mainly in the provinces of Hebei, eastern Shanxi, Shandong, Liaoning, Jilin and Heilongjiang. Like hoe money, the style of knife money varied in time and place: older blades were larger and curved, later issues were smaller and angular; blades from the state of Qi were pointed, whereas those from the city of Ming had a blunt tip.

During the fourth century BC, some of the independent feudal states began replacing cowrie, hoe and knife money with round coins. These early coins, cast with a distinctive round hole in the centre, varied in size and weight.

In 221 BC, the ruler of the western state of Qin, having defeated all the other rulers, named himself China's first emperor. Qin Shi Huangdi united the country and reformed China's coinage, eliminating the circulation of cowrie, hoe and knife money and issuing a new round cast-bronze coin with a square hole called *ban liang* (half *liang* or Chinese cash). Coins of this shape and size, similar to that of a Canadian quarter, would continue to be cast in China until 1912.

The round coin, imitation cowries, and spade and knife money pictured on the cover are part of the National Currency Collection of the Bank of Canada.

Photography by Gord Carter.

Contents

The International Monetary System: An Assessment and Avenue for Reform

Eric Santor, International Economic Analysis Department, and Lawrence Schembri, Adviser

- *The international monetary system comprises the policies and official arrangements related to the international balance of payments, in particular, exchange rate arrangements.*

- *The current system is in need of reform. It has not facilitated the timely and symmetric adjustment in the real exchange rate necessary to accommodate the integration of China and other emerging-market economies into the global economy. This lack of adjustment contributed to the global financial crisis and recession, and is hindering the global recovery because it is forestalling the required rotation of global demand.*

- *The G-20 countries should co-operate to promote the system's transition to one in which all systemically important countries and currency areas have market-determined flexible exchange rates supported by coherent macroeconomic and financial sector policy frameworks to ensure that global growth is strong, sustainable and balanced.*

The current international monetary system (IMS) has been described as a "non-system" or "hybrid," owing to the lack of a coherent set of exchange rate policies among systemically important economies. In practice, the system has not been able to adjust efficiently to large shocks, such as the integration of China into the global economy, thus allowing the occurrence of large and unsustainable current account imbalances. Indeed, many observers attribute the 2007–09 financial crisis and the current weak recovery of the global economy, in part, to the system's inherent instability—in particular, to the lack of timely and symmetric adjustment of real exchange rates to these imbalances.[1] The authorities in many emerging-market economies (EMEs), in particular, have acted to constrain capital flows and exchange rate movements, thereby preventing a necessary rotation of global demand.

These failings have not gone unnoticed—the Group of 20 (G-20) has identified the need to reform the IMS (G-20 2010). Many proposals have been put forward to address the current system's ubiquitous flow and stock imbalances, including an expanded role for Special Drawing Rights (Williamson 2009) and the promotion of alternative reserve currencies to the U.S. dollar (Zhou 2009). But such proposals are simply coping mechanisms, designed to maintain the status quo, rather than effective solutions for the system's inherent problems (Carney 2009). Broader proposals, such as moving to a single world reserve currency—with the International Monetary Fund (IMF) as lender of last resort—are patently impractical. Missing from the current debate is a well-articulated vision for an efficient and resilient IMS—one consistent with both the individual and collective best interests of all countries. Also absent

---

[1] See Fischer (2009), Carney (2009), Rajan (2010), Obstfeld and Rogoff (2009) and Banque de France (2011).

are the strategies needed to move toward such a vision from the current predicament.

This article begins with an assessment of the existing IMS, noting its strengths, but also highlighting its ongoing weaknesses. We then propose an avenue for the reform of the IMS, in which all systemically important countries and currency areas adopt market-based and convertible floating exchange rates supported by appropriate monetary, fiscal and financial sector policies.[2, 3] We also explore the roles of the G-20 countries and major international financial institutions in facilitating this transition.

An IMS in which flexible exchange rates predominate would be much more resilient and stable since it would allow more timely and symmetric, and thus more efficient, adjustment to shocks, thereby precluding persistent external imbalances and large accumulations of reserves.[4] These changes would facilitate the transformation of the IMS into a more market-based and decentralized system that would better accommodate the emerging multi-polar global economy.

Characterizing the International Monetary System

The IMS comprises the policies and official arrangements related to the international balance of payments (Carney 2009; Lipsky 2010). Specifically, the IMS consists of arrangements for (i) exchange rates, (ii) current payments and capital flows, and (iii) international reserves. It is also (iv) a collection of institutions, rules, standards and conventions that govern its operation.

Exchange rates

The current IMS has been characterized as a hybrid or non-system, because the systemically important
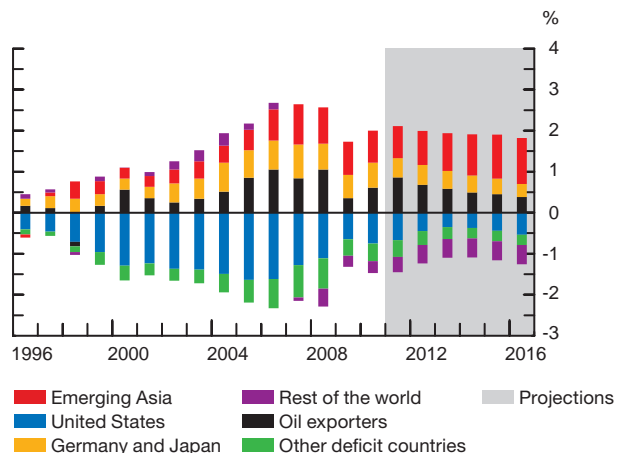
countries do not share the same exchange rate regime or nominal anchor. Roughly two-thirds of the 40 largest countries in the world (measured by market-based GDP) have floating exchange rates, and one-third have exchange rates that are managed or fixed, representing roughly three-quarters and one-quarter of global GDP, respectively. Until recently, the number of countries with floating exchange rates had been increasing, but this trend has stopped, if not reversed.

Current payments and capital flows

Most advanced countries have convertible currencies and open capital accounts. Practices among EMEs differ, but until the 2007–09 crisis, the trend had been to remove controls and liberalize these payments and capital flows.
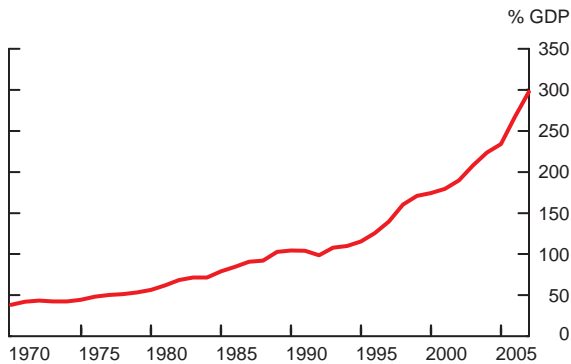
Since 2000, these arrangements have contributed to the buildup of persistent and large current account surpluses and deficits, as a percentage of both global and national GDP (**Chart 1**). For example, the U.S. current account deficit peaked at almost 6 per cent of GDP in 2006, and Chinese surpluses at 10.1 per cent of GDP in 2007. While the financial crisis led to a reduction in these imbalances, this was largely the result of the cyclical downturn in advanced economies, which depressed their demand for imported goods and services. Current account imbalances are therefore expected to persist over the coming years as these economies recover. Even more striking is the rapid increase in capital flows

Chart 1: Global imbalances
Current account as a percentage of global GDP

Sources: Board of Governors of the U.S. Federal Reserve and IMF staff estimates

---

2 Fiscal policy must be sustainable and therefore coherent with monetary and exchange rate policies. Indeed, the overall coherence of medium-term frameworks for exchange rate and monetary, fiscal and financial sector policies is critical for maintaining internal and external stability.

3 In this vision, small economies, for which the advantages of an independent monetary policy are outweighed by transactions costs in both trade and finance, would have a permanent fixed exchange rate, either a hard peg or a common currency.

4 Historically, there are several examples of large and persistent current account imbalances being driven by market forces and having a positive welfare impact; for example, large flows of foreign direct investment into Australia's resource sector and into Canada's infrastructure at the end of the nineteenth century. It is important to note, however, that these imbalances were accompanied by significant appreciations in the real exchange rate.

Chart 2: Total G-20 gross foreign assets and liabilities

% GDP

Note: Excludes Saudi Arabia from calculation of the G-20 total
Sources: IMF *International Financial Statistics*, Lane and Milesi-Ferretti (2007) and authors' calculations
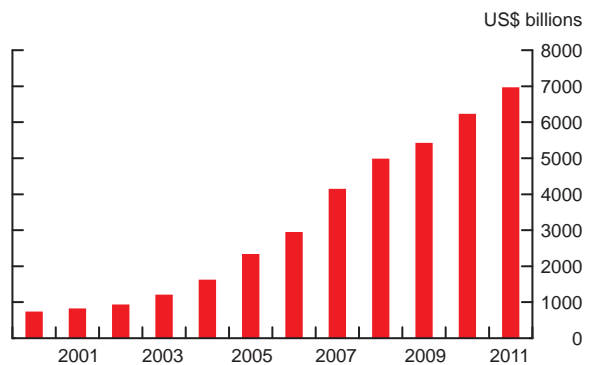
Last observation: 2007

Chart 3: Reserves in emerging and developing economies

US$ billions

Source: IMF *International Financial Statistics*   Last observation: June 2011

Chart 4: BRIC reserves

% GDP

■ Brazil   ■ India   ■ Russia   ■ China

Source: IMF *International Financial Statistics*

and gross holdings of foreign assets and liabilities. From 1970 to 2007, gross foreign assets and liabilities increased from 40 per cent of GDP to nearly 300 per cent of GDP for G-20 countries (**Chart 2**). Moreover, changes in gross foreign assets and liabilities have swamped movements in the current account. This dramatic rise in gross flows, which reflects the increase in financial globalization, is one of main structural changes in the IMS in recent years. These closer financial links among countries facilitated the transmission and magnified the impact of the financial crisis.
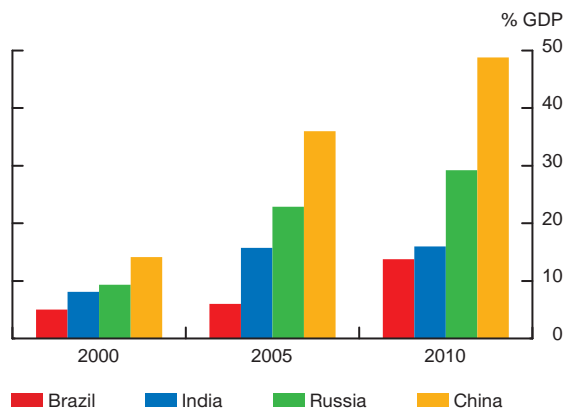
International reserves

Many advanced economies with a flexible exchange rate have reduced their holdings of reserves as a percentage of GDP over the post–Bretton Woods era. In contrast, a notable feature of the IMS over the past decade has been the rapid accumulation of reserves in emerging economies, concurrent with the increase in current account imbalances. From 2000 onward, these reserves have risen from less than US$1 trillion to almost US$7 trillion (**Chart 3**), much of which is invested in U.S. government debt. By most metrics, this buildup far exceeds that needed for precautionary purposes. For example, reserves have increased significantly as a percentage of GDP for the BRIC economies (Brazil, Russia, India and China) (**Chart 4**). Substantial and persistent current account imbalances and the resulting extraordinary accumulation of reserves speak directly to the lack of timely and symmetric adjustment of real exchange rates in the IMS.

Institutions

The IMS is overseen by a complex and evolving set of institutions that seek to establish and promote compliance with a wide variety of rules, standards and conventions in terms of both macroeconomic and financial sector policies (**Figure 1**).[5] The overarching goal of this global financial architecture is to preserve global financial and monetary stability. The key institutions that monitor and oversee the system—the IMF, the Bank for International Settlements (BIS), the Financial Stability Board (FSB) and the G-20—identified many of the risks that subsequently materialized during the 2007–09 crisis. But they failed to appreciate the true magnitude of

---

[5] The choice of an exchange rate regime is a sovereign decision of the individual country. The IMF and other institutions can only exercise surveillance over a country's policies and make non-binding policy recommendations.

	← Less exchange rate flexibility		More exchange rate flexibility →	
Exchange rate regime	Gold standard →	Bretton Woods →	Hybrid: Fixed to flexible →	Market based: Flexible exchange rates with inflation targeting
Institutions that comprise the global financial architecture	Central banks (BIS)	IMF World Bank	G-7/G-20 FSB →	
Capital controls	Capital mobility	Widespread capital controls	Managed capital flows conditional on exchange rate regime	Capital mobility
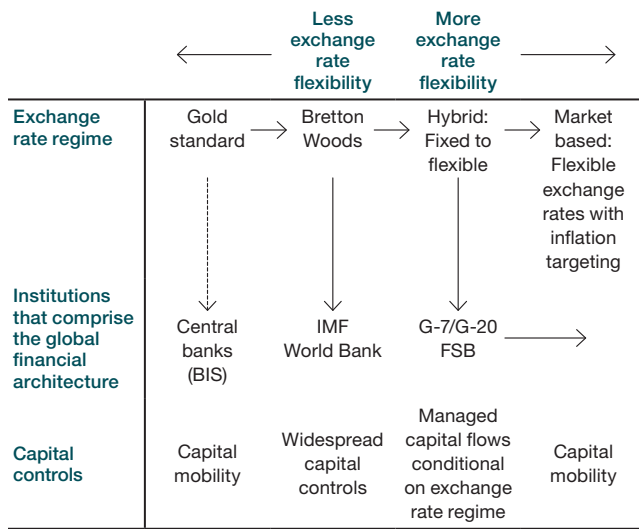
the risks and their consequences if realized, or to gain the political traction necessary to reform members' policies and allow adjustment to external imbalances. They were thus ineffective in preventing the buildup of the macroeconomic and financial vulnerabilities that contributed to the crisis.[6]

Assessing the System: Some Strengths, but Many Weaknesses

Strengths

The current IMS has facilitated an enormous expansion of global growth, trade and financial integration: since the 1970s, annual growth in global GDP has averaged more than 3 per cent, global trade has increased by nearly double the rate of GDP, and gross foreign assets and liabilities by more than three times. Most importantly, this expansion has included the integration of China and India—nearly one-third of the world's population—into the global economy. Between 1980 and 2010, China's economy rose from the world's twelfth largest to the second largest, as its size increased more than twelvefold. Globalization, particularly in the form of trade and foreign direct investment, has allowed China not only to benefit from access to markets, technology transfer and increased specialization, but also to realize its comparative advantage in producing labour-intensive manufactured goods. Despite the

banking, sovereign debt and currency crises that the IMS has experienced since the breakdown of the Bretton Woods system, it has generally functioned well in supporting increased trade and capital flows.

Weaknesses

In theory, the fact that countries have different nominal anchors for their monetary policy frameworks (price stability versus exchange rate stability) is not necessarily a problem for the stability of the IMS, as long as market-based adjustment of real exchange rates can take place in response to shocks, either via movements in the nominal exchange rate or through changes in domestic wages and prices. Without a commitment to allow real exchange rate adjustment, the current IMS is plagued by two fundamental and pervasive asymmetries.

Under fixed exchange rates, surplus countries can thwart the conventional adjustment mechanism with sterilization

First, there is no effective channel through which market pressures can, in a timely manner, force countries with managed or fixed exchange rates and with balance of payments surpluses to permit equilibrating adjustment in their real exchange rates. In particular, under fixed exchange rates, surplus countries can thwart the conventional adjustment mechanism by sterilizing the impact of the balance of payments surplus on the money supply. Sterilization involves central bank sales of government bonds or central bank bond issuance to the private sector (or to a state-owned or -controlled banking sector). This is typically done in the presence of capital controls, which essentially prevent the private sector from undoing the central bank's sterilized foreign exchange intervention.[7]

Second, a unilateral depreciation of the real exchange rate is very difficult to achieve under a fixed exchange rate regime, because domestic prices and wages are generally less flexible downward than upward. Depreciation consequently requires large and painful gaps in output and employment, as have been experienced in Greece and Spain in the aftermath of the crisis.

---

6 The IMF's Independent Evaluation Office (2011) notes many of the shortcomings in Fund surveillance during the lead-up to the crisis.

7 In a similar vein, the reserve ratio for banks can be increased to absorb the excess liquidity resulting from exchange rate interventions. For instance, China has increased its reserve ratio from 17 per cent to 21.5 per cent over the past year.

These two effects contribute to one of the major weaknesses of the IMS—the lack of symmetric adjustment. Simply, surplus countries can thwart adjustment far longer than deficit countries, and thus adjustment is often borne disproportionately by the latter, with harsh consequences. When it occurs on a global scale, the lack of symmetric adjustment can lead to deficient global demand.

The increasing proportion of the global GDP in the hands of countries that have managed or fixed exchange rate regimes and capital controls, and that are accumulating and sterilizing reserves, represents an unprecedented threat to the stability of the IMS and the global economy. In fact, countries whose currencies comprise more than 50 per cent of the U.S. real effective exchange rate (REER) are actively thwarting adjustment. To gauge the effect of these interventionist policies, it is useful to compare their impact on real exchange rates by comparing the movements in the REER of East Asian EMEs that have largely maintained this set of policies with those of Latin American EMEs that have largely embraced flexible exchange rates, open capital accounts and inflation targeting. While the Latin American EMEs have experienced a significant appreciation in their real exchange rates, as the Balassa-Samuelson hypothesis would predict, the real exchange rates of the Asian EMEs have remained relatively unchanged (**Chart 5**). As countries thwart adjustment in their exchange rates, that adjustment can be displaced onto countries with open capital markets and floating exchange rates (**Chart 6**).[8]

This lack of symmetric adjustment has caused large and persistent imbalances in trade and in current accounts, as well as unbalanced economic growth. For example, China, as well as other EMEs that compete with China in third markets, has tried to forestall REER adjustment to its current account and balance of payments surpluses. China has resisted growing international pressure to increase the flexibility of its heavily managed exchange rate, which has allowed it to maintain an undervalued exchange rate and to promote export-led employment and output growth. China and other countries in similar circumstances are frustrating adjustment in their real exchange rates by accumulating reserves and sterilizing the impact on the domestic money supply. In China, the resultant financial repression and disintermediation, caused by the banking sector's absorption of central bank

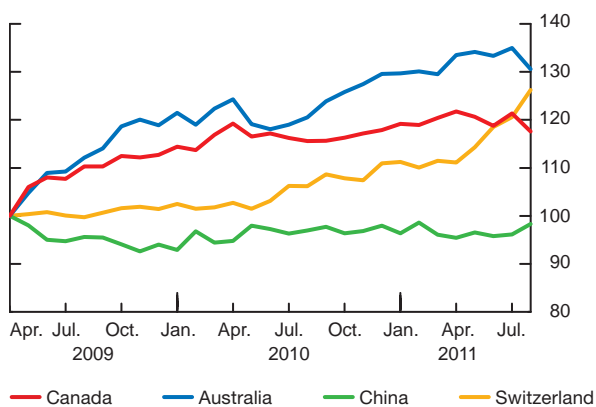**Chart 5: Real effective exchange rates**
Index: 1990=100

■ Latin America　■ East Asia
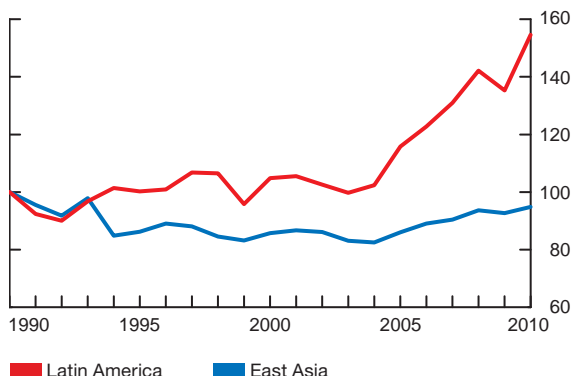
Sources: JP Morgan, IMF *World Economic Outlook* and authors' calculations　　　　Last observation: 2010

Chart 6: Displaced adjustment in real effective exchange rates
Index: April 2009=100

■ Canada　■ Australia　■ China　■ Switzerland

Source: BIS　　　　Last observation: August 2011

sterilization bonds, precludes efficient intermediation, leading many Chinese, especially the owners of small and medium-sized enterprises, to boost their savings to self-finance their investments (**Chart 7**).

These structural imbalances contributed to the financial crisis by channelling excess global savings into the U.S. capital market and exposing regulatory and supervisory gaps in the U.S. financial system. These flows set off a search for yield and created undue pressure to generate highly rated assets (Bernanke 2011), partly by creating incentives to exploit cracks in the regulatory framework and undertake destabilizing regulatory arbitrage across domestic and international regulatory jurisdictions (Rajan 2010; Obstfeld and Rogoff 2009).

8　Although the displaced adjustment seen in Chart 6 was not the sole factor behind the appreciation of the Canadian, Australian and Swiss currencies, it had an impact.

Chart 7: Excess Chinese savings

% GDP

■ Households	■ Companies	■ Government
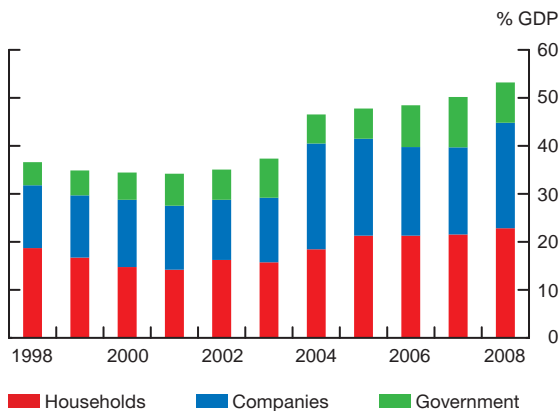
Note: Companies include non-financial corporations and financial institutions
Source: National Bureau of Statistics of China   Last observation: 2008

There is evidence that these flow imbalances are now re-emerging because real exchange rates have remained relatively unchanged. These flow imbalances are causing an ever-increasing stock imbalance. EME reserves, which are enormous, are largely invested in U.S.-dollar assets. Consequently, they represent a substantial source of vulnerability in the global economy because of their sheer size and lack of diversification. Reserve holders would incur significant capital losses if the U.S. dollar were to depreciate.[9]

To address the flow and stock imbalances that threaten the global economy, the IMS needs to be reformed. In the next section, we outline our vision for the IMS and make the case for its adoption.

Renewing the International Monetary System

Given the pervasive imbalances, currency tensions and vulnerabilities, there is a strong need to reform the international monetary system to permit more symmetric and timely exchange rate adjustment, especially among systemically important countries. Such an adjustment would facilitate the equilibration of trade and capital flows and prevent the re-emergence of large external imbalances.

---

[9] In addition, this enormous reserve accumulation is creating a modern version of the Triffin dilemma, as countries accumulate more and more assets of lower and lower quality, thus reducing the probability that they will be paid back in full. This accumulation also increases the likelihood that their currency will eventually appreciate, causing capital losses on their holdings of foreign-currency assets. See Gourinchas, Rey and Truempler (2011) for a fuller discussion.

The various configurations of the IMS over the past 140 years have not managed to produce a system that is resilient to large shocks over the long run. **Figure 1** depicts the evolution of the IMS since 1870, from the gold standard and gold-exchange standard, to the Bretton Woods system (fixed, but adjustable, exchange rates, based on the U.S. dollar as the international reserve currency), and to the current "hybrid" system. The IMS is not static: the current system represents an evolution along a continuum of possible systems.

The respective configurations of the IMS have one common feature: in each case, fixed exchange rate regimes were unable to withstand large shocks. Since surplus countries can thwart exchange rate adjustment far more easily than deficit countries, the latter often bear a disproportionate share of the adjustment. The history of the IMS also shows that the system functions only if systemically important countries follow policies that are consistent with the system. For example, in principle, the gold-exchange standard should have allowed for symmetric adjustment; however, throughout the 1920s, France and the United States, as surplus countries, prevented adjustment through sterilized intervention (pushing deflation onto the United Kingdom, the main deficit country). Similarly, symmetric adjustment failed to occur in the 1960s during the Bretton Woods era and over the past decade. The main lesson from history is that it is not the choice of reserve asset that matters, but whether countries follow policies that permit real exchange rate adjustment.

A new vision is therefore required, one that embodies the following characteristics:

- promotes timely and symmetric adjustment to shocks,
- prevents large external imbalances and crises,
- provides sufficient global liquidity, and
- maintains confidence in the system.

The last two characteristics are critical for achieving global economic stability in the context of increasing financial globalization. The IMS also needs to be supported by an appropriately constituted set of international institutions, which comprise the global financial architecture that can effectively promote the desirable characteristics listed above.

To achieve an IMS in which timely and symmetric exchange rate adjustment is market based and thus inherent to the system, all systemically important countries and currency areas should have flexible exchange rates in conjunction with a monetary policy

framework consistent with achieving price stability—normally defined as low, stable and predictable inflation. Such a monetary policy framework must itself be underpinned by a fiscal policy that ensures sustainable public finances and therefore preserves the central bank's ability to achieve price stability, and by a financial sector policy that promotes financial system stability via resilient financial institutions and markets to ensure the efficient intermediation of savings and the effective transmission of monetary policy.

Evidence from a number of countries has shown that a flexible, market-determined exchange rate adjusts quickly in response to economic shocks, thereby acting as a "shock absorber" by mitigating the impact of the shock on real economic activity and welfare.[10] It is important to note that this adjustment is symmetric: it occurs for both negative and positive shocks, and it occurs for shocks to both assets and goods markets. A flexible exchange rate thus increases the resilience of the domestic economy to both external and internal shocks.

Flexible exchange rates,
by themselves,
are not a panacea

Flexible exchange rates, by themselves, are not a panacea, however. Indeed, the experience from the period immediately after the collapse of the Bretton Woods system clearly indicates that, in the absence of a domestic nominal anchor for inflation expectations supported by a coherent framework for macroeconomic and financial sector policy, flexible exchange rates can be volatile and display overshooting behaviour.[11]

Although it took almost 20 years after the collapse of the Bretton Woods system, public authorities eventually recognized that a well-functioning, flexible exchange rate regime must be supported by central banks with a mandate of price stability and by fiscal authorities that allow central banks the operational independence to pursue this goal. As a consequence, the volatility and dramatic misalignments

of exchange rates among Western economies that occurred in the 1970s and 1980s declined.

The adoption of formal inflation targets by New Zealand and Canada in 1990 and 1991, respectively, has led to broader acceptance of this monetary policy framework by an expanding set of countries that includes the United Kingdom, Sweden and Norway, as well as South Africa, Israel, the Czech Republic and most countries in Latin America.[12] Indeed, the widespread adoption of monetary policy frameworks consisting of a flexible exchange rate with a price-stability mandate ushered in the era of the "Great Moderation," as the combination of shock-absorbing flexible exchange rates and well-anchored inflation expectations provided a solid basis for strong economic performance in many of these countries.

Rose (2007) and Mihov and Rose (2008) obtain three important results concerning the monetary policy framework of a formal inflation target and a flexible exchange rate that support our proposed vision for the IMS. First, they find that such a framework reduces exchange rate volatility because it provides a credible anchor for inflation expectations. Second, and of particular importance for EMEs, they find that countries with this framework experience fewer "sudden stops" or reversals in capital flows. A flexible exchange rate adjusts to mitigate the volatility of capital flows by appreciating (depreciating) when net inflows are positive (negative). It also provides two-way risk to domestic borrowers and foreign investors, thus reducing moral hazard. Unlike a pegged exchange rate, a flexible rate does not offer a promise, either explicit or implicit, to maintain the exchange rate unchanged. The exchange rate risk must therefore be borne by the parties involved, providing them with an incentive to manage their foreign exchange exposures prudently or hedge their exposures. This provides an impetus for the development of hedging markets and eliminates the need for capital controls. Third, they find that the durability of this monetary policy framework is historically unprecedented: it has been in place for almost 20 years, and no country has abandoned an inflation-targeting framework.

The currency crises of the 1990s in Europe, Mexico, Russia and East Asia spurred the wider adoption of this monetary policy framework. These crises demonstrated the difficulty of maintaining pegged

---

10 See, for example, Murray, Schembri and St-Amant (2003), for a recent study of Canada, and Broda (2004) for a study of a set of developing countries.

11 The most notable example occurred during the Reagan-Volcker era in the early to mid-1980s, when the U.S. dollar appreciated almost 40 per cent on a real effective basis and then depreciated by almost the same absolute amount.

12 In addition, while some countries or regions, such as the United States, Japan and the euro area, do not have explicit inflation targets, their central banks are independent and have a price-stability mandate.

exchange rate regimes in the absence of widespread capital controls. Such "soft" pegs proved to be subject to self-fulfilling speculative attacks because they offered attractive one-way bets to speculators who knew that politicians were not willing to sacrifice domestic economic goals to preserve the pegged rate.

From this negative experience with soft-pegged exchange rates and the contrasting positive experience with flexible exchange rates came the bipolar view of exchange rate regimes (Obstfeld and Rogoff 1995; Fischer 2001). The bipolar view maintained that countries had two choices when it came to exchange rate regimes: adopt either a hard, fixed exchange rate regime (such as a common currency, as among euro-area countries; the use of a foreign currency, such as the dollar or euro, as with Ecuador or Bosnia; or a currency board, as in Bulgaria) or a flexible exchange rate regime with a monetary policy goal of price stability.[13]
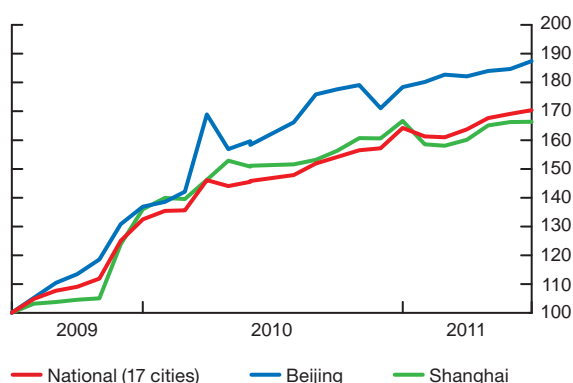
While the bipolar view provided a coherent vision for the IMS, it did not address transitional issues, especially with respect to financial market development. Many countries were not prepared to jump to this vision immediately, and the lack of articulation and consensus surrounding the evolution of the policy frameworks proved to be a serious obstacle. Moreover, in the face of instability in the IMS, many EMEs felt compelled to maintain the status quo. However, moving toward a more flexible market-based exchange rate can help alleviate the distortions created by a heavily managed, undervalued exchange rate.[14] In particular, if the exchange rate were allowed to appreciate gradually, it would boost domestic purchasing power, which would increase domestic consumption, and raise imports, since they would become relatively less expensive. Concerns that such a move would have a large negative impact on export competitiveness are often overstated, because the increase in purchasing power would mitigate wage demands. Furthermore, domestic prices and wages would adjust to reduce the impact of any appreciation. Finally, increased exchange rate flexibility would increase monetary policy autonomy, allowing domestic authorities more scope to control rising inflation without resorting to altering reserve ratios

Chart 8: House prices in China
Soufun Index, July 2009 = 100

— National (17 cities)　　— Beijing　　— Shanghai

Sources: Soufun, Bank of Canada and authors' calculations
Last observation: June 2011

and other types of credit controls, which are generally less effective. In the absence of nominal exchange rate flexibility, adjustment of the real exchange rate would tend to occur through rising prices and increasing inflationary pressures as market forces eventually assert themselves, as is the case in China. Higher inflation could, however, undermine financial stability, because asset prices would also increase. For example, housing prices in China have risen more than 60 per cent since July 2009 (**Chart 8**).

In asset markets, increased exchange rate flexibility would reduce the need for the rapid accumulation of reserves and sterilization. Thus, banks, in turn, could redirect deposits away from purchases of sterilization bonds and toward loans to households and firms, thereby reducing the pressure on them to self-finance with excess savings. Greater exchange rate flexibility would also create an incentive for the development of financial markets: first to hedge exchange rate risk, and second to create a short-term money market or government securities market to serve as a basis for broader capital-market activities. Such a market would allow the central bank to eventually adopt a short-term interest rate instrument for monetary policy—a more efficient way of influencing monetary conditions than the existing quantitative instruments or controls, namely, the reserve ratio.

As financial markets and banking activities develop, capital controls can be loosened, but again, the flexible exchange rate would help stabilize capital flows since it would tend to appreciate (depreciate) when there are net inflows (outflows), reducing the incentives for such flows. Many Latin American countries

---

13 Note that a hard, fixed exchange rate effectively defines the domestic monetary policy framework in the context of capital mobility because maintaining the fixed rate essentially becomes the goal of monetary policy.

14 Chinese Premier Wen Jiabao stated, "In the case of China, there is a lack of balance, coordination and sustainability in the economic development" (Wen 2010).

have experienced an expansion and deepening of their financial markets, including the development of local-currency bond markets, following the adoption of a more flexible market-based exchange rate. This policy has been underpinned by strong fiscal and monetary policy frameworks, as well as robust financial sector regulation and supervision. This gradual path to liberalization and financial market development would also have substantial benefits for China.

Given China's leadership position in East Asia, among EMEs and in the G-20, the movement toward a flexible exchange rate regime would help to promote widespread adoption of this regime and would have important positive effects on the global economy. Reform of the global financial architecture would also help countries pursue this avenue of policy reform.

Renewing the Global Financial Architecture

As a result of the crisis, comprehensive plans are being put in place to overhaul domestic and international financial structures. Although this reform process is ongoing, much has been accomplished. The G-20 succeeded the G-7 as the premier economic forum. With its broader representation, which includes the major EMEs, and the regular participation of leaders, the G-20 can provide the leadership and legitimacy to increase the political traction of economic reform.

The G-20 can provide the leadership and legitimacy to increase the political traction of economic reform

The FSB superseded the Financial Stability Forum. In addition to the name change, its membership has expanded from the G-10 to beyond the G-20 to include other countries with significant financial sectors, such as the Netherlands and Singapore. The FSB has been given a clear mandate to take the lead on financial sector reform in several areas, including Basel III standards for bank capital and liquidity, shadow banking, financial market infrastructure and systemically important financial institutions. The representation of the IMF has also been renewed to better reflect the composition of global economic activity and thus give EMEs a larger voice and a greater stake in preserving global economic

and financial stability. The IMF has also been given more resources and has expanded its precautionary facilities to allow countries access to contingent funding should an adverse event occur. In general, all of the key global institutions have been strengthened and have become more representative of the global economy, with the intention that EMEs will take more responsibility for maintaining the stability of the IMS.

The global financial architecture that is emerging from the crisis has, now more than ever, the capacity to promote and support the transition of the current IMS to one with greater exchange rate flexibility, more timely and symmetric adjustment to external imbalances, and greater global financial stability. Broadly speaking, the G-20 leaders can provide direction, the IMF can offer significant analytic capacity, and the FSB can coordinate financial sector reforms with all of the key players at the table. The BIS, which provides a useful forum for co-operation between central banks, could also play an expanded role.

Because the international financial institutions cannot impose economic policies on sovereign countries, the G-20 has established the Framework for Strong, Sustainable and Balanced Growth to encourage its systemically important member countries to co-operate and move toward greater exchange rate flexibility and more sustainable macroeconomic and financial policy frameworks for their individual and collective benefit. To this end, the Framework incorporates a mutual assessment process of macroeconomic and financial sector policies to increase the global coherence of medium-term policy frameworks. Balanced growth implies that current account and fiscal positions should shrink to ensure that global growth is also sustainable. An important outcome of this process is a G-20 Action Plan to identify and coordinate the needed macroeconomic and financial reforms.

Concluding Remarks

The international monetary system is not working and needs reform. The lack of exchange rate adjustment not only led to the emergence of the substantial external imbalances in the United States and China, which contributed to the financial crisis of 2007–09, but is also forestalling the rotation of global demand needed to strengthen the global economic recovery. Indeed, the very weak and protracted recovery in the United States and Europe is aggravating their fiscal and financial problems.

Although the IMS has been plagued throughout its history with serious adjustment problems after major shocks, the precarious status quo is not an option. The G-20 countries and the major international financial institutions must take steps to achieve an IMS founded on coherent medium-term macroeconomic and financial policy frameworks that are consistent with strong, sustainable and balanced global growth. These frameworks would include increasingly flexible and more market-determined nominal and real exchange rates for all systemically important countries. To help the IMS advance to this desired end state, the institutions overseeing the global financial architecture will also need to become better governed, more coordinated and more effective.

Over the long term, as the global economy evolves into a multi-polar one with several regions of significant economic activity, flexible exchange rates embedded within strong macroeconomic and financial policy frameworks will help to ensure a smooth transition to a more resilient IMS—one that is better able to cope with the implications of increased financial globalization. In addition, having currencies that are convertible and market determined will facilitate welfare-enhancing competition among currencies to determine those that will have reserve-currency status, based, in part, on the strength of their policy frameworks.

Literature Cited

Banque de France. 2011. *Financial Stability Review: Global Imbalances and Financial Stability*. Report No. 15 (February).

Bernanke, B. 2011. "International Capital Flows and the Returns to Safe Assets in the United States, 2003–2007." *Financial Stability Review: Global Imbalances and Financial Stability* 15 (February): 13–26.

Broda, C. 2004. "Terms of Trade and Exchange Rate Regimes in Developing Countries." *Journal of International Economics* 63 (1): 31–58.

Carney, M. 2009. "The Evolution of the International Monetary System." Speech to the Foreign Policy Association, New York City, 19 November.

Fischer, S. 2001. "Exchange Rate Regimes: Is the Bipolar View Correct?" *Journal of Economic Perspectives* 15 (2): 3–24.

———. 2009. "Preparing for Future Crises." In *Financial Stability & Macroeconomic Policy*. Proceedings of the annual symposium held by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, 21 August.

Gourinchas, P.-O., H. Rey and K. Truempler. 2011. "The Financial Crisis and the Geography of Wealth Transfers." National Bureau of Economic Research Working Paper No. W17353. Available at <http://ssrn.com/abstract=1918682>.

Group of Twenty (G-20). 2010. "The G-20 Toronto Summit Declaration." June 26–27.

Independent Evaluation Office of the International Monetary Fund. 2011. *IMF Performance in the Run-Up to the Financial and Economic Crisis: IMF Surveillance in 2004–07*. Washington, D.C.: International Monetary Fund.

Lane, P. and G. M. Milesi-Ferretti. 2007. "The External Wealth of Nations Mark II: Revised and Extended Estimates of Foreign Assets and Liabilities, 1970–2004." *Journal of International Economics* 73 (November): 223–50. Available at <http://www.philiplane.org/EWN.html>.

Lipsky, J. 2010. "Reconsidering the International Monetary System." In *Macroeconomic Challenges: The Decade Ahead*. Proceedings of the annual symposium held by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, 28 August.

Mihov, I. and A. Rose. 2008. "Is Old Money Better than New? Duration and Monetary Regimes." *Economics* 2 (Special Issue): 1–24. Available at <http://www.economics-ejournal.org>.

Murray, J., L. Schembri and P. St-Amant. 2003. "Revisiting the Case for Flexible Exchange Rates in North America." *The North American Journal of Economics and Finance* 14 (2): 207–40.

Literature Cited (cont'd)

Obstfeld, M. and K. Rogoff. 1995. "The Mirage of Fixed Exchange Rates." *Journal of Economic Perspectives* 9 (4): 73–96.

———. 2009. "Global Imbalances and the Financial Crisis: Products of Common Causes." In *Asia and the Global Financial Crisis*, 131–72. Proceedings of a conference held by the Federal Reserve Bank of San Francisco, 19–20 October.

Rajan, R. 2010. *Fault Lines: How Hidden Fractures Still Threaten the World Economy*. Princeton: Princeton University Press.

Rose, A. 2007. "A Stable International Monetary System Emerges: Inflation Targeting Is Bretton Woods, Reversed." *Journal of International Money and Finance* 26 (5): 663–81.

Wen, J. 2010. "Consolidate the Upward Momentum and Promote Sustained Growth." Speech to the World Economic Forum Annual Meeting of the New Champions 2010, Tianjin, China, 13 September.

Williamson, J. 2009. "Understanding Special Drawing Rights (SDRs)." Peterson Institute for International Economics Policy Brief No. 09-11.

Zhou, X. 2009. "Reform the International Monetary System." People's Bank of China.

Liquidity Provision and Collateral Haircuts in Payments Systems

James Chapman, Financial Stability Department; Jonathan Chiu and Miguel Molico, Funds Management and Banking Department

- *Central banks play a pivotal role in well-functioning payments systems by providing liquidity via collateralized lending.*

- *Collateral haircuts limit the exposure of the central bank's balance sheet to credit, market and liquidity risk, and have potential policy implications because of their influence on financial market participants and on the relative liquidity of financial assets.*

- *Central banks must consider how their haircut policies affect the market's asset allocation, the relative price and liquidity of assets, and the likelihood of participants to default.*

- *Under extraordinary circumstances, there is a rationale for the central bank to temporarily reduce haircuts or broaden the list of eligible collateral to mitigate the shortage of liquidity in the market.*

To maintain monetary and financial stability, central banks may need to provide liquidity to financial institutions, financial markets and financial market infrastructure.[1] This includes the extension of intraday and overnight loans to promote a safe and efficient payments system. When extending these loans, the central bank typically requires collateral (assets pledged as security) to protect its balance sheet against the risk of default by the borrower (credit risk).[2] In addition, the central bank also applies a discount (haircut) to the market value of the instruments pledged, as protection against a decline in the value of the collateral (market risk) and a decline in the asset's marketability (liquidity risk).[3] Given the pivotal role of the central bank in the financial system, the design of its collateral policy is important. In particular, what should the size of haircuts be and how should that size vary under different financial system conditions?

Haircuts are typically chosen to protect lenders in the event of a borrower's default

This article examines how haircuts on collateral pledged to the central bank are determined. Haircuts are typically chosen to protect lenders against market and liquidity risk in the event of a borrower's

---

1. Financial market infrastructure includes exchanges, central counterparties, and payment and settlement systems. Under the Payment Clearing and Settlement Act, the Bank of Canada is responsible for the regulatory oversight of clearing and settlement systems with a view to controlling systemic risk.
2. A list of eligible securities and margin requirements for the Bank of Canada's Standing Liquidity Facility as of 7 September 2010 is available at <http://www.bankofcanada.ca/wp-content/uploads/2011/08/securities_160710.pdf>.
3. Note, however, that a central bank is better able to absorb liquidity risk than private agents, owing to its ability to create liquidity.

default. Recent research argues that the central bank should also consider the impact of its haircut policy on the risk-management incentives and asset-portfolio decisions of its counterparties and on the relative market liquidity of securities.[4, 5] These effects arise from the central bank's unique ability to create liquid settlement assets and because it is a major participant in the financial system, owing to the size of its balance sheet and the fact that it can undertake much larger market interventions than private sector institutions.

The article begins with a discussion of the central bank's role in providing liquidity in payments systems, as well as the role of collateral and haircuts in central bank lending. Particular institutional details are provided for the Canadian context. The next section focuses on the distinguishing features of a central bank's haircut policy relative to private sector practices. This is followed by a review of recent research on the determination of the optimal level of haircuts, using a model that explicitly incorporates the central bank's unique position in the payments system. The article concludes with an examination of the implications of this research for central bank haircut policy.

Liquidity Provision in Payments Systems

Financial institutions send payments to each other continuously throughout the day on their own behalf and on behalf of their clients. These payment flows arise from the underlying economic activities of payments-system participants and their clients. Although particular arrangements and institutional details differ from country to country and from system to system, central banks generally play an important role in the provision of liquidity to these payments systems and in their oversight. Central banks typically require collateral and face a fundamental trade-off between risk and liquidity when setting their haircut policies.

The Large Value Transfer System (LVTS) is the main settlement system in Canada. It is used to complete large-value, time-critical payments arising from the day-to-day business activities of households, corporations and financial institutions, and to settle Canadian-dollar obligations arising from securities and foreign exchange transactions.[6]

The Bank of Canada plays a central role in the day-to-day operation and oversight of the LVTS. Participants in the system use claims on the Bank (reserve accounts at the Bank of Canada) to settle their net payment obligations with other direct participants. The Bank also supplies liquidity to the system via collateralized intraday and overnight loans to LVTS participants (see **Box**). The Bank is also responsible for overseeing the operation and risk controls of the LVTS, given the LVTS's designation as a systemically important system.

Intraday finality and the Bank of Canada's residual guarantee

The LVTS processes payments in real time, and settlement of the system occurs on a multilateral net basis at the end of the day. During the day, processed payments are final and irrevocable. This immediate intraday finality is achieved because the Bank of Canada explicitly guarantees that settlement will occur, even in the event of multiple defaults.[7] This is facilitated through the use of collateral to secure participants' intraday net debit (negative) positions and by a residual guarantee provided by the Bank.

The Bank of Canada explicitly guarantees that settlement will occur, even in the event of multiple defaults

Since payments are not fully funded by senders' settlement balances or fully collateralized by the senders' and/or recipients' collateral holdings, this central bank guarantee of intraday finality provides an implicit intraday credit to system participants.

---

4 The idea that collateral policy can play an important role in influencing the development of financial markets has been known for some time. It has influenced Bank of Canada actions to promote the development of Canadian financial markets since the 1950s and has underpinned technical advice provided by the International Monetary Fund and the World Bank, among others, for many years.

5 Market liquidity refers to the ability to quickly sell an asset without causing a significant movement in its price.

6 Arjani and McVanel (2006) provide an overview of the structure of the LVTS and its relationship to the Canadian financial system. The LVTS is owned and operated by the Canadian Payments Association and is overseen by the Bank of Canada. See also the information contained in the backgrounder on the LVTS at <http://www.bankofcanada.ca/about/backgrounders/large-value-transfer-system-lvts/>.

7 If multiple participants default and if the collateral assigned to the system is insufficient to cover the total value of their obligations, the Bank of Canada would take possession of the available collateral and become an unsecured creditor of the defaulting institutions for the amount remaining after the collateral has been sold.

Collateral Requirements Under Tranche 1 and Tranche 2 of the LVTS

The Large Value Transfer System (LVTS) is a real-time, electronic wire transfer system that processes large-value, time-critical payments quickly and continuously throughout the day. It assures participants and their customers that, once a payment message has passed the system's risk-control tests, the transaction is settled on the books of the Bank of Canada on the same day, regardless of what happens to any of the participants afterward.

The sender of an LVTS payment can choose between two payment streams: Tranche 1 and Tranche 2. Each tranche has a corresponding risk-control limit.

A participant can send a Tranche 1 payment as long as the net amount that it owes (the balance of all its Tranche 1 payments sent and received) is less than the collateral that it has pledged to the Bank of Canada for Tranche 1 activity. Should the participant default during the course of that day, this collateral would be used to cover any net negative position in this payment stream. For this reason, Tranche 1 payments are known as "defaulter-pays."

Under Tranche 2, each participating institution begins the day by granting a bilateral credit limit (BCL) (which can be zero) to every other participating institution—the largest net exposure that it is prepared to accept from another participant on that day. These BCLs limit the negative position that each participant can take vis-à-vis other participants. In addition, each participant (as a sender) has a multilateral net debit cap, calculated as the sum of all BCLs extended to it, multiplied by a specified system-wide percentage set by the Canadian Payments Association. This multilateral cap limits the exposure of each participant in relation to the whole system and therefore the potential magnitude of any default. Each participating financial institution pledges collateral to the Bank of Canada in an amount equal to the largest BCL that it has extended to any other institution, multiplied by the specified system-wide percentage. If a participating institution fails, its collateral is seized and used to meet its obligation. If the defaulter's collateral is insufficient, the loss-allocation procedures distribute any remaining obligation proportionally among survivors on the basis of the BCLs granted by survivors to the failed institution. The survivors' obligation is not unlimited: it will never be larger than the amount of collateral they have pledged to the Bank. The collateral pledged by system participants is always sufficient to cover the failure of the institution with the largest possible amount owing to the system, i.e., the institution with the largest net debit cap. Tranche 2 has been described as "survivors-pay," since surviving financial institutions may be called upon to absorb the losses associated with a failure (after the defaulter's collateral is seized and used to meet its obligations). In the event of multiple defaults, the same process is used: first the defaulters' collateral is used, then the survivors face a loss allocation. If an obligation remains, the Bank will contribute the funds to settle the system under its residual guarantee.

Tranche 2 payments make up the great majority of the volume and value of payment transfers in the LVTS, principally because of savings in collateral relative to Tranche 1 operations. Nevertheless, participants who want immediate real-time settlement equivalency prefer Tranche 1 payments, even though they come with higher collateral costs.

The Standing Liquidity Facility

The Bank also facilitates settlement of the LVTS by routinely extending overnight collateralized loans through the Standing Liquidity Facility (SLF) to direct LVTS participants that are experiencing temporary shortfalls in their settlement balances at the end of the day. These advances provide participants with access to a reliable source of liquidity should they need to fund their end-of-day payment obligations, thus helping them to transfer payments efficiently among themselves during the day.[8] The interest rate on the overnight loan is set at the upper limit of the Bank of Canada's operating band for the overnight interest rate (the Bank Rate).[9]

Collateral services

The Bank establishes eligibility criteria for the various asset types that can be pledged intraday to collateralize bilateral credit limits or overnight through the SLF. It provides the LVTS system operator, the Canadian Payments Association, with valuations of the securities that are pledged as collateral. The Bank values and monitors the collateral pledged by LVTS participants on an ongoing basis.

The Role of Collateral and Haircuts in Central Bank Lending

Why lend against collateral?

Lenders require the pledging of assets to secure a loan, primarily to protect themselves against the risk of default by the borrower. As a public institution, a central bank requires collateral to protect its balance sheet against such risk.[10]

While small financial losses would likely have minor consequences, the negative effect of a significant loss on the central bank's balance sheet could potentially affect its ability to achieve its monetary policy and financial stability objectives. It could also undermine public confidence in the central bank. Any central bank financial loss from defaulted loans will have a negative impact on the balance sheet of either the central bank or the government.[11] While the exact timing of the loss can be altered by short-term portfolio and debt management (e.g., central bank sterilization or public debt issuance), the loss must ultimately be absorbed by some form of costly taxation or asset sales and would therefore affect the financial health of the consolidated balance sheet of the government and the central bank and, accordingly, the economy.[12]

A significant loss on the central bank's balance sheet could potentially affect its ability to achieve its monetary policy and financial stability objectives

The central bank also requires collateral to support neutrality in the conduct of liquidity operations and to avoid having to differentiate the pricing of its lending in day-to-day operations based on the creditworthiness of the counterparties. Secured lending allows the central bank to lend at a single interest rate to a set of heterogeneous counterparties, which facilitates the communication and transmission of monetary policy.[13]

What is a haircut?

The amount of collateral required for a loan is determined by the haircut applied by the lender. The haircut "trims" the market value of the collateral by a certain percentage to determine the maximum loan amount. For example, when the haircut is 1 per cent, the borrower needs to post collateral (e.g., treasury bills) with a value of $100 with the central bank in order to acquire a loan of $99.

---

8   For more information on the function of the Bank of Canada as "lender of last resort," see <http://www.bankofcanada.ca/financial-system/lender-of-last-resort/>.

9   This provides LVTS participants with a pricing incentive to first look for other sources of funding. For further details, see the backgrounder on the Bank Rate at <http://www.bankofcanada.ca/about/backgrounders/bank-rate/>. For more information on the implementation of monetary policy in the LVTS environment, see the Bank's primer at <http://www.bankofcanada.ca/wp-content/uploads/2010/07/lvts_primer_2010.pdf>.

10 The Bank of Canada Act (<http://www.bankofcanada.ca/wp-content/uploads/2010/07/act_loi_boc_bdc.pdf>) requires the Bank to take collateral when making loans or providing advances.

11 The central bank can finance the loss by permanently expanding the monetary base (i.e., a seigniorage tax), limiting the flexibility needed to achieve price and financial stability. Alternatively, by decreasing remittances of earnings (or through capital injections from the government), the loss can be transferred to the government. The government would have to finance the loss by raising taxes or by cutting spending, hence, tightening its budget and potentially lowering economic welfare.

12 Johnson and Zelmer (2007) examine, in the context of central bank balance-sheet management under new accounting standards, how a negative capital position on the Bank's balance sheet might pose risks to the credibility of monetary policy and raise doubts about its ability to act as lender of last resort.

13 There are strict "membership criteria" for becoming a direct counterparty to the Bank of Canada, either as a direct LVTS participant or as a primary dealer.

Why use haircuts?

In the event that a borrower defaults, a haircut minimizes the lender's losses from the market and liquidity risks associated with the collateral pledged. A lender's key concerns are what the secondary price of the collateral would be when the lender needs to sell the collateral to reclaim the funds loaned and how long it would take to liquidate it. A haircut is therefore determined based on the price volatility and liquidity of the pledged collateral asset. More-volatile or less-liquid assets usually receive a higher haircut to protect the lender from a price decline or liquidation costs. In the case of a default where the price of the collateral drops so low that the haircut provides insufficient coverage, the lender will suffer a loss. Thus, when determining the size of a haircut, there is a trade-off: a lower haircut can relax the borrower's liquidity constraint, but it also increases the lender's potential loss.

A haircut also influences the behaviour of potential counterparties. A borrower facing a haircut is less inclined to default because the collateral posted with the lender is worth more than the loan amount it has received. Hence, the borrower is more likely to repay the loan and retrieve the collateral.

Central Bank vs. Private Sector Haircuts

The traditional risk-management approach used to determine haircuts is based on examining the historical volatility of the value of the collateral and setting a haircut to limit the lender's risk exposure, based on a given risk tolerance.[14] This approach is appropriate for a private sector lender that has little market power and wants to maximize expected return and minimize exposure to counterparty default.

The central bank's unique ability to create and supply an unlimited amount of settlement assets to the payments system ensures that it is never constrained in its provision of liquidity.[15] Nevertheless, any losses incurred will be shared by society and will likely generate costly economic distortions. The central bank must therefore balance access to liquidity with prudent risk management.

The central bank also plays a prominent role in the overnight money market and other core funding markets. The study presented in the following section argues that the central bank's haircut policy can therefore directly affect liquidity in these markets and indirectly influence market participants' choice of asset portfolios, as well as the pricing of credit and liquidity spreads. The central bank is concerned not only about its own exposure to credit risk, but also about the efficiency and stability of the financial system. Consequently, in setting its haircut policy, the central bank must consider the impact of the policy on the financial system and its participants.[16]

> *The central bank is concerned not only about its own exposure to credit risk, but also about the efficiency and stability of the financial system*

A Model of Central Bank Haircuts

In a recent study, Chapman, Chiu and Molico (2011) develop a stylized model of the Canadian large-value payments system to examine the various effects that changes to central bank collateral haircuts can have on the economy.

The model economy

In the model, financial institutions (FIs) send and receive payments through a payments system on behalf of their clients. FIs are uncertain of the exact amount and timing of these payments and must therefore efficiently manage their liquidity based on a forecast of their payment needs. At the beginning of each day, FIs try to forecast their liquidity needs and to optimize their asset portfolios accordingly by trading with each other in a financial market. In the model, an FI can hold only two possible assets in its portfolio: a *liquid* asset, which has a lower rate of return but is (or can be easily exchanged for) the means of settling payment obligations; and an *illiquid* asset, which has a higher expected rate of return but is not generally accepted in the settlement of payment obligations, and has a more volatile price (is

---

14 See, for example, García and Gençay (2006) and the related risk-management literature cited.

15 Payments in the LVTS are settled via the account balances that participants hold at the Bank. Since the Bank of Canada can supply virtually unlimited amounts of these deposits and guarantees settlement even in the event of multiple defaults on a single day, sufficient balances will be available for settlement purposes under all circumstances.

16 As discussed in Zorn and García (2011), the collateral policy of central banks not only reflects the objective of protecting the central bank's balance sheet from financial losses arising from a counterparty default, but also supports the central bank's objectives for monetary policy and financial stability, including encouraging better risk-management practices by financial institutions.

riskier). This scenario highlights the trade-off between liquidity risk and asset return. When FIs foresee the need to send a high volume of payments, they will choose to hold a higher amount of the liquid asset than the illiquid asset to facilitate payment. If they expect to receive more payments than they send, they will choose to hold a lower amount of the liquid asset to earn a higher rate of return. Given the uncertain nature of the payment flows, however, and their realized payment needs, some banks may be holding too little of the liquid asset.

If FIs are unable to fund a shortage of the liquid asset through the intraday money market, then the central bank lending facility comes into play. In the study, the authors focus on this specific case. In this situation, there is credit risk because central bank loans are subject to potential default by the borrowers. To limit its financial exposure, the central bank requires borrowers to pledge their asset holdings as collateral. In the model, since the only FIs that require central bank funding are those that need to send payments but hold illiquid assets, these are the assets pledged as collateral. Since the market price of these illiquid assets can fluctuate, the central bank applies a haircut to their value when determining the amount of the loan. This protects its balance sheet in case there is a default by the borrower that is accompanied by a decrease in the price of the pledged asset. If there is no default, the borrower repays the loan and retrieves its collateral. If there is a default, the collateral is surrendered to the central bank, which will sell it to recover its losses.

The authors analyze the effects of central bank haircuts on several key aspects of the financial system: the amount of liquidity in the system, equilibrium asset prices, the exposure of the central bank's balance sheet, the asset-portfolio decisions of financial institutions and their incentives to default.

The following three channels interact and jointly determine the equilibrium effects of the haircut policy on the economy:

(i) Payments-system liquidity
Changes in the central bank's haircut policy affect the borrowing constraint faced by payments-system participants that need funding. They directly alter the borrowing and payment decisions of the affected FIs and, ultimately, improve or restrict underlying consumption and production in the economy. By lowering the haircut, the central bank reduces the discount on collateral and relaxes the liquidity constraint of FIs that are holding the illiquid asset but

need liquidity. This facilitates payment flows and supports economic activity.

(ii) Asset allocation
Changes to the haircut on a specific asset will affect its relative liquidity with respect to other assets. In the model, a lower haircut on the illiquid asset allows its holder relatively easier access to liquidity if needed.[17] This, in turn, affects the portfolio decisions of FIs when they forecast their liquidity needs and, ultimately, affects the prices and allocation of financial assets. A lower haircut will increase the attractiveness of the illiquid asset, which alters the relative prices of the liquid and illiquid assets and induces agents to accumulate more of the illiquid asset. This change in portfolio can be welfare reducing, if the initial holding of liquid and illiquid assets is optimal, or it can be welfare enhancing, if the initial portfolio is not socially optimal and agents are holding too much of the liquid asset.

(iii) Central bank exposure
For a given probability of default by the borrowers, lowering the haircut on a risky asset could increase the exposure of the central bank's balance sheet to possible losses. In addition, a lower haircut could potentially increase the probability of default because it makes the decision to default less costly relative to repaying the loan. Haircuts set too low could therefore give FIs the incentive to default. This increased probability of default affects the central bank's ability to control price stability because of a potential unintended expansion of its balance sheet. In the model, this translates into an expansion of the monetary base that can increase expected inflation, which is costly to society.[18]

Discussion and Conclusion

While the model presented here is stylized, it highlights the essential points that the central bank must consider when determining its haircut policy. First, it stresses the traditional trade-off between the provision of liquidity and the credit, market and liquidity risks. When providing liquidity via collateralized loans, the central bank must consider the implications of restricting the eligibility of assets, as well as the impact of its haircut policy on the exposure of its

---

17 Since the model deals with only two assets, this change in relative liquidity is between these two assets. In the more realistic case, a change in the haircut on one asset will change its liquidity in relation to that of all other assets.

18 Since the government's fiscal authority is not explicitly modelled, the loss is financed by an inflation tax.

balance sheet to credit risk. Second, the model highlights the potential of the central bank's haircut policy to affect the allocations of asset portfolios, asset prices and borrowers' incentives to default, which, as noted above, can have real welfare consequences.[19]

The model highlights the potential of the central bank's haircut policy to affect the allocations of asset portfolios, asset prices and borrowers' incentives to default

In its day-to-day operations, under normal circumstances, the central bank typically restricts the list of eligible assets accepted in payments systems to high-quality assets with relatively low credit and liquidity risk; it aims for "market-neutrality," whereby its operations do not significantly influence relative asset prices.[20] Under extraordinary circumstances, such as a liquidity-induced crisis, however, it may be reasonable for the central bank to temporarily broaden the collateral pool and accept less-marketable assets in its day-to-day operations, thus freeing more conventional liquid collateral for use in other operations.[21] When the scarcity of liquidity threatens the well-functioning of the financial system (for example, during a period of financial stress), the benefits of providing additional central bank liquidity by reducing haircuts or expanding the list of eligible assets can outweigh the potential costs associated with the distortion of asset portfolios and the higher risk exposure of the central bank's balance sheet. A similar view—that central banks should take the liquidity of the financial system into account when setting haircuts—is found in the report of the

Bank for International Settlements' Committee on the Global Financial System (CGFS 2010) on the cyclical behaviour of private sector margins and haircuts.

Policy-makers face the challenge of a growing demand for high-quality collateral

Policy-makers also face the challenge of a growing demand for high-quality collateral. Modern financial systems tend to utilize more collateral because of the increased private use of collateral, and because of the need to post additional collateral with payment and settlement systems.[22] The G-20 countries committed to have all standardized over-the-counter derivatives contracts cleared by central counter-parties (CCPs) by the end of 2012 to help strengthen the global financial system. Such an increase in CCP activity has the potential to increase the need for collateral.[23] In addition, revisions to the core principles for financial market infrastructure, currently being considered by the Bank for International Settlements' Committee on Payment and Settlement Systems and by the International Organization of Securities Commissions, will further increase the demand for collateral by financial market participants. The haircuts set by central banks are important parameters in determining the ability of financial systems to make the most efficient use of high-quality collateral.

---

19 For a discussion on the collateral-management practices of LVTS participants during relatively normal times, see D'Souza (2009).

20 Chailloux, Gray and McCaughrin (2008) provide a comparison of collateral frameworks in different jurisdictions. They discuss the eligibility of assets as collateral and the implications of collateral policy for the market's choice of asset portfolios and for the pricing of credit and liquidity spreads.

21 During the recent financial crisis, the Bank of Canada expanded the list of assets that could be pledged as collateral in LVTS intraday operations and under the SLF, while strengthening its requirements for transparency and credit ratings. Direct participants in the LVTS were able to pledge their non-mortgage loan portfolio (NMLP) to cover the system's collateral requirements, as well as those of the SLF and the Term Loan Facility. As of 1 April 2010, LVTS participants may use their NMLP for up to 20 per cent of their total collateral value. The Bank also conducted term purchase and resale agreements with primary dealers against an expanded set of eligible securities.

22 An example of the increased private use of collateral is the larger size of institutional cash pools, which demand collateral in their day-to-day money market investments (Pozsar 2011).

23 If the settlement of over-the-counter financial transactions moves to CCPs, then the aggregate value of collateral needed may decrease because of efficiency gains in multilateral netting. But if this move decreases the overall amount of netting, then the need for collateral could increase (Duffie and Zhu 2010).

Literature Cited

Arjani, N. and D. McVanel. 2006. "A Primer on Canada's Large Value Transfer System." Bank of Canada. Available at <http://www.bankofcanada.ca/wp-content/uploads/2010/05/lvts_neville.pdf>.

Chailloux, A., S. Gray and R. McCaughrin. 2008. "Central Bank Collateral Frameworks: Principles and Policies." IMF Working Paper No. WP/08/222.

Chapman, J., J. Chiu and M. Molico. 2011. "Central Bank Haircut Policy." *Annals of Finance* 7 (3): 319–48.

Committee on the Global Financial System (CGFS). 2010. "The Role of Margin Requirements and Haircuts in Procyclicality." CGFS Papers No. 36.

D'Souza, C. 2009. "Collateral Management in the LVTS by Canadian Financial Institutions." *Bank of Canada Review* (Summer): 3–14.

Duffie, D. and H. Zhu. 2010. "Does a Central Clearing Counterparty Reduce Counterparty Risk?" Rock Center for Corporate Governance at Stanford University Working Paper No. 46.

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.

Johnson, G. and M. Zelmer. 2007. "Implications of New Accounting Standards for the Bank of Canada's Balance Sheet." Bank of Canada Discussion Paper No. 2007-2.

Pozsar, Z. 2011. "Institutional Cash Pools and the Triffin Dilemma of the U.S. Banking System." IMF Working Paper No. WP/11/190.

Zorn, L. and A. García. 2011. "Central Bank Collateral Policy: Insights from Recent Experience." *Bank of Canada Review* (Spring): 37–45.

Extracting Information from the *Business Outlook Survey*: A Principal-Component Approach

Lise Pichette and Lori Rennison, Canadian Economic Analysis

- *Since 1997, the* Business Outlook Survey *(BOS) has provided the Bank of Canada with valuable and timely information for the conduct of monetary policy.*

- *Recent work using principal-component analysis to extract information common to the BOS indicators is reviewed, as is testing of the out-of-sample forecasting performance of various models using this information.*

- *Results suggest that summarizing the common movements among BOS indicators may provide useful information for forecasting near-term growth in business investment. For growth in real gross domestic product, however, the survey's balance of opinion on future sales growth appears to be more informative.*

Since the autumn of 1997, the Bank of Canada's regional offices have conducted quarterly consultations with businesses across Canada. These consultations, referred to as the *Business Outlook Survey* (BOS), are structured around a questionnaire that covers topics of importance to the Bank, notably business activity, pressures on production capacity, prices and inflation, and credit conditions.[1] The responses to these qualitative questions (e.g., whether sales volumes will increase at a greater, lesser or the same rate over the next 12 months as over the past 12 months), together with the explanations that accompany them, allow senior economics staff at the Bank's regional offices to provide a macro-level assessment of the economy using the various demand- and supply-side signals from the survey. This assessment supplements the more quantitative approaches used by the Bank to evaluate the economic situation and outlook by providing insights into what businesses are seeing and planning.[2]

A key advantage of the BOS is its timeliness. Consultations take place around the middle of each quarter, and the results are published the week before the Bank's next fixed date for announcing monetary policy decisions. This is well ahead of the release of the National Income and Expenditure Accounts for that quarter. The high demand for timely information about the economy has led the survey results to become a well-monitored

---

1 For a detailed description of the survey, see Martin (2004) and *Backgrounder on Questions in the* Business Outlook Survey *Concerning Past Sales and Credit Conditions* (<http://www.bankofcanada.ca/wp-content/uploads/2011/07/bos_backgrounder_jan2008.pdf>).

2 See Macklem (2002) and Jenkins and Longworth (2002) for a description of how the BOS fits into the Bank's monetary policy decision-making process.

Table 1: *Business Outlook Survey* indicators

Survey question	Horizon	Type of signal
Balance of opinion[a] on past sales growth	Past 12 months	Demand-side
Balance of opinion on future sales growth	Next 12 months	Demand-side
Balance of opinion on investment in machinery and equipment	Next 12 months	Demand-side
Balance of opinion on employment	Next 12 months	Supply-side, indirect demand-side
Ability to meet an unexpected increase in demand[b]	Current	Supply-side, cost structure
Percentage of firms facing labour shortages	Current	Supply-side, cost structure
Balance of opinion on input prices	Next 12 months	Supply-side, cost structure
Balance of opinion on output prices	Next 12 months	Supply-side, margins
Balance of opinion on credit conditions	Past 3 months	Financial markets, demand-side

a. Percentage of firms responding "greater" or "higher" minus percentage of firms reporting "lesser" or "lower"
b. Percentage of firms responding "some" or "significant" difficulty

information source for the press and the financial community since their public release in 2004. Information published in the BOS helps to refine the Bank's view on the economic outlook, and is often cited in the Bank of Canada's *Monetary Policy Report*.

Our understanding of the statistical precision of individual BOS indicators has been strengthened by the research of de Munnik, Dupuis and Illing (2009) and de Munnik (2010) (**Box 1**). An evaluation of the survey's ability to predict economic variables, however, has been limited by its relatively short sample period. The initial assessment of the information content of the BOS by Martin (2004) was based on graphical and correlation analysis of 24 observations. Since then, the sample period has become long enough to include at least one full economic cycle with periods of expansion and slowdown, as well as a steep recession and a recovery, providing richer information for empirical analysis.

This article summarizes recent work that contributes to our understanding of the survey's information content by extending the early work by Martin (2004) in two key ways. First, since all BOS questions are designed to capture some aspect of economic activity and, therefore, are interrelated, principal-component analysis was used to extract the common underlying variations among the indicators. Second, the information content of these common movements was assessed, using regression analysis and a forecasting assessment. The first test of this measure's usefulness was whether it can help predict growth in real gross domestic product (GDP) and, if yes, whether it outperforms the survey question on future sales expectations—the question most

closely tied to measuring GDP. The second test was whether the common movements of indicators provide clearer signals for any one component of economic activity. Given that the BOS is a survey of firms, business investment was a natural element to consider.

Extracting Common Information from BOS Indicators

For analytical and communication purposes, the responses to most BOS questions are expressed in terms of a balance of opinion or as a proportion of respondents (for questions on labour shortages and firms' ability to meet demand). This practice has been useful for interpreting the survey results with respect to specific aspects of economic activity. Each quarter, staff in the Bank's regional offices assess and amalgamate these signals from the survey regarding aggregate demand, aggregate supply and financial markets (**Table 1**), informed by the broader discussions that take place with firms during the interviews.

To statistically evaluate the survey's information content and its ability to predict real economic variables, common movements from the various BOS indicators were extracted using a data-reduction technique—principal-component analysis (PCA). In addition to capturing a common source of variation, using this shared underlying component to represent the fuller BOS data set in a forecasting assessment is an appealing alternative to using individual indicators, since it conserves degrees of freedom and lessens concerns about issues of multicollinearity.

Box 1: Statistical Accuracy of the *Business Outlook Survey*

Martin (2004) reports that the statistical properties of the *Business Outlook Survey* (BOS) are difficult to determine, given the small sample size of 100 firms and the non-random, quota-sampling approach, which involves setting objectives for the number of firms selected by region, industry and size in order to be representative of the Canadian economy. De Munnik, Dupuis and Illing (2009) and de Munnik (2010) have made significant progress in dealing with this issue.

Using a Monte Carlo simulation framework, de Munnik, Dupuis and Illing (2009) construct an artificial data set of firms and their responses and estimate the impact of the Bank's non-random sampling on the accuracy and coverage of the survey.[1] More specifically, they develop a method for modelling a complex, non-random sampling process and for computing relevant measures of the confidence intervals. This allows them to replicate the survey's firm-selection process. **Table 1-A** shows how each quota or constraint affects the estimate of the population parameter compared with the simple case of random sampling. When investigated individually, only the quotas for industry and firm size are found to widen the confidence intervals (rows 3 and 4). Results for the fully constrained model (row 8), however, show no evidence that the Bank's firm-selection process produces significantly biased estimates and/or wider confidence intervals than random selection. In other words, although the quota constraints result in biases on the parameter estimates when controlled individually, these biases are small and appear to be largely offset when the model is calibrated using average historical responses. With respect to survey coverage, the authors find that the BOS method of firm selection restricts the survey sample but does not create bias in the estimate.

De Munnik (2010) extends this analysis by outlining the statistical properties of questions expressed as population proportion versus balance of opinion, and demonstrates how the design of the question affects the calculation of the confidence intervals. He also shows that the confidence bands around both types of question can change from survey to survey when the underlying distribution of responses becomes more or less concentrated in particular categories (such as "higher," "the same" or "lower"). In particular, he illustrates that the confidence intervals around balance-of-opinion questions are larger when there is more dispersion in the responses.

These studies have improved understanding of the survey's statistical accuracy and, together with the qualitative stories that accompany the responses, allow staff at the Bank's regional offices to better interpret and describe survey results from quarter to quarter.

Table 1-A: Comparison of simulation results

Selection model	Bias vs. "pseudo-population"	Confidence interval 95% (66%)
1. Random sample	0.06	16.6 (8.2)
2. Regional quota	2.00	16.6 (8.2)
3. Industry quota	-2.07	17.5 (8.6)
4. Firm-size quota	-2.78	17.7 (8.7)
5. Rotation constraint	0.17	16.7 (8.2)
6. Familiarity constraint	-0.23	17.0 (8.4)
7. Non-response constraint	-0.10	16.7 (8.4)
8. Fully constrained model	-0.23	16.8 (8.3)

---

[1] De Munnik, Dupuis and Illing (2009) do not analyze the coverage of the survey, but a revised version of the paper that includes this analysis is available upon request.

Box 2: Technical Details of Principal-Component Analysis

The first principal component is obtained by maximizing its contribution to the variance of a set of p variables (x). It is expressed as follows:

$$\alpha_1' \, x = \alpha_{11} \, x_1 + \alpha_{12} \, x_2 + \cdots + \alpha_{1p} \, x_p =$$

$$\sum_{j=1}^{p} \alpha_{1p} \, x_p,$$

where x has a covariance matrix Σ. To derive this first principal component, the vector α_1 maximizes $var(\alpha_1' x) = \alpha_1' \Sigma \alpha_1$ subject to $\alpha_1' \alpha_1 = 1$. Using the technique of Lagrange multipliers:

$$\max_{\alpha_1} \alpha_1' \Sigma \alpha_1 - \lambda(\alpha_1' \alpha_1 - 1).$$

The first-order condition is

$$(\Sigma - \lambda I_p)\alpha_1 = 0,$$

where I_p is the $(p \times p)$ identity matrix, λ is an eigenvalue of Σ, and α_1 is the corresponding eigenvector. To maximize $var(\alpha_1' x)$, the vector of coefficients in the first principal component, α_1, is the eigenvector associated to the largest eigenvalue:

$$var(\alpha_1' x) = \alpha_1' \Sigma \alpha_1 = \alpha_1' \lambda \alpha_1 = \lambda \alpha_1' \alpha_1 = \lambda.$$

The kth principal component is derived by maximizing $var(\alpha_k' x)$ subject to $\alpha_k' \alpha_k = 1$ and $\text{cov}(\alpha_k' x, \alpha_l' x) = 0$ for all $k \neq l$.

Chamberlin (2007) makes similar use of PCA on survey measures of economic activity in the United Kingdom to develop an alternative forecast of GDP.

As explained by Jolliffe (2002, 1), "The central idea of principal component analysis is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set." This method generates a new set of variables—principal components—that are linear combinations of the original variables. Principal components are artificial variables that account for most of the variance in the observed variables contained in the data set, and they are all orthogonal to each other. (See **Box 2** for a brief technical description of PCA.)

By definition, the number of principal components that can be found is the same as the number of variables considered, but, in general, most of the variance in the data set will be accounted for by fewer principal components. For this analysis, only the first principal component (PC1) was retained. This underlying variable was extracted from the nine published BOS indicators that pertain to firms' views on their own business situation or plans (**Table 1**).[3]

While data for most indicators are available from 1998Q3, the questions on credit conditions and on the ability of firms to meet an unexpected increase in demand were added more recently. Extracting information from all nine indicators therefore required the length of the sample be limited to 2001Q4 to 2011Q2.[4, 5]

Correlation with Economic Data

Results from the correlation analysis of PC1 and real economic variables are shown in **Table 2**. For each economic variable, PC1 is compared with the survey indicator that most closely corresponds to the same economic concept.

Panel A presents the peak correlations with real GDP growth. The correlation results are moderate to moderately strong for both PC1 and the most relevant

---

3 This analysis excluded the question relating to firms' expectations regarding consumer price index inflation over the next two years.

4 The question on the ability to meet demand was added in 1999Q3 and that on credit conditions in 2000Q3. The question on credit was modified in 2001Q4 to reduce its horizon from the past 12 months to the past 3 months.

5 While the information content analyzed used the sample spanning 2001Q4 to 2011Q2, the sample was also extended back to 2000Q3, using the first formulation of the credit-conditions question. Since results were very similar, this extended sample is used for the charts to illustrate the behaviour of the underlying variable through the 2001 slowdown.

Table 2: Peak correlations
Sample: 2001Q4 to 2011Q2

BOS indicators	Quarter-over-quarter			Year-over-year		
	Panel A: Real GDP growth					
PC1	0.54 (t - 1)	0.56 (t)	0.50 (t + 1)	0.63 (t)	0.73 (t + 1)	0.63 (t + 2)
Balance of opinion on future sales	0.49 (t)	0.69 (t + 1)	0.54 (t + 2)	0.61 (t + 2)	0.68 (t + 3)	0.55 (t + 4)
	Panel B: Real consumption growth					
PC1	0.51 (t - 2)	0.45 (t - 1)	0.31 (t)	0.63 (t)	0.64 (t + 1)	0.47 (t + 2)
Balance of opinion on future sales	0.11 (t - 1)	0.32 (t)	0.27 (t + 1)	0.23 (t + 1)	0.33 (t + 2)	0.31 (t + 3)
	Panel C: Real business-investment growth					
PC1	0.61 (t - 1)	0.83 (t)	0.73 (t + 1)	0.79 (t + 1)	0.79 (t + 2)	0.63 (t + 3)
Balance of opinion on investment in machinery and equipment	0.59 (t)	0.67 (t + 1)	0.57 (t + 2)	0.55 (t)	0.70 (t + 1)	0.65 (t + 3)

indicator—the balance of opinion on future sales growth.[6] While the question on future sales asks firms to characterize the expected change in sales growth over the next 12 months (i.e., momentum as opposed to growth), it is reasonable to expect that firms' predictions about future momentum could also contain information about current or near-term growth. The strongest correlation between PC1 and quarterly GDP growth is reached contemporaneously (in quarter t), while, for the balance of opinion on future sales growth, the highest correlation coefficient is obtained one quarter ahead (t + 1). Given the 12-month horizon of many BOS questions, it is also worth noting that the correlation for the question on sales outlook with the year-over-year growth of real GDP reaches a peak three quarters ahead, while it is only one quarter ahead for PC1. **Chart 1** plots the balance of opinion on future sales and PC1, together with real GDP growth. It shows that both the individual question and the common factor seem to track developments in aggregate economic activity relatively well.

Panels B and C of **Table 2** report the correlation results for specific components of the GDP: consumption and business investment. The results for quarterly growth in real consumption are weak or moderate, while those for growth in real business investment are moderately strong to strong. The weak correlation with consumption may be partly explained by the survey sample, which comprises

the business sector rather than the consumer sector, and even within the business sector, not only firms selling to consumers, but also firms selling to other businesses or exporting. The underlying variable, PC1, has a higher correlation with business investment than the survey question on the expected direction of change in investment in machinery and equipment over the next 12 months. This suggests that extracting the common movements from all the survey questions might lead to a better indicator for quarterly growth in business investment than this single question on investment intentions.

The strong correlation between PC1 and business investment is interesting, since very few indicators of investment are available

Chart 2 shows that PC1 closely tracks fluctuations in business investment over the sample period. The strong correlation between this underlying variable derived from the BOS results and business investment is interesting, since very few indicators of investment are available ahead of official statistical data. Moreover, the correlation coefficient continues to be moderately strong one quarter ahead, suggesting that the BOS also contains forward-looking information regarding business investment.
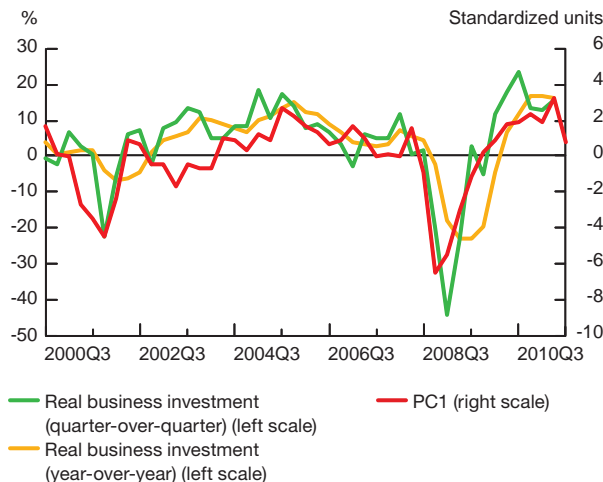
6   The scale used to evaluate the correlation coefficients is that used in Martin (2004): strong, > 0.80; moderately strong, 0.60 to 0.80; moderate, 0.40 to 0.60; weak, 0.20 to 0.40; insignificant, < 0.20.

Chart 1: Aggregate economic activity and BOS indicators
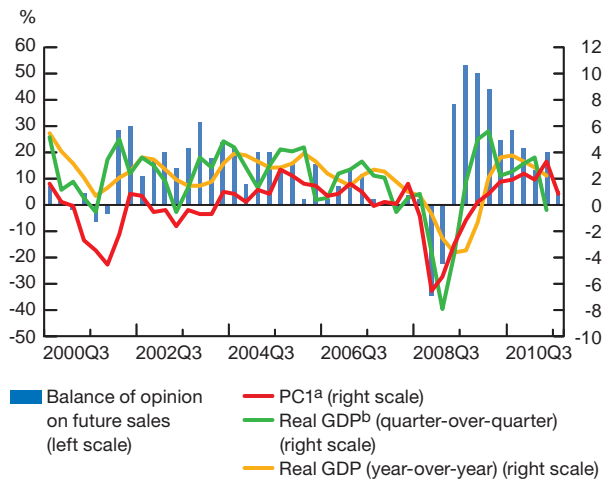
a. PC1 is measured in standardized units.
b. Real GDP is measured as a percentage.
Sources: Statistics Canada, Bank of Canada and authors' calculations
Last observation: 2011Q3

Legend:
- Balance of opinion on future sales (left scale)
- PC1[a] (right scale)
- Real GDP[b] (quarter-over-quarter) (right scale)
- Real GDP (year-over-year) (right scale)

Chart 2: Business investment and the underlying BOS variable

Sources: Statistics Canada and authors' calculations
Last observation: 2011Q3

Legend:
- Real business investment (quarter-over-quarter) (left scale)
- Real business investment (year-over-year) (left scale)
- PC1 (right scale)

Forecasting Assessment

Regression analysis and a forecasting exercise were carried out to evaluate (i) whether the underlying variable extracted from the BOS indicators using PCA can provide information beyond that contained in the past values of the economic variables, and (ii) whether it provides more information than is contained in the answers to the individual survey questions on future sales growth and investment intentions.

Various simple models were examined and compared based on the root mean square errors (RMSE) computed using a series of one-step-ahead forecasts for each equation.[7] Specifically, each equation is estimated for a sample spanning the period 2001Q4 to 2006Q1, and a forecast is produced for 2006Q2.[8] One observation is then added to the estimation period for the next-quarter forecast, and this is repeated up to 2011Q2. The ratio of the RMSE for each equation, relative to a benchmark case that includes only the lags of the dependent variable, is reported. For example, an RMSE ratio below one implies that the inclusion of the common component obtained from BOS results improves the forecast derived from an equation that takes into account only the latest information on the variable of interest.

Table 3 summarizes the estimation results for quarterly real GDP growth. The first five rows report results for three different specifications (equations 1 to 3) estimated on the full sample (2001Q4 to 2011Q2). The coefficient on PC1 (equation 2) is significant, and the adjusted R^2 increases compared with equation 1, which includes only lagged GDP. PC1 is incorporated only at time t in this equation, because lags were not statistically significant. Nonetheless, since data from the BOS are available almost two months before the release of the national accounts, the results can be useful for forecasting. The results for equation 3, however, indicate that the balance of opinion on future sales growth remains a better indicator than the underlying variable, with the adjusted R^2 increasing to 0.56. The balance of opinion on future sales is significant only contemporaneously despite correlation results that suggested that expectations of future sales contained more forward-looking information.

The last row of **Table 3** reports the results of the out-of-sample forecast exercise. The RMSE ratios for both equations 2 and 3 are below one, indicating that the inclusion of information from the BOS improves the forecast from that of equation 1. However, the improvement is only marginal for PC1, and the difference between equations 2 and 3 is found to be significant, according to the Diebold-Mariano test. Thus, the underlying variable extracted from the set of BOS responses does not outperform the balance of opinion on future sales in forecasting real economic activity.

---

7   The prediction is for the current quarter before the release of the national accounts.
8   The data used in this exercise were the latest available, published on 31 August 2011.

Table 3: Estimation results for real GDP growth (quarter-over-quarter)
Sample: 2001Q4 to 2011Q2

Variables included	Equation 1	Equation 2	Equation 3
Constant	0.79 (1.84)[a]	1.11 (2.56)	-0.33 (-0.75)
GDP growth (t - 1)	0.60 (4.63)	0.42 (2.84)	0.58 (5.45)
PC1 (t)		0.41 (2.19)	
Balance of opinion on future sales (t)			0.07 (4.34)
Adjusted R^2	0.35	0.41	0.56
RMSE ratio	1.00	0.97	0.82

a. t-statistics are in parentheses.

Table 4: Estimation results for growth in real business investment (quarter-over-quarter)
Sample: 2001Q4 to 2011Q2

Variables included	Equation 1	Equation 2	Equation 3	Equation 4
Constant	1.76 (1.05)[a]	2.97 (2.51)	4.20 (3.71)	-2.36 (-1.22)
Growth in investment (t - 1)	0.66 (5.22)	0.26 (2.41)		0.40 (2.93)
PC1 (t)		4.16 (6.31)	3.72 (4.57)	
PC1 (t - 1)			1.82 (2.25)	
Balance of opinion on investment in machinery and equipment (t)				0.39 (3.32)
Adjusted R^2	0.41	0.71	0.71	0.54
RMSE ratio	1.00	0.70	0.71	0.89

a. t-statistics are in parentheses.

Table 4 reports the estimation results for growth in real business investment, using the same approach. In this case, the values of the adjusted R^2 and the RMSE ratio are quite impressive when the underlying variable extracted from the BOS results is included as an explanatory variable. Based on the estimates from equation 3, PC1 alone, without the lagged growth of business investment, produces results very similar to those of equation 2. While the survey question on investment intentions for machinery and equipment is found to have explanatory power

(equation 4), results for equations 2 and 3 indicate that PC1 outperforms the survey question.[9] From these results, it appears that a measure of the underlying information from all BOS indicators provides more useful signals for monitoring the growth of near-term investment than the question regarding investment intentions for machinery and equipment over the next 12 months in isolation.

Robustness

Results for PCA, presented in the previous section, are relatively robust to weights. Indeed, alternative approaches to extracting common movements, including a simple average of the nine indicators and factor analysis, generated series that were highly correlated with PC1.[10] Moreover, real-time estimates of the first principal component were examined, since weights may fluctuate when the sample is changing. Sensitivity analysis suggests that, although the weights vary, the underlying variable extracted using PCA remains virtually the same.[11]

Discussion and Conclusions

The research reviewed here used principal-component analysis to evaluate the information content of the BOS. It also assessed the information content of the first principal component relative to that of individual survey questions in an out-of-sample forecasting exercise. This is the first empirical assessment of the BOS information content since the initial correlation analysis by Martin (2004), and it has provided several notable contributions.

First, the results suggest that the first principal component appears to be a useful indicator of economic activity, particularly for providing information on investment spending—a variable that is typically difficult to predict and for which there are very few indicators. This may not be surprising, since the BOS is a survey of firms, and all its questions provide some signals relating to the probability of investing. For instance, if the outlook of firms regarding sales, employment and investment improves, if more firms are operating at or above their production capacity, and if more firms report an easing in credit conditions, then it is reasonable to expect higher

---

9 The results from both equations 2 and 3 are statistically different from those of equation 4 at the 5 per cent significance level, according to the Diebold-Mariano test.
10 Correlation coefficients were greater than 0.98. (This is true for the series expressed both in levels or first differences.)
11 Results of this sensitivity analysis are available from the authors upon request.

investment activity. The outlook for prices can also play a role in firms' near-term investment spending; for example, if firms are expecting higher input costs and are therefore spending to become more efficient, or if higher prices are stimulating activity in particular sectors where investment projects become profitable (as was the case with the boom in commodity prices in the 2000s).

Second, this analysis found that the individual survey questions on future sales growth and intentions for investment in machinery and equipment provide useful information in a forecasting context for real GDP growth and the growth of business investment, respectively. The inclusion of these variables in their respective regressions improved upon a simple auto-regressive model. In the case of business investment, however, the single question was found to be less informative than the measure of common movements.

This work contributes to our understanding of the survey's information content, but the reliability and robustness of the results will need to be evaluated over time as the sample period grows. Moreover, promising statistical assessments do not preclude careful examination of the movements in each BOS indicator every quarter, or the qualitative assessment of the messages that accompany firms' responses, both of which make a valuable contribution to monetary policy. Whether in terms of common movements or individual indicators, information gathered from business surveys is often best used with informed judgment rather than according to mechanical rules. The information obtained from individual survey indicators and from the qualitative assessment carried out by the Bank's regional offices remain important elements in BOS analysis. As emphasized in Martin (2004, 10), "The BOS interview format allows for a broader understanding of current business perceptions through confidential discussions with business representatives, which provide invaluable information that cannot be measured quantitatively."

Literature Cited

Chamberlin, G. 2007. "Forecasting GDP Using External Data Sources." *Economic & Labour Market Review* 1 (8): 18–23.

de Munnik, D. 2010. "Statistical Confidence Intervals for the Bank of Canada's Business Outlook Survey." Bank of Canada Discussion Paper No. 2010-7.

de Munnik, D., D. Dupuis and M. Illing. 2009. "Computing the Accuracy of Complex Non-Random Sampling Methods: The Case of the Bank of Canada's Business Outlook Survey." Bank of Canada Working Paper No. 2009-10.

Jenkins, P. and D. Longworth. 2002. "Monetary Policy and Uncertainty." *Bank of Canada Review* (Summer): 3–10.

Jolliffe, I. 2002. *Principal Component Analysis*. 2nd ed. New York: Springer.

Macklem, T. 2002. "Information and Analysis for Monetary Policy: Coming to a Decision." *Bank of Canada Review* (Summer): 11–18.

Martin, M. 2004. "The Bank of Canada's Business Outlook Survey." *Bank of Canada Review* (Spring): 3–18.

Modelling the Counterfeiting of Bank Notes: A Literature Review

Ben Fung and Enchuan Shao, Currency Department

- *Ensuring that the threat of counterfeiting remains low is critical to maintaining the public's confidence in bank notes as a means of payment.*

- *In the past 50 years, Canada has experienced two major episodes of counterfeiting at levels that threatened public confidence. The Bank of Canada has since developed a comprehensive anti-counterfeiting strategy that has brought counterfeiting under control and that supports public confidence by staying ahead of counterfeiters.*

- *Research that models the behaviour of relevant parties—the counterfeiters, the merchants and the central bank—helps to improve our understanding of their respective decisions: whether or not to produce counterfeit notes, whether or not to verify and accept bank notes offered, and what level of security to apply to bank notes. This research also sheds light on the importance of policies against counterfeiting.*

In June 2011, the Bank of Canada unveiled its new $100 and $50 notes to the public. The Bank's new series of polymer bank notes, which incorporates innovative security features that are not only easy to verify but also difficult to counterfeit, marks a significant advance in counterfeiting deterrence.

Issuing new, more-secure bank notes is one of four components in the Bank's comprehensive anti-counterfeiting strategy, which was developed during the most recent episode of high levels of counterfeiting.[1] At the episode's peak in 2004, the number of counterfeit notes detected per million notes in circulation reached 470, which was the highest among industrialized countries. The high levels of counterfeiting threatened to undermine Canadians' confidence in using bank notes. The anti-counterfeiting strategy developed by the Bank and its partners has been very effective. By 2010, detected counterfeits had dropped to 35 parts per million. Recent survey results suggest that bank notes continue to be an important means of payment in Canada, and thus it is important for the Bank to continue to ensure that Canadians can use bank notes with the highest confidence.[2]

This article aims to contribute to our understanding of counterfeiting and its policy implications by reviewing the literature on the subject. We begin by discussing the general characteristics of counterfeiting. We then briefly review several models of counterfeiting and study their implications for the incentive to counterfeit, social welfare and anti-counterfeiting policies. This research is still at an early stage, and more work is necessary before these models can be used to explain many of the observed characteristics of counterfeiting.

---

1   For a description of the Bank's four-part anti-counterfeiting strategy, see Bank of Canada (2011, 16).
2   According to a 2009 survey of Canadian consumers, nearly half of all transactions are conducted with bank notes, accounting for 20 per cent of the total value of transactions (Arango, Hogg and Lee, forthcoming).

Stylized Facts About Counterfeiting

Analysis of the available data on counterfeiting in Canada and other countries can provide a general understanding of the problem, as well as facilitate our discussion of recent research. Counterfeiting is usually measured in terms of the number of counterfeit notes detected in circulation in one year, typically in comparison to the total number of genuine notes in circulation, or parts per million (PPM).[3]
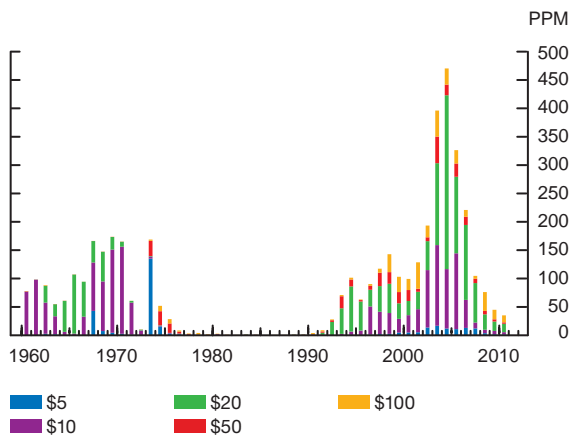
Counterfeiting tends to be episodic

Since 1960, Canada has experienced two major episodes of increased counterfeiting (**Chart 1**).[4] During the first episode, from 1960 to around 1975, counterfeiters targeted mainly $5, $10 and $20 notes. Toward the end of this episode, however, the counterfeiting of $50 and $100 notes began to rise, even as counterfeiting in general was trending down. From 1976 to 1991, counterfeiting remained subdued.

The second major episode of counterfeiting activity began in 1992, with counterfeiters initially targeting $20 notes. Around 1996, the number of bogus

Chart 1: Counterfeit Canadian bank notes passed for every one million genuine notes in circulation, 1960–2010

Sources: Bank of Canada and the Royal Canadian Mounted Police

$10 notes began to rise rapidly, and this was soon followed by an increased number of counterfeit $100 notes. The counterfeiting of $10 notes increased substantially following the issue of a new series of $10 notes in 2001 whose security features were not sufficient to deter counterfeiters. Also in 2001, a new and highly deceptive $100 counterfeit note began to be passed in considerable numbers, and the resulting publicity heightened the public's concern about counterfeiting.[5] Since the peak of the counterfeiting problem in 2004, the Bank's anti-counterfeiting strategy has been successful. In particular, the security features of the new series of bank notes issued between 2004 and 2006 were considerably improved, and the less-secure notes previously issued were aggressively withdrawn from circulation. By 2010, the number of counterfeit notes detected annually had dropped well below the Bank's current medium-term target of 50 PPM, suggesting that the most recent counterfeiting episode may have ended.

> *By 2010, the number of counterfeit notes detected annually had dropped well below the Bank's current medium-term target*

Given the episodic nature of counterfeiting, it would be important to understand what factors lead to a sharp rise in counterfeiting, and what anti-counterfeiting policies are effective in preventing and reducing it.

Counterfeiting tends to vary across countries

In recent years, counterfeiting has been a problem in Canada, the United Kingdom, Mexico and the euro area, while remaining at low levels in Switzerland, Australia and South Korea (**Chart 2**).[6] Indeed, in some countries, such as Australia, counterfeiting has been almost negligible for the past 20 years. The differences may lie with the security features of the current bank notes, the deterrence efforts of law-enforcement authorities, the extent of

---

3 A preferred measure would be the outstanding stock of counterfeit notes in circulation, which would better reflect the extent of the counterfeiting problem. Unfortunately, such a measure cannot be observed directly. Chant (2004) discusses various methods of estimating the stock of counterfeits in circulation.

4 Chart 1 also depicts the number of counterfeit notes detected for each denomination as a ratio of the total number of genuine notes in circulation. For a discussion of the counterfeiting of various series of Canadian bank notes, see Moxley, Meubus and Brown (2007).

5 The number of counterfeit $20 notes detected during the second episode was also relatively high. Since the early 1970s, the $20 note has become the most popular denomination in circulation, and demand rose sharply in the mid-1990s because it is the main denomination dispensed by automatic teller machines. It now accounts for over 40 per cent of all notes in circulation.

6 Many countries, including the United States, do not publish their counterfeiting data in a form that facilitates comparisons across countries.

Chart 2: Counterfeit bank notes passed for every one million genuine notes in circulation in selected countries
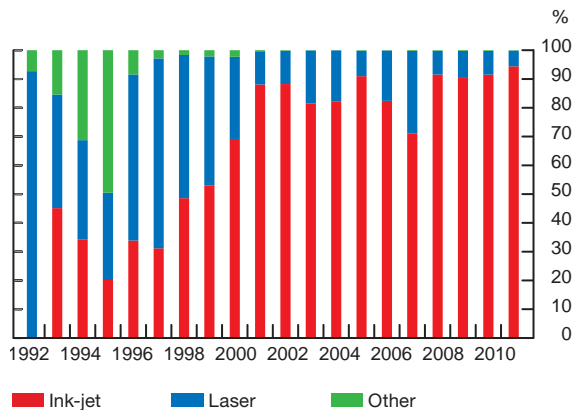

PPM

2004
2005
2006
2007
2008
2009
2010
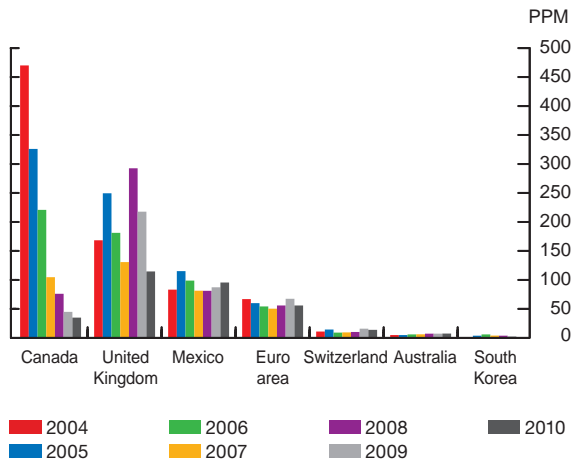
Source: Authors' calculations based on information available on websites of central banks and law-enforcement agencies

Chart 3: Share of counterfeits by printing method

%

Ink-jet   Laser   Other

Source: Royal Canadian Mounted Police

criminals' access to counterfeiting technology or the publicity that successful counterfeiters receive. Understanding why some countries have experienced a more serious counterfeiting problem than others and what factors explain this difference across countries could be instructive.

Counterfeiting tends to change with technology

Advances in security technology allow the central bank to produce more-secure notes, but other technological innovations can enable counterfeiters to produce higher-quality counterfeits at a relatively low cost. Prior to the 1980s, the production of counterfeits required engraved plates and offset presses that necessitated expensive, large-scale operations. Counterfeiting was therefore usually part of organized criminal activity, and large numbers of counterfeit notes were produced at a time. Police often seized large quantities of uncirculated counterfeits. The availability of colour copiers, and later of personal computers with scanners and ink-jet and laser printers, heightened the potential threat. Over 99 per cent of counterfeits detected in Canada are now produced by colour printers and copiers (**Chart 3**). Moreover, the share of uncirculated counterfeits seized by law-enforcement authorities relative to counterfeits passed into circulation has been low. It would therefore be interesting to study how technology affects the behaviour of counterfeiters and merchants, as well as the security design of bank notes.

Models of Counterfeiting

There has been almost no empirical work on counterfeiting because of the limited availability of counterfeiting data and related statistics. We therefore focus our discussion on theoretical studies that model the behaviour of the relevant economic agents. To date, there have been only a small number of these studies. They can be grouped into two categories—partial-equilibrium and general-equilibrium models—depending on how money is introduced.[7]

Partial-equilibrium models

Models in the first category do not explicitly specify the demand for money. It is assumed to be exogenous and does not depend on the actions of agents in the model. These models are therefore referred to as partial-equilibrium models. They are used to study the interactions of counterfeiters, merchants and the central bank, and often derive implications that can be compared with actual counterfeiting data. We review two studies that use this approach.

Lengwiler (1997) models the strategic interaction between a central bank and counterfeiters. He finds that it is optimal for a central bank to choose a design with a high level of security for high-denomination notes because the expected cost of producing a secure note is relatively small compared with the expected losses to those accepting counterfeits. If genuine notes are very costly to counterfeit, then forgeries are not profitable. In contrast, the expected

---

7 The term "money" usually refers to the medium of exchange used to facilitate transactions between individual parties. In this article, money refers to bank notes issued by the monetary authority.

losses to consumers from counterfeits of low-denomination notes are small compared with the cost of producing secure notes. The central bank therefore produces less-secure low-denomination notes, and counterfeiters respond by forging a large number of these notes. The study implies that central banks tend to issue more-secure high-denomination notes, and thus counterfeiting rates are higher for low-denomination notes.

> *The denominations that are most counterfeited tend to change over time*

Empirical support for Lengwiler's main findings is mixed. In the United States, it is true that the security features of $1 bills have been changed only slightly in the past few decades. Contrary to Lengwiler's assertion, however, their very low face value also makes them unattractive to counterfeiters. In Canada, some previous series of bank notes did have better security features on the high-denomination notes, consistent with Lengwiler's argument. However, in the *Canadian Journey* series issued from 2004 to 2006, all denominations carry the same security features. Moreover, the denominations that are most counterfeited also tend to change over time. For example, $10 notes were the most counterfeited in the early 2000s, when they were less secure than others, which supports Lengwiler's findings. Since 2008, however, the counterfeiting of $100 notes has been higher than that of other denominations, which contradicts the argument that low-denomination notes are more often counterfeited.

Quercioli and Smith (2009) investigate the strategic interaction between merchants and counterfeiters in an environment where counterfeiters select the quality of counterfeit notes to produce and merchants verify the notes offered to avoid counterfeits. The security level of notes is given, since the central bank is not explicitly modelled. High-quality counterfeits are less likely to be detected but are more costly to produce. Counterfeiters also face expected costs of legal punishment. Verification by merchants is imperfect and costly, and a greater effort by merchants results in better detection.

Quercioli and Smith find that the behaviour of merchants and counterfeiters varies with note denomination. They derive three main implications from their model. First, there is no counterfeiting of low-denomination notes, since the expected gain is not large enough to cover the expected costs.

Counterfeiting occurs only in the case of high-denomination notes. Second, merchants choose to exert more effort when verifying high-denomination notes because the losses from accepting counterfeits are larger. At the same time, counterfeiters also produce higher-quality counterfeits for high-denomination notes, since the marginal gain to quality is higher. While both verification effort and counterfeit quality increase as the denomination of the notes rises, the authors find that the rise in counterfeit quality can dominate the increase in the verification effort under certain conditions. In other words, the ratio of passed counterfeit notes to all counterfeit notes (seized and passed) rises with the denomination. Third, the authors find that the counterfeiting rate, measured as the fraction of counterfeit notes to the total notes in circulation, displays a hump-shaped distribution across denominations, which is consistent with available counterfeiting data.

> *If people think that only high-denomination notes are counterfeited, they may avoid them and use more low-denomination notes*

While the above analyses derive interesting and useful implications regarding counterfeiting, some of which are consistent with the stylized facts, both studies assume that the demand for money is fixed and thus will not influence the decision to counterfeit. However, if people think that only high-denomination notes are counterfeited, they may avoid them and use more low-denomination notes. In this case, counterfeiters may prefer to produce low-denomination counterfeits because they are more easily passed. Indeed, when a large number of $100 counterfeit notes were detected in Canada in 2001, up to one in ten retailers in some regions displayed signs informing customers that they would no longer accept $100 notes. Counterfeiters likely found it more difficult to pass $100 counterfeits and thus might have had less incentive to produce them. This suggests the need for a more complete modelling approach, which we discuss in the next section.

General-equilibrium models

Models in the second category are referred to as general-equilibrium models because the model environment that generates money as the medium of exchange is explicitly specified, and the demand

for money depends on the interactions of agents in the model.

The relatively few general-equilibrium models of counterfeiting share certain common features. There are two types of money: (i) genuine money, which is supplied by the monetary authority and lasts forever; and (ii) counterfeit money produced privately, which may last for one or more time periods. In each period, buyers and sellers have a chance to meet and trade. If buyers do not have genuine money, they can produce counterfeit money at a cost. Sellers will accept genuine money, but they may or may not accept counterfeit money. If sellers refuse to trade with buyers that use counterfeit money, they will have to wait until the next period to meet another buyer.

Once counterfeiting has been established as an equilibrium outcome, the model can be used to study [its] effects and the effectiveness of policies

An important first step in modelling counterfeiting in a general-equilibrium framework is to examine whether both genuine and counterfeit money will be accepted in equilibrium, given the economic environment of the model. In this case, some buyers use genuine money while others decide to produce counterfeit money. Sellers accept money as payment, even though there is a possibility of receiving counterfeits. Once counterfeiting has been established as an equilibrium outcome, the model can be used to study the effects of counterfeiting on social welfare and to assess the effectiveness of policies in reducing counterfeiting.

In earlier models by Kultti (1996), Green and Weber (1996), Williamson (2002), Monnet (2005), and Cavalcanti and Nosal (2007), counterfeiting is defined as the private provision of money. Counterfeit money can last for more than one period and so will have value if people are willing to accept it as a medium of exchange. Sellers may knowingly accept counterfeit money in a bilateral trade when there is a shortage of genuine money.[8] This is because they may have to

wait a long time before they meet a buyer with genuine money, and they expect other sellers to accept counterfeit money in such a situation.

Monnet (2005) argues that central banks should aim to reduce counterfeiting, even if counterfeit money has positive value as private money. In his model, genuine notes are less costly to produce than counterfeits because of factors such as economies of scale. Thus, from a social-welfare viewpoint, it is better to have trades intermediated by genuine money. He also finds that if the cost of counterfeiting is low, more counterfeits will be in circulation, and thus the money stock will rise. Counterfeiting can therefore lead to inflationary pressures, and high inflation can impose a large social cost.[9]

In practice, however, society does not consider counterfeits as private money. Private money is issued by a reputable private institution, such as a commercial bank, and is backed by the issuer's assets. Its design differs from that of government-issued money and so is easily identified. In contrast, counterfeiters aim to produce notes that look like government-issued notes so that unsuspecting merchants will accept them. It is against the law to produce and pass such counterfeits. Therefore, sellers will not accept or recirculate counterfeits if they know the notes are fake. One way to capture these features in a model is to assume that counterfeit money can last for only one period and will be completely confiscated by the authorities at the end of each period. Counterfeit money therefore has no value because it cannot serve as a medium of exchange. Thus, it can pass to sellers only when they do not recognize that it is a counterfeit.[10]

For counterfeiting to exist as an equilibrium outcome when counterfeit money cannot circulate across periods, the method that buyers and sellers use to trade with each other is critical. In Nosal and Wallace (2007) and Li and Rocheteau (2011), a buyer can make a take-it-or-leave-it offer to the seller, which specifies the quantity of goods traded and the amount of money paid. Both papers find that counterfeiting does not occur in a monetary equilibrium because buyers holding authentic notes can signal that they are using genuine money by offering to pay a small sum of money in exchange for a very small quantity of goods. No counterfeiter will make

---

8 Earlier models of money tended to assume that money is indivisible in the sense that an agent can carry only one unit of money in each period, and thus there could be a shortage of money. Wallace and Zhou (1997) constructed a model to explain the currency shortages that were common in many parts of the world until the mid-nineteenth century. While a shortage of money is unlikely in modern economies, the development of private digital money, such as Facebook credits and Bitcoins, could reflect a shortage of "money" in other areas of trade, particularly online commerce, in which bank notes are not accepted.

9 Friedman (1960) also argues that counterfeiting and the private issue of money can lead to inflation.
10 The inability to recognize even poor-quality counterfeits can occur in situations where cashiers are inexperienced or where they choose not to verify notes.

such an offer because the gain from the trade is too small to cover the cost of counterfeiting. Therefore, no one will produce counterfeits. Li and Rocheteau (2011) also consider the case in which both buyers and sellers take turns to make a take-it-or-leave-it offer to the other party. When sellers make the offer, they are willing to trade with a buyer even when they cannot verify the quality of money, since the buyer may be holding genuine notes. Thus, counterfeits can exist alongside genuine money in equilibrium. Fung and Shao (2011) consider a model environment in which sellers post offers to attract buyers, and buyers search for sellers based on these offers. In this case, sellers will not be able to use their offers to screen buyers, and all sellers will post the same offers.[11] As a result, counterfeiting can occur in equilibrium since a seller will attract both counterfeiters and buyers with genuine money.

> *A policy that deters counterfeiting*
> *is needed to maintain the public's*
> *confidence in money*

According to Nosal and Wallace (2007), the threat of counterfeiting itself can have a significant negative impact on the economy. The possibility of counterfeiting threatens the use of money in equilibrium when no one is willing to accept money as a medium of exchange. Therefore, a policy that deters counterfeiting is needed to maintain the public's confidence in money. Li and Rocheteau (2011) argue that, while counterfeiting does not pose a threat to the acceptance of money as a medium of exchange, it lowers the volume of economic transactions, since sellers are concerned about receiving counterfeits. This, in turn, will reduce production and further decrease economic transactions. Anti-counterfeiting policies can prevent these effects by mitigating the threat of counterfeiting.

Policies against counterfeiting

In addition to the direct losses experienced by sellers who accept worthless notes, counterfeiting has indirect costs for society, such as the costs of producing more-secure notes and providing effective law enforcement, as well as the potential loss of confidence in bank notes.

Green and Weber (1996) study the effectiveness of introducing a new series of bank notes that is very difficult to counterfeit, such as one with advanced security features. Only old-style money can be counterfeited at some cost. The old-style money in circulation is gradually replaced by new-style money, as is done in the United States. The authors find that the introduction of new-style money may not always be effective in reducing counterfeiting because sellers may knowingly accept counterfeits (e.g., if there is a currency shortage, as explained earlier). Nevertheless, if counterfeits are too costly to produce, they will eventually go out of circulation. Thus, if old notes are not immediately withdrawn when a new series of more-secure bank notes is issued, additional anti-counterfeiting measures, such as enhanced law enforcement that increases the cost of counterfeiting, are also needed to effectively reduce counterfeiting.

Many of the other studies considered in this article also discuss the importance of policies against counterfeiting, and their insights can be summarized as follows:

(i) Policies aimed at raising the cost of counterfeiting, such as increased bank note security or enhanced law enforcement, can reduce the incentive to produce counterfeits. Policies aimed at improving note verification by merchants and withdrawing counterfeit notes from circulation can also stop these notes from entering into circulation or from being recirculated.

(ii) The implementation of anti-counterfeiting measures is crucial, even when the level of counterfeiting is low, because the threat of counterfeiting can affect the public's acceptance of bank notes as a means of payment, as well as reduce production, economic transactions and social welfare.

(iii) A comprehensive anti-counterfeiting strategy that encompasses a range of measures, including the regular issuance of a new series of bank notes, is more effective in deterring counterfeiting than any one single measure.

Conclusion

This article has summarized some relevant insights from studies that model counterfeiting. The policy implications of these studies support our understanding of counterfeiting and the Bank's comprehensive anti-counterfeiting strategy.

Many of the questions raised by the stylized facts derived from the experience with counterfeiting

---

11 Sellers also face a capacity constraint since they can serve only a limited number of buyers in each period. Buyers may therefore have to take into account how likely it is that they will be served by the seller they visit. If buyers do not pay attention to this constraint, as in Guerrieri, Shimer and Wright (2010), counterfeiting may not occur.

remain unanswered, however. Future research using general-equilibrium models of counterfeiting with less-restrictive assumptions would be useful in addressing these matters. It would also be very helpful to have more data available for empirical work. Such work would help us to better understand why counterfeiting episodes break out and why experience varies across countries. It would also improve our ability to assess the effectiveness of various policies against counterfeiting.

Literature Cited

Arango, C., D. Hogg and A. Lee. "Why Is Cash (Still) So Entrenched? Insights from the 2009 Methods of Payment Survey." Bank of Canada Discussion Paper (forthcoming).

Bank of Canada. 2011. *Annual Report 2010*. Ottawa: Bank of Canada.

Cavalcanti, R. and E. Nosal. 2007. "Counterfeiting as Private Money in Mechanism Design." Federal Reserve Bank of Cleveland Working Paper No. 07-16.

Chant, J. 2004. "The Canadian Experience with Counterfeiting." *Bank of Canada Review* (Summer): 41–54.

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

Fung, B. and E. Shao. 2011. "Counterfeit Quality and Verification in a Monetary Exchange." Bank of Canada Working Paper No. 2011-4.

Green, E. and W. Weber. 1996. "Will the New $100 Bill Decrease Counterfeiting?" *Federal Reserve Bank of Minneapolis Quarterly Review* 20 (3): 3–10.

Guerrieri, V., R. Shimer and R. Wright. 2010. "Adverse Selection in Competitive Search Equilibrium." *Econometrica* 78 (6): 1823–62.

Kultti, K. 1996. "A Monetary Economy with Counterfeiting." *Journal of Economics* 63 (2): 175–86.

Lengwiler, Y. 1997. "A Model of Money Counterfeits." *Journal of Economics* 65 (2): 123–32.

Li, Y. and G. Rocheteau. 2011. "On the Threat of Counterfeiting." *Macroeconomic Dynamics* 15 Supplement: 10–41.

Monnet, C. 2005. "Counterfeiting and Inflation." European Central Bank Working Paper No. 512.

Moxley, J., H. Meubus and M. Brown. 2007. "The *Canadian Journey*: An Odyssey into the Complex World of Bank Note Production." *Bank of Canada Review* (Autumn): 47–55.

Nosal, E. and N. Wallace. 2007. "A Model of (the Threat of) Counterfeiting." *Journal of Monetary Economics* 54 (4): 994–1001.

Quercioli, E. and L. Smith. 2009. "The Economics of Counterfeiting." Available at the Social Sciences Research Network <http://ssrn.com/abstract=1325892>.

Wallace, N. and R. Zhou. 1997. "A Model of a Currency Shortage." *Journal of Monetary Economics* 40 (3): 555–72.

Williamson, S. 2002. "Private Money and Counter-feiting." *Federal Reserve Bank of Richmond Economic Quarterly* 88 (3): 37–57.

Bank of Canada Publications

Unless noted otherwise, all publications are available in print and on the Bank's website: bankofcanada.ca

Monetary Policy Report
(quarterly: January, April, July and October)

Financial System Review (June/December)

Bank of Canada Review
(quarterly: February, May, August and November)

Business Outlook Survey
(quarterly: January, April, July and October)

Senior Loan Officer Survey
(quarterly: January, April, July and October)*

Speeches and Statements by the Governor

Bank of Canada Banking and Financial Statistics
(monthly)*

Weekly Financial Statistics (published each Friday)*

Renewal of the Inflation-Control Target: Background Information

Annual Report

A History of the Canadian Dollar
James Powell (available at Can$8 plus applicable taxes) (2005)

By All Accounts: Outside Perspectives on the Bank of Canada**
A portrait of the Bank from the perspective of outside observers, showing how Canadians have perceived the performance of their central bank over the decades through the eyes of those who monitor its work on the public's behalf. (2010)

Beads to Bytes: Canada's National Currency Collection**
This volume explores the role of money in society through the lens of the National Currency Collection, an extraordinary repository of coins, bank notes and related paraphernalia from around the world. (2008)

More Than Money: Architecture and Art at the Bank of Canada**
A tour of the head office complex, highlighting the architecture, interior design and decoration, as well as elements of restoration and preservation. It also features pieces from the Bank's art collection. (2007)

The Art and Design of Canadian Bank Notes**
A journey behind the scenes to explore the demanding world of bank note design. (2006)

The Bank of Canada: An Illustrated History**
To celebrate the Bank's 70th anniversary, this book depicts the history of the Bank from 1935. (2005)

The Transmission of Monetary Policy in Canada
(1996)*

The Thiessen Lectures (January 2001)
Lectures delivered by Gordon G. Thiessen, Governor of the Bank of Canada 1994 to 2001

A Festschrift in Honour of David Dodge's Contributions to Canadian Public Policy (2009)

Bilingualism at the Bank of Canada
(published annually)

Planning an Evolution: The Story of the Canadian Payments Association, 1980–2002
James F. Dingle (June 2003)

Conference Proceedings
Conference volumes published up to and including April 2005 are available on the Bank's website. Print copies can be purchased for Can$15 plus applicable taxes. Papers and proceedings from Bank of Canada conferences, seminars and workshops held after April 2005 are now published exclusively on the Bank's website.

Technical Reports, Working Papers and Discussion Papers*
Technical Reports, Working Papers and Discussion Papers are published in the original language only, with an abstract in both official languages.

Discussion papers deal with finished work on technical issues related to the functions and policy-making of the Bank. They are of interest to specialists and other central bankers.

For further information contact:

Publications Distribution
Communications Department
Bank of Canada
234 Wellington St.
Ottawa, ON
Canada K1A 0G9
Telephone: 1 800 303-1282
Email address: publications@bankofcanada.ca

---

*   Available only on the Bank's website.
** Each Can$25 plus shipping costs. Sample pages are available on the Bank's website.