Volume 2, Number 1 March 2006

IJCB

INTERNATIONAL JOURNAL OF CENTRAL BANKING

IJCB INTERNATIONAL JOURNAL OF CENTRAL BANKING Volume 2, Number 1 March 2006

International Journal of Central Banking

Volume 2, Number 1 March 2006

The contents of this journal, together with additional materials provided by article authors, are available without charge at ***www.ijcb.org.***

The *International Journal of Central Banking* is published quarterly (ISSN: 1815-4654). Online access to the publication is available free of charge at **www.ijcb.org**. Individual print subscriptions are available at an annual rate of $100 (USD).

Print subscription orders may be placed online at www.ijcb.org, by phone (202-452-3245), via fax (202-728-5886), or by e-mail (editor@ijcb.org).

Renewals, claims, address changes, and requests for permission to reprint material from this journal should be addressed to:

International Journal of Central Banking
Publications Fulfillment, MS-127
Federal Reserve Board
Washington, DC 20551

Phone: 202-452-3245
Fax: 202-728-5886
E-mail: editor@ijcb.org

ISSN: 1815-4654

SIGMA: A New Open Economy Model for Policy Analysis*

Christopher J. Erceg, Luca Guerrieri, and Christopher Gust
Federal Reserve Board

In this paper, we describe a new multicountry open economy SDGE model named "SIGMA" that we have developed as a quantitative tool for policy analysis. We compare SIGMA's implications to those of an estimated large-scale econometric policy model (the FRB/Global model) for an array of shocks that are often examined in policy simulations. We show that SIGMA's implications for the near-term responses of key variables are generally similar to those of FRB/Global. Nevertheless, some quantitative disparities between the two models remain due to certain restrictive aspects of SIGMA's optimization-based framework. We conclude by using long-term simulations to illustrate some areas of comparative advantage of our SDGE modeling framework.

JEL Codes: E32, F41.

In the wake of the Lucas critique and the early real business cycle (RBC) literature, a wide gap emerged between the models examined in the academic literature and those utilized in policy analysis by most central banks. While central bank policy models retained

*We appreciate comments and suggestions from Tamim Bayoumi, Ralph Bryant, Pablo Burriel (discussant), Matthew Canzonieri, Lawrence Christiano, Behzad Diba, Martin Eichenbaum, Joe Gagnon, Jordi Galí, Fabio Ghironi, Dale Henderson, Ben Hunt, Steve Kamin, Doug Laxton, Andrew Levin, Stephen Murchison, Paolo Pesenti, Alessandro Rebucci, Trevor Reeve, Stephanie Schmitt-Grohe, Christopher Sims (discussant), Ralph Tryon, Martin Uribe, and participants in workshops at the Bank of Canada, Duke University, the Bank of Finland, the Federal Reserve Board, the Federal Reserve Bank of Dallas, Georgetown University, and the University of Montreal. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. Corresponding author: Erceg, telephone 202-452-2575, fax 202-872-4926. E-mail addresses: christopher.erceg@frb.gov, luca.guerrieri@frb.gov, christopher.j.gust@frb.gov.

a heavy emphasis on fitting short-run properties of the data, academic models increasingly heeded the methodological imperative of the RBC literature that required optimizing-agent foundations and model-consistent expectations. But the focus of the latter on coherent theoretical underpinnings came at the expense of empirical realism.

In recent years, there has been a surge of interest in developing optimization-based models that are more suited to fitting the data. Consistent with this more empirical orientation, "state-of-the-art" stochastic dynamic general equilibrium (SDGE) models have evolved to include a large array of nominal and real rigidities. Important work by Christiano, Eichenbaum, and Evans (2005) showed that their optimization-based model, which includes both sticky nominal wages and various types of adjustment costs in the expenditure components, is quite successful in accounting for the estimated effects of a monetary policy shock. Smets and Wouters (2003) demonstrated that the forecasting ability of a similar model appears comparable to that of an unconstrained Bayesian vector autoregression.

A salient motivation of this recent research has been to enhance the latitude of the SDGE models to contribute to policy analysis. In this vein, a number of central banks and other institutions such as the International Monetary Fund are in the process of developing microfounded models that can provide quantitative input into the policy process.[1] But while recent empirical work validating certain features of the microfounded models is encouraging, it remains an open question whether these models can yield plausible implications across the rather broad spectrum of shocks considered routinely in policy work.

In this paper, we address this question using a new multicountry open economy SDGE model (SIGMA) that we have developed for quantitative policy analysis. Our new model has its antecedents in the seminal open economy modeling framework of Obstfeld and Rogoff (1995). However, it includes many of the nominal and real frictions that have been identified as empirically important in the closed economy models of Christiano, Eichenbaum, and Evans (2005) and Smets and Wouters (2003), such as habit persistence in consumption

[1] The IMF's new SDGE model "GEM" is described by Laxton and Pesenti (2003). The Bank of England's new model "BEQM" is described by Harrison et al. (2005).

and adjustment costs in investment. Moreover, it incorporates analogous frictions relevant in an open economy framework, including both local currency pricing (e.g., Betts and Devereux [1996] or Devereux and Engel [2002]) and costs of adjusting trade flows.

Our approach consists of comparing the implications of SIGMA with those of the FRB/Global model, a large-scale econometric model used extensively in policy analysis at the Federal Reserve Board.[2] Our comparisons involve examining the impulse response functions to a number of shocks often considered in policy work. These include domestic monetary and fiscal shocks, a taste shock to home (U.S.) consumption demand, a shock to the risk premium in the uncovered interest parity equation, and alternative shocks to foreign demand. The large-scale models form an important benchmark for evaluating the new SDGE models both because they provide a reasonably accurate reduced-form characterization of the data and because they embed the priors of policymakers who have applied these models to policy questions for several decades.[3] While we do not exclude an eventual departure from the responses of the large-scale econometric models if the data provide sufficient grounds for doing so, it seems crucial that the new SDGE models not a priori rule out such responses due to arbitrary structural restrictions in the theoretical framework.

[2]While the FRB/Global model essentially has neoclassical properties in the long run, the behavioral equations are formulated to allow considerable flexibility in accounting for the short-run properties of the data. This model assumes that expectations are formed adaptively, i.e., expectations are derived from small-scale vector autoregressions. In our analysis, we focus on the implications for the U.S. block of the model, which consists of about 80 estimated behavioral equations and 300 identities; the foreign sector consists of 29 country blocks and roughly 4,000 equations. Brayton et al. (1997) provide a description of the model's basic structure. The U.S. block of the FRB/Global model, augmented with a small external sector, can also be run as an independent model (FRB/US) under either adaptive or rational expectations. See Brayton and Tinsley (1996) for an overview; Kiley (2001) describes the specification and estimation of the business investment sector; and Elmendorf and Reifschneider (2002) provide an application to fiscal policy.

[3]We also provide some comparisons to estimates from structural vector autoregressions (SVARs). However, comparisons with FRB/Global have the advantage that we can consider a larger set of shocks than typically analyzed in the SVAR literature. Moreover, estimates from the SVAR literature often show considerable divergence, making it difficult to gauge the appropriate benchmark.

Our model comparisons indicate that the short-run responses of SIGMA are qualitatively similar to FRB/Global for a large array of macroeconomic variables, including output, prices, interest rates, real exchange rates, and the trade balance. We highlight two features of our framework that are instrumental in giving SIGMA greater flexibility to generate responses that are more closely aligned with FRB/Global quantitatively, including more persistent responses of both real and nominal variables. These features include "information frictions," which posit agents as having incomplete information about the persistence of shocks, and non-Ricardian households.

With information frictions, agents use the Kalman filter to estimate the nature of shocks affecting the economy and to make projections. This simple learning mechanism typically implies gradual responses to underlying shocks that are similar to adaptive expectations models, while retaining the appealing property of model-consistent expectations. As emphasized by Erceg and Levin (2003), it implies that model dynamics may depend crucially on the credibility and transparency of policy changes.

Our model breaks Ricardian equivalence by assuming that some fraction of households simply consume their current after-tax disposable income.[4] Accordingly, the short-run fiscal multiplier associated with temporary increases in government spending exceeds unity and the response of private consumption is positive. The inclusion of *both* non-Ricardian agents and of information frictions in our model can account for a highly persistent fiscal multiplier similar to that in FRB/Global.[5]

But our comparisons also reveal some noticeable quantitative disparities between the SIGMA and FRB/Global models. First, import prices adjust much more quickly and completely to exchange rate changes in SIGMA, with full exchange rate passthrough after a few quarters. Second, SIGMA tends to imply substantially greater volatility in the expenditure components of GDP than FRB/Global. Finally, SIGMA tends to imply smaller and less persistent spillover effects from foreign shocks to the domestic economy (though the

[4]Our approach is similar to that of Galí, López-Salido, and Vallés (2004) and Mankiw (2000).

[5]The highly persistent fiscal multiplier and positive consumption response are in line with the empirical findings of Fatás and Mihov (2001) and of Blanchard and Perotti (2002).

disparity is small if the increase in foreign demand is investment driven).

We argue that these differences are particularly significant insofar as they stem from features of our SDGE framework that appear robust to reasonable departures from our benchmark calibration; specifically, they reflect certain restrictive theoretical constraints that are likely to hold in a broad class of current open economy SDGE models. For instance, high exchange rate passthrough after a few quarters is likely to characterize any model in which the desired price markup is either fixed (as in SIGMA) or exhibits only modest variation. In turn, the large volatility in expenditure components is partly attributable to high passthrough, and to a marked sensitivity of private absorption to very persistent changes in real interest rates. Finally, insofar as SIGMA relies on trade linkages to account for spillover effects from foreign demand disturbances—a feature common to most open economy SDGE models—it may understate the importance of spillovers arising through other channels, including financial linkages.

Notwithstanding some limitations, the SDGE framework possesses some key advantages over existing econometric models as a tool for policy analysis. The last section of our paper focuses on some of these advantages in the context of longer-term simulations of SIGMA, including simulations of cuts in distortionary tax rates, and of a productivity acceleration. One advantage of the SDGE framework is that it facilitates assessing how structural features of the economy affect its responses to shocks. For example, it is often of interest to consider how the effects of a tax cut depend on the elasticity of household labor supply, or on the extent to which households are Ricardian in their consumption behavior. Such experiments are more difficult to design in typical large-scale econometric models in which there may exist no clear linkage between structural features and reduced-form parameters. Another major advantage of the SDGE framework for policy questions is that it fully articulates the channels through which the economy returns to its balanced growth path following initial "imbalances." For instance, in the case of the productivity acceleration, we illustrate the economic forces that induce the trade balance to eventually move into surplus following an initial trade deficit, and discuss how different initial perceptions about the underlying shock influence adjustment dynamics.

The remainder of this paper is organized as follows. Section 1 presents our basic open economy model. The calibration is discussed in section 2. Section 3 compares short-run impulse response functions in SIGMA and FRB/Global for an array of shocks frequently considered in policy analysis. Section 4 examines long-run simulations in SIGMA under alternative calibrations. Section 5 concludes and provides a discussion of directions for future research.

1. The Model

Our model consists of two countries that may differ in size, but are otherwise isomorphic.[6] Hence, our exposition below focuses on the "home" country. Each country in effect produces a single domestic output good, although we adopt a standard monopolistically competitive framework to rationalize stickiness in the aggregate price level. While household utility depends on consumption of both the domestic output good and imported goods, it is convenient to assume that a competitive distribution sector purchases both inputs and simply resells a final consumption good to households. Similarly, we assume that competitive distributors combine the domestic output good with imports to produce a final investment good.

To introduce non-Ricardian consumption behavior, we assume that there are two types of households. One type of household maximizes welfare subject to an intertemporal budget constraint. These households own the entire capital stock, accumulate capital subject to adjustment costs, and exhibit habit persistence in their consumption decisions. They are also regarded as monopolistic competitors in the labor market to account for aggregate wage stickiness. The other type of household (the "hand-to-mouth" household) simply consumes its entire after-tax disposable income.

[6]For expositional purposes, we focus on a two-country variant of SIGMA. However, in actual policy analysis we typically use a variant with seven country blocks that includes the United States, the euro area, Japan, Canada, Mexico, Developing Asia, and a rest-of-the-world sector. Moreover, the expanded model incorporates features designed to account for the effects of oil shocks, which include allowing oil to enter both the consumption bundle of households, and the production function of firms (see Guerrieri 2005). For the shocks that we analyze in this study, the effects on the home country (i.e., the United States) are quantitatively very similar in the two-country model discussed below as in the larger policy model.

1.1 Firms and Price Setting

Production of Domestic Intermediate Goods. There is a continuum of differentiated intermediate goods (indexed by $i \in [0,1]$) in the home country, each of which is produced by a single monopolistically competitive firm. As in Betts and Devereux (1996), intermediate goods firms charge different prices at home and abroad (i.e., they practice local currency pricing). In the home market, firm i faces a demand function that varies inversely with its output price $P_{Dt}(i)$ and directly with aggregate demand at home Y_{Dt}:

$$Y_{Dt}(i) = \left[\frac{P_{Dt}(i)}{P_{Dt}}\right]^{\frac{-(1+\theta_p)}{\theta_p}} Y_{Dt}, \tag{1}$$

where $\theta_p > 0$, and P_{Dt} is an aggregate price index defined below. Similarly, in the foreign market, firm i faces the demand function

$$X_t(i) = \left[\frac{P^*_{Mt}(i)}{P^*_{Mt}}\right]^{\frac{-(1+\theta_p)}{\theta_p}} M^*_t, \tag{2}$$

where $X_t(i)$ denotes the foreign quantity demanded of home good i, $P^*_{Mt}(i)$ denotes the price that firm i sets in the foreign market (denominated in foreign currency), P^*_{Mt} is the foreign import price index, and M^*_t is aggregate foreign imports (we use an asterisk to denote foreign variables).

Each producer utilizes capital services $K_t(i)$ and a labor index $L_t(i)$ (defined below) to produce its respective output good. The production function is assumed to have a constant elasticity of substitution (CES) form:

$$Y_t(i) = \left(\omega_K^{\frac{\rho}{1+\rho}} K_t(i)^{\frac{1}{1+\rho}} + \omega_L^{\frac{\rho}{1+\rho}} (Z_t L_t(i))^{\frac{1}{1+\rho}}\right)^{1+\rho}. \tag{3}$$

The production function exhibits constant-returns-to-scale in both inputs, and technological progress Z_t is given by

$$Z_t = \exp(g_z t + z_t), \tag{4}$$

where z_t is a country-specific shock to the level of technology and g_z, the deterministic rate of technological growth, is assumed to be the same in both countries. Firms face perfectly competitive factor markets for hiring capital and labor. Thus, each firm chooses $K_t(i)$ and

$L_t(i)$, taking as given both the rental price of capital R_{Kt} and the aggregate wage index W_t (defined below). Firms can costlessly adjust either factor of production. Thus, the standard static first-order conditions for cost minimization imply that all firms have identical marginal cost per unit of output, MC_t.

We assume that the home and foreign prices of the intermediate goods are determined by Calvo-style staggered contracts (see Calvo 1983). In each period, a firm faces a constant probability, $1 - \xi_p$, of being able to reoptimize its price at home ($P_{Dt}(i)$) and $1 - \xi_{p,x}$ probability of being able to reoptimize its price abroad ($P^*_{Mt}(i)$). These probabilities are assumed to be independent across firms, time, and countries. If a firm is not allowed to optimize its prices, we follow Christiano, Eichenbaum, and Evans (2005) and assume the firm must reset its home price based on lagged aggregate inflation (i.e., $P_{Dt}(i) = \pi_{t-1} P_{Dt-1}(i)$, where $\pi_t = P_{Dt}/P_{Dt-1}$).[7] In foreign markets, if a firm cannot reoptimize its price, the price is changed according to an analogous rule (i.e., $P^*_{Mt}(i) = \pi^*_{Mt-1} P^*_{Mt-1}(i)$, where $\pi^*_{Mt} = P^*_{Mt}/P^*_{Mt-1}$). This form of lagged indexation is a mechanism for introducing inflation inertia into the key price-setting equations.

When firm i is allowed to reoptimize its price in the domestic market in period t, the firm maximizes

$$\widetilde{\mathbb{E}}_t \sum_{j=0}^{\infty} \xi_p^j \psi_{t,t+j} \left[V_{Dt+j} P_{Dt}(i) Y_{Dt+j}(i) - MC_{t+j} Y_{Dt+j}(i) \right]. \tag{5}$$

The operator $\widetilde{\mathbb{E}}_t$ represents the conditional expectation based on the information available to agents at period t. The firm discounts profits received at date $t+j$ by the state-contingent discount factor $\psi_{t,t+j}$; for notational simplicity, we have suppressed all of the state indices.[8]

[7] In alternative calibrations of SIGMA, we also consider the specification used by Yun (1996) and Erceg, Henderson, and Levin (2000) where $P_{Dt}(i) = \pi P_{Dt-1}(i)$ so that $V_{Dt+j} = \pi^j$ in the profit functional defined below. For this alternative calibration, prices are updated according to $P_{Mt}(i) = \pi^* P_{Mt-1}(i)$ in foreign markets.

[8] We define $\xi_{t,t+j}$ to be the price in period t of a claim that pays one dollar if the specified state occurs in period $t+j$ (see the household problem below); then the corresponding element of $\psi_{t,t+j}$ equals $\xi_{t,t+j}$ divided by the probability that the specified state will occur.

Also, V_{Dt+j} is defined by

$$V_{Dt+j} = \prod_{h=1}^{j} \pi_{t+h-1}. \tag{6}$$

The first-order condition for setting the contract price of good i in the home market is

$$\widetilde{\mathbb{E}}_t \sum_{j=0}^{\infty} \psi_{t,t+j} \xi_p^j \left(\frac{V_{Dt+j} P_{Dt}(i)}{(1+\theta_p)} - MC_{t+j} \right) Y_{Dt+j}(i) = 0. \tag{7}$$

Defining a similar profit functional to equation (5) for a firm's optimal choice of its contract price in the foreign market at date t, the associated first-order condition is

$$\widetilde{\mathbb{E}}_t \sum_{j=0}^{\infty} \psi_{t,t+j} \xi_{p,x}^j \left(\frac{e_{t+j} V_{Mt+j}^* P_{Mt}^*(i)}{(1+\theta_p)} - MC_{t+j} \right) X_{t+j}(i) = 0. \tag{8}$$

In equation (8), e_t is the price of a unit of foreign currency in terms of the home currency (so that a rise in e_t corresponds to a depreciation of the home currency), and V_{Mt+j}^* is defined as

$$V_{Mt+j}^* = \prod_{h=1}^{j} \pi_{Mt+h-1}^*. \tag{9}$$

Production of the Domestic Output Index. Because households have identical Dixit-Stiglitz preferences, it is convenient to assume that a representative aggregator combines the differentiated intermediate products into a composite home-produced good Y_{Dt}:

$$Y_{Dt} = \left[\int_0^1 Y_{Dt}(i)^{\frac{1}{1+\theta_p}} \, di \right]^{1+\theta_p}. \tag{10}$$

The aggregator chooses the bundle of goods that minimizes the cost of producing Y_{Dt}, taking the price $P_{Dt}(i)$ of each intermediate good $Y_{Dt}(i)$ as given. The aggregator sells units of each sectoral output index at its unit cost P_{Dt}:

$$P_{Dt} = \left[\int_0^1 P_{Dt}(i)^{\frac{-1}{\theta_p}} \, di \right]^{-\theta_p}. \tag{11}$$

We also assume a representative aggregator in the foreign economy who combines the differentiated home products $X_t(i)$ into a single index for foreign imports:

$$M_t^* = \left[\int_0^1 X_t\left(i\right)^{\frac{1}{1+\theta_p}}\, di\right]^{1+\theta_p}, \tag{12}$$

and sells M_t^* at price P_{Mt}^*:

$$P_{Mt}^* = \left[\int_0^1 P_{Mt}^*\left(i\right)^{\frac{-1}{\theta_p}}\, di\right]^{-\theta_p}. \tag{13}$$

Production of Consumption and Investment Goods. Final consumption goods are produced by a representative "consumption goods distributor." This firm combines purchases of domestically produced goods with imported goods to produce a final consumption good (C_t) according to a constant-returns-to-scale CES production function:

$$C_t = \left(\omega_C^{\frac{\rho_C}{1+\rho_C}} C_{Dt}^{\frac{1}{1+\rho_C}} + \left(1-\omega_C\right)^{\frac{\rho_C}{1+\rho_C}} \left(\varphi_{Ct} M_{Ct}\right)^{\frac{1}{1+\rho_C}}\right)^{1+\rho_C}, \tag{14}$$

where C_{Dt} denotes the consumption goods distributor's demand for the index of domestically produced goods, M_{Ct} denotes the distributor's demand for the index of foreign-produced goods, and φ_{Ct} reflects costs of adjusting consumption imports. The form of the production function mirrors the preferences of households over consumption of domestically produced goods and imports. Accordingly, the quasi-share parameter ω_C may be interpreted as determining household preferences for home relative to foreign goods, or equivalently, the degree of home bias in household consumption expenditure. Finally, the adjustment cost term φ_{Ct} is assumed to take the quadratic form

$$\varphi_{Ct} = \left[1 - \frac{\varphi_{M_C}}{2}\left(\frac{\frac{M_{Ct}}{C_{Dt}}}{\frac{M_{Ct-1}}{C_{Dt-1}}} - 1\right)^2\right]. \tag{15}$$

This specification implies that it is costly to change the proportion of domestic and foreign goods in the aggregate consumption bundle, even though the level of imports may jump costlessly in response to changes in overall consumption demand. It aims to capture the intuitively appealing notion that households may have limited ability in the short run to vary the mix of domestic relative to foreign goods in producing consumption services, even if longer-run substitution possibilities are more favorable. From an empirical perspective, our specification is consistent with evidence which suggests that imports adjust slowly in response to relative price changes, but respond rapidly to changes in real activity, e.g., McDaniel and Balistreri (2003).

Given the presence of adjustment costs, the representative consumption goods distributor chooses (a contingency plan for) C_{Dt} and M_{Ct} to minimize its discounted expected costs of producing the aggregate consumption good:

$$\min_{C_{Dt+k}, M_{Ct+k}} \widetilde{\mathbb{E}}_t \sum_{k=0}^{\infty} \psi_{t,t+k} \Bigg\{ (P_{Dt+k} C_{Dt+k} + P_{Mt+k} M_{Ct+k}) + P_{Ct+k} \Bigg[C_{t+k} - \left(\omega_C^{\frac{\rho_C}{1+\rho_C}} C_{Dt+k}^{\frac{1}{1+\rho_C}} + (1-\omega_C)^{\frac{\rho_C}{1+\rho_C}} (\varphi_{Ct+k} M_{Ct+k})^{\frac{1}{1+\rho_C}} \right)^{1+\rho_C} \Bigg] \Bigg\}. \tag{16}$$

The distributor sells the final consumption good to households at a price P_{Ct}, which may be interpreted as the consumption price index (or equivalently, as the shadow cost of producing an additional unit of the consumption good).

We model the production of final investment goods in an analogous manner. Thus, the representative "investment goods distributor" produces a final investment good by combining its purchases of domestically produced goods with purchases of foreign-produced goods according to a constant-returns-to-scale CES production function:

$$I_t = \left(\omega_I^{\frac{\rho_I}{1+\rho_I}} I_{Dt}^{\frac{1}{1+\rho_I}} + (1-\omega_I)^{\frac{\rho_I}{1+\rho_I}} (\varphi_{It} M_{It})^{\frac{1}{1+\rho_I}} \right)^{1+\rho_I}, \tag{17}$$

where I_{Dt} denotes the investment goods distributor's demand for the index of domestically produced goods, M_{It} denotes the distributor's demand for the index of foreign goods, and φ_{It} reflects costs of adjusting imports of investment goods. As in the case of consumption goods, the quasi-share parameter ω_I may be interpreted as determining the degree of home bias in the production of final investment goods. The adjustment cost φ_{It} takes a form that is analogous to the adjustment cost specification for imported consumption goods, so that

$$\varphi_{It} = \left[1 - \frac{\varphi_{M_I}}{2} \left(\frac{\frac{M_{It}}{I_{Dt}}}{\frac{M_{It-1}}{I_{Dt-1}}} - 1 \right)^2 \right]. \tag{18}$$

Investment goods distributors solve an intertemporal cost minimization problem isomorphic to that of consumption goods distributors. Each distributor sells the final investment good to households at a price P_{It}, which may be interpreted as the investment price index. This price may differ from the price index of the consumption good P_{Ct} because of differences in import composition, even in the absence of the adjustment costs for consumption and investment imports.

1.2 Households and Wage Setting

We assume a continuum of monopolistically competitive households (indexed on the unit interval), each of which supplies a differentiated labor service to the intermediate goods-producing sector (the only producers demanding labor services in our framework). It is convenient to assume that a representative labor aggregator (or "employment agency") combines households' labor hours in the same proportions as firms would choose. Thus, the aggregator's demand for each household's labor is equal to the sum of firms' demands. The aggregate labor index L_t has the Dixit-Stiglitz form

$$L_t = \left[\int_0^1 \left(\zeta_t N_t \left(h \right) \right)^{\frac{1}{1+\theta_w}} dh \right]^{1+\theta_w}, \tag{19}$$

where $\theta_w > 0$ and $N_t(h)$ is hours worked by a typical member of household h. Also, ζ_t is the size of a household of type h and evolves according to $\zeta_t = g_n \zeta_{t-1}$ (effectively, ζ_t and g_n determine the size and growth rate of the population). The aggregator minimizes the cost of producing a given amount of the aggregate labor index, taking each household's wage rate $W_t(h)$ as given, and then sells units of the labor index to the production sector at their unit cost W_t:

$$W_t = \left[\int_0^1 W_t(h)^{\frac{-1}{\theta_w}} dh\right]^{-\theta_w}. \tag{20}$$

It is natural to interpret W_t as the aggregate wage index. The aggregator's demand for the labor services of a typical member of household h is given by

$$N_t(h) = \left[\frac{W_t(h)}{W_t}\right]^{-\frac{1+\theta_w}{\theta_w}} L_t/\zeta_t. \tag{21}$$

We assume that there are two types of households: (i) households that make intertemporal consumption, labor supply, and capital accumulation decisions in a forward-looking manner by maximizing utility subject to an intertemporal budget constraint (FL households, for "forward-looking" households) and (ii) the remainder that simply consume their after-tax disposable income (HM households, for "hand-to-mouth" households). The HM households receive no capital rental income or profits, and choose to set their wage to be the average wage of optimizing households. Given that households of each type grow at the same rate, the share of each type of household in the population is fixed. We denote the share of FL households by ς and the share of HM households by $1-\varsigma$.

We consider first the problem faced by FL households. The utility functional for an optimizing representative member of household h is

$$\widetilde{\mathbb{E}}_t \sum_{j=0}^{\infty} \beta^j \left\{ \frac{1}{1-\sigma}\left(C_{t+j}(h) - \frac{\varkappa}{\varsigma}\frac{C^O_{t+j-1}}{\zeta_{t+j-1}} - \nu_{ct}\right)^{1-\sigma} \right.$$
$$\left. + \frac{\chi_0 Z^{1-\sigma}_{t+j}}{1-\chi}(1-N_{t+j}(h))^{1-\chi} + \frac{\mu_0}{1-\mu}\left(\frac{MB_{t+j+1}(h)}{P_{Ct+j}}\right)^{1-\mu}\right\}, \tag{22}$$

where the discount factor β satisfies $0 < \beta < 1$. As in Smets and Wouters (2003), we allow for the possibility of external habits, where each household member cares about its consumption relative to lagged aggregate consumption per capita of optimizing agents,

$\frac{C^O_{t-1}}{\varsigma\zeta_{t-1}}$. The period utility function depends on each member's current leisure $1 - N_t(h)$, his or her end-of-period real money balances, $\frac{MB_{t+1}(h)}{P_{Ct}}$, and a preference shock, ν_{ct}. We allow for preferences over leisure to shift with the level of technology so that the model is consistent with balanced growth, even if the subutility function over consumption is not logarithmic.[9]

Household h faces a flow budget constraint in period t which states that its combined expenditure on goods and on the net accumulation of financial assets must equal its disposable income:

$$\begin{aligned} P_{Ct}C_t(h) + P_{It}I_t(h) + MB_{t+1}(h) - g_n^{-1}MB_t(h) + \int_s \xi_{t,t+1}B_{Dt+1}(h) \\ -g_n^{-1}B_{Dt}(h) + P_{Bt}B_{Gt+1} - g_n^{-1}B_{Gt} + \frac{e_tP^*_{Bt}B_{Ft+1}(h)}{\phi_{bt}} - g_n^{-1}e_tB_{Ft}(h) \\ = (1-\tau_{Nt})W_t(h)N_t(h) + \Gamma_t(h) + TR_t(h) - T_t(h) \\ +(1-\tau_{Kt})g_n^{-1}R_{Kt}K_t(h) + P_{It}\tau_{Kt}\delta g_n^{-1}K_t(h) - P_{Dt}\phi_{It}(h). \end{aligned} \tag{23}$$

The presence of the population growth parameter g_n in the household's budget constraint reflects that equation (23) is expressed in per capita terms as well as the assumption that new household members are born without any initial holdings of bonds, capital, or money. Final consumption goods are purchased at a price P_{Ct}, and final investment goods at a price P_{It}. Investment in physical capital augments the per capita capital stock $K_{t+1}(h)$ according to a linear transition law of the form

$$K_{t+1}(h) = (1-\delta)g_n^{-1}K_t(h) + I_t(h), \tag{24}$$

where δ is the depreciation rate of capital.

Financial asset accumulation of a typical member of FL household h consists of increases in nominal money holdings $(MB_{t+1}(h) - g_n^{-1}MB_t(h))$ and the net acquisition of bonds. We assume that agents within a country can engage in frictionless trading of a complete set of contingent claims, while trade in international assets is restricted to a non-state-contingent nominal bond. The

[9]This statement is only strictly true in the absence of permanent country-specific technology shocks. In this case, a permanent increase in technology in the home country that does not occur abroad will be associated with a permanent deterioration in the home country's terms of trade that moves the home economy off its balanced growth path.

term $P_{Bt}B_{Gt+1} - g_n^{-1}B_{Gt}$ represents each household member's net purchases of domestic government bonds, while $\int_s \xi_{t,t+1}B_{Dt+1}(h) - g_n^{-1}B_{Dt}(h)$ is net purchases of state-contingent domestic bonds. We denote $\xi_{t,t+1}$ as the price of an asset that will pay one unit of domestic currency in a particular state of nature at date $t+1$, while $B_{Dt+1}(h)$ represents the quantity of such claims purchased by a typical member of household h at time t. Thus, the gross outlay on new state-contingent domestic claims is given by integrating over all states at time $t+1$, while $B_{Dt}(h)$ indicates the value of the household's existing claims (on a per capita basis) given the realized state of nature.

In equation (23), $B_{Ft+1}(h)$ represents the quantity of a non-state-contingent bond purchased by a typical member of household h at time t that pays one unit of foreign currency in the subsequent period, P_{Bt}^* is the foreign currency price of the bond, and e_t is the exchange rate expressed in units of home currency per unit of foreign currency. We follow Turnovsky (1985) and assume there is an intermediation cost ϕ_{bt} paid by households in the home country for purchases of foreign bonds, which ensures that net foreign assets are stationary in the model.[10] More specifically, the intermediation costs depend on the ratio of economy-wide holdings of net foreign assets to nominal GDP, P_tY_t, and are given by

$$\phi_{bt} = \exp\left(-\phi_b\left(\frac{e_tB_{Ft+1}}{P_tY_t}\right) + \nu_{bt}\right). \tag{25}$$

In the above, ν_{bt} is a mean-zero stochastic process, which we interpret as a risk-premium shock or shock to the uncovered interest-rate parity condition. Abstracting from this shock, if the home economy has an overall net lender position internationally, then a household will earn a lower return on any holdings of foreign bonds. By contrast, if the economy has a net debtor position, a household will pay a higher return on its foreign liabilities.

Each member of FL household h earns after-tax labor income, $(1-\tau_{Nt})W_t(h)N_t(h)$, where τ_{Nt} is a stochastic tax on labor income. The household leases capital to firms at the after-tax rental rate

[10] This intermediation cost is asymmetric, as foreign households do not face these costs; rather, they collect profits on the monopoly rents associated with these intermediation costs.

$(1-\tau_{Kt})R_{Kt}$, where τ_{Kt} is a stochastic tax on capital income. The household receives a depreciation write-off of $P_{It}\tau_{Kt}\delta$ per unit of capital. Each member also receives an aliquot share $\Gamma_t(h)$ of the profits of all firms and a lump-sum government transfer, $TR_t(h)$, and pays a lump-sum tax $T_t(h)$. Following Christiano, Eichenbaum, and Evans (2005), we assume that it is costly to change the level of gross investment from the previous period, so that the acceleration in the capital stock is penalized:

$$\phi_{It}(h) = \frac{1}{2}\phi_I \frac{(I_t(h) - g_z g_n I_{t-1}(h))^2}{I_{t-1}(h)}. \tag{26}$$

In every period t, each member of FL household h maximizes the utility functional (22) with respect to its consumption, investment, (end-of-period) capital stock, money balances, holdings of contingent claims, and holdings of foreign bonds, subject to its labor demand function (21), budget constraint (23), and transition equation for capital (24). In doing so, a household takes as given prices, taxes and transfers, and aggregate quantities such as lagged aggregate consumption and the aggregate net foreign asset position.

Forward-looking (FL) households set nominal wages in staggered contracts that are analogous to the price contracts described above. In particular, with probability $1-\xi_w$, each member of a household is allowed to reoptimize its wage contract. If a household is not allowed to optimize its wage rate, we assume each household member resets its wage according to

$$W_t(h) = \omega_{t-1} W_{t-1}(h), \tag{27}$$

where $\omega_t = W_t/W_{t-1}$ and in steady state $\omega = \pi g_z$.[11] Each member of household h chooses the value of $W_t(h)$ to maximize its utility functional (22), yielding the following first-order condition:

$$\begin{aligned} &\widetilde{\mathbb{E}}_t \textstyle\sum_{j=0}^{\infty} \beta^j \xi_w^j \{ \tfrac{1-\tau_{Nt}}{(1+\theta_w)} \tfrac{\Lambda_{t+j}}{P_{t+j}} V_{wt+j} W_t(h) \\ &-\chi_{0t+j} Z_{t+j}^{1-\sigma} (1 - N_{t+j}(h))^{-\chi} \} N_{t+j}(h) = 0, \end{aligned} \tag{28}$$

[11] In alternative specifications, we also consider $W_t(h) = \omega W_{t-1}(h)$.

where Λ_t is the marginal value of a unit of consumption, and V_{wt+j} is defined as

$$V_{wt+j} = \prod_{h=1}^{j} \omega_{t+h-1}. \tag{29}$$

Roughly speaking, equation (28) says that the household chooses its contract wage to equate the present discounted value of working an additional unit of time to the discounted marginal cost.

Finally, we consider the determination of consumption and labor supply of the hand-to-mouth (HM) households. A typical member of an HM household simply equates his or her nominal consumption spending to his or her current after-tax disposable income, which consists of labor income plus net lump-sum transfers from the government:

$$P_{Ct}C_t\left(h\right) = (1-\tau_{Nt})W_t\left(h\right)N_t\left(h\right) + TR_t(h) - T_t\left(h\right). \tag{30}$$

The HM households set their wage to be the average wage of the forward-looking households. Since HM households face the same labor demand schedule as the forward-looking households, each HM household works the same number of hours as the average for forward-looking households.

1.3 Monetary Policy

We assume that the central bank follows an interest rate reaction function similar in form to the historical rule estimated by Orphanides and Wieland (1998) over the Volcker-Greenspan period. Thus, the short-term nominal interest rate is adjusted so that the ex post real interest rate rises when inflation exceeds its constant target value, or when output *growth* rises above some target value. With some allowance for interest rate smoothing, monetary policy is described by the following interest rate reaction function:

$$i_t = \gamma_i i_{t-1} + \overline{r} + \overline{\pi}_t + \gamma_\pi(\pi_t^{(4)} - \overline{\pi}) + \gamma_y(y_t - y_{t-4} - g_y) + \epsilon_{it}. \tag{31}$$

In the above, i_t is the annualized nominal interest rate, $\pi_t^{(4)}$ is the four-quarter inflation rate of the GDP deflator (i.e., $\pi_t^{(4)} = \sum_{j=0}^{3}\pi_{t-j}$), and $\overline{r}$ and $\bar{\pi}$ are the steady-state real interest rate and the central bank's constant inflation target (both expressed at annual rate). Also, $y_t - y_{t-4}$ is the four-quarter growth rate of output, and g_y is its corresponding steady-state value.

1.4 Fiscal Policy

Some of the domestically produced good is purchased by the government. Government purchases (G_t) are assumed to have no direct effect on the utility of a household.[12] We also assume that government purchases as a fraction of output, $g_t = \frac{P_{Dt}G_t}{P_tY_t}$, follow an exogenous stochastic process.

The government can issue debt B_{Gt+1} to finance a deficit so that its budget constraint is given by

$$\begin{aligned} P_{Bt}B_{Gt+1} - B_{Gt} = \; & P_{Dt}G_t + TR_t - T_t - \tau_{Nt}W_tL_t \\ & -(\tau_{Kt}R_{Kt} - \delta P_{It})K_t \\ & -(MB_{t+1} - MB_t). \end{aligned} \quad (32)$$

In equation (32), we have aggregated the capital stock, money and bond holdings, and transfers and taxes over all households so that, for example, $T_t = \zeta_t \int_0^1 T_t(h)dh$. As noted above, labor and capital taxes are determined exogenously, while we assume that real transfers as a fraction of domestic output, $tr_t = \frac{TR_t}{P_tY_t}$, evolve according to an exogenous stochastic process. Given that the central bank uses the nominal interest rate as its policy instrument, the level of seigniorage revenues is determined by nominal money demand.

Lump-sum taxes are adjusted in a manner that the government satisfies an intertemporal solvency constraint, requiring that the present discounted value of the government debt stock tends toward zero in the long run. In particular, we assume that the real lump-sum tax rate, $\tau_t = \frac{T_t}{P_tY_t}$, is determined according to the following reaction function:

$$\tau_t = \nu_0\tau_{t-1} + \nu_1(b_{Gt+1} - b_G) + \nu_2(b_{Gt+1} - b_{Gt}), \quad (33)$$

[12] The model's dynamics would be unchanged if we had assumed instead that government purchases entered separately in the utility function, although the welfare consequences would be different.

where $b_{Gt+1} = \frac{B_{Gt+1}}{P_t Y_t}$ and b_G is the government's target value for the ratio of government debt to nominal output.[13]

1.5 Resource Constraint and Net Foreign Assets

The home economy's aggregate resource constraint can be written as

$$Y_{Dt} = C_{Dt} + I_{Dt} + G_t + \phi_{It}, \tag{34}$$

and ϕ_{It} is the adjustment cost on investment aggregated across all households. Total exports may be allocated to either the consumption or the investment sector abroad:

$$M_t^* = M_{Ct}^* + M_{It}^*. \tag{35}$$

Finally, at the level of the individual firm,

$$Y_t(i) = Y_{Dt}(i) + X_t(i) \qquad \forall i. \tag{36}$$

The evolution of net foreign assets can be expressed as

$$\frac{e_t P_{B,t}^* B_{F,t+1}}{\phi_{bt}} = e_t B_{F,t} + e_t P_{Mt}^* M_t^* - P_{Mt} M_t. \tag{37}$$

This expression can be derived from the budget constraint of the FL households after imposing the government budget constraint, the consumption rule of the HM households, the definition of firm profits, and the condition that domestic bonds (B_{Dt+1}) are in zero net supply.[14]

Finally, we assume that the structure of the foreign economy (the "rest of the world") is isomorphic to that of the home country.

[13]We found that the inclusion of the term $(b_{Gt+1} - b_{Gt})$ was instrumental in yielding a determinate rational expectations equilibrium over a large region of the parameter space.

[14]The derivation of the evolution of net foreign assets also requires that $P_{Ct}C_t = P_{Dt}C_{Dt} + P_{Mt}M_{Ct}$ and $P_{It}I_t = P_{Dt}I_{Dt} + P_{Mt}M_{It}$. Given that these conditions are satisfied even in the presence of adjustment costs for imports, the import adjustment cost terms do not appear in equation (37).

1.6 Description of Shocks and the Optimal Filtering Problem

As discussed above, our model includes shocks to the level of home productivity (Z_t), real government spending (g_t), real transfers (tr_t), labor tax rate (τ_{Nt}), capital tax rate (τ_{Kt}), preferences (ν_{ct}), and similar shocks to the foreign country (which may be regarded as alternative sources of foreign demand shocks from the perspective of the home country). In addition, it includes a shock to the financial intermediation technology (ν_{bt}), which can be interpreted as a risk premium shock.

We assume that each shock in the model has two underlying components with different levels of persistence. Agents observe the shock, but they cannot observe the separate components. It is helpful to illustrate the nature of the forecasting problem with reference to a particular shock, since the treatment of the other shocks is basically isomorphic. Thus, focusing on the case of the productivity growth shock examined in section 4 below, and recalling from equation (4) that productivity growth can be written as $\Delta\log(Z_t) = g_z + \Delta z_t$, our approach attributes changes in z_t to two separate components:

$$\Delta z_t = \Delta z_{Pt} + \Delta z_{Tt}. \tag{38}$$

The first component, Δz_{Pt}, is a highly persistent shock that shifts the "trend" level of productivity growth, and the second, Δz_{Tt}, is a transient shock to productivity growth. The bivariate process determining the evolution of each component may be represented as

$$\begin{bmatrix} \Delta z_{Pt} \\ \Delta z_{Tt} \end{bmatrix} = \begin{bmatrix} \rho_p & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta z_{Pt-1} \\ \Delta z_{Tt-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{Pt} \\ \varepsilon_{Tt} \end{bmatrix}, \tag{39}$$

where the persistence parameter ρ_p is assumed to be slightly below unity, and the transient shock is assumed to be i.i.d. in the case of this particular shock (for other shocks, e.g., to government spending, we allow the transient component to be somewhat persistent, though much less persistent than the permanent shock). Thus, an innovation to the growth rate of the transient component has a permanent effect on the level of productivity, but no effect on the future growth rate. Moreover, the innovations ε_{Pt} and ε_{Tt} are mutually uncorrelated with variances v_P and v_T, respectively.

Agents observe the current level of productivity in the economy, and hence observe Δz_t, but cannot observe the growth rate of the underlying components Δz_{Pt} and Δz_{Tt}. Thus, agents solve a signal extraction problem to forecast the future level of productivity. Given that agents know the underlying parameters of the bivariate process for productivity growth, the Kalman filter can be used to obtain an optimal solution.

The expected level of productivity at a future date K periods ahead depends only on the current level of productivity and on the expected growth rate of the permanent component:

$$\begin{aligned}\widetilde{\mathbb{E}}_t \log(Z_{t+K}) &= Kg_z + \log(Z_t) + \widetilde{\mathbb{E}}_t(z_{Pt+k} - z_{Pt}) \\ &= Kg_z + \log(Z_t) + \sum_{J=1}^{K} \widetilde{\mathbb{E}}_t \Delta z_{Pt+J}. \qquad (40)\end{aligned}$$

The Kalman filtering algorithm implies that the expected growth rate of the permanent component is updated according to

$$\widetilde{\mathbb{E}}_t \Delta z_{Pt} = \rho_p \widetilde{\mathbb{E}}_{t-1} \Delta z_{Pt-1} + k_g \rho_p (\Delta z_t - \rho_p \widetilde{\mathbb{E}}_{t-1} \Delta z_{Pt-1}). \qquad (41)$$

Thus, agents update their assessment of the persistent component of the productivity growth process by the product of the forecast error innovation and a constant coefficient. This coefficient, which is proportional to the scalar Kalman gain parameter k_g, is an increasing function of the signal-to-noise ratio $\frac{v_P}{v_T}$ (the ratio of the variances of the persistent and transitory components of the productivity growth process).

2. Solution Method and Calibration

Because the levels of technology and the population are nonstationary, real variables (including output and the expenditure components of GDP) are also nonstationary. Accordingly, prior to solving the model, we scale real variables in the home and foreign country by the common deterministic trends in technology and population size. Nominal variables are scaled to account both for growth in the corresponding real variable and for the steady-state inflation rate.

We solve the model by log-linearizing the equations (specified in terms of the transformed variables) around the steady state associated with common growth rates of technology and population in the two countries. To obtain the reduced-form solution of the model, we use the numerical algorithm of Anderson and Moore (1985), which provides an efficient implementation of the method proposed by Blanchard and Kahn (1980) (see also Anderson 1997).[15]

2.1 Calibration of Parameters

The model is calibrated at a quarterly frequency. Structural parameters are set at identical values for each of the two countries, except for the parameters determining population size (as discussed below). We assume that the discount factor $\beta = 0.997$ and the rate of technological growth $g_z = 1.0037$. These values are consistent with a steady-state annualized real interest rate $\overline{r}$ of about 3 percent.

The utility functional parameter σ is set equal to 2, while the parameter determining the degree of habit persistence in consumption $\varkappa = 0.8$. We set $\chi = 10$, implying a Frisch elasticity of labor supply of 1/5, which is considerably lower than if preferences were logarithmic in leisure, but well within the range of most empirical estimates. The utility parameter χ_0 is set so that employment comprises one-third of the household's time endowment, while the parameter μ_0 on the subutility function for real balances is set at an arbitrarily low value (although given the separable specification, variation in real balances has no impact on other variables). We choose $\varsigma = 0.5$ so that 50 percent of households are Ricardian FL agents.[16]

The depreciation rate of capital δ is set at 0.025 (consistent with an annual depreciation rate of 10 percent). We fix the price and wage markup parameters so that $\theta_p = \theta_w = 0.20$, similar to the estimated values obtained by Rotemberg and Woodford (1999) and

[15]The steady state around which we linearize depends on the relative level of technology in each country, which we initialize to unity (so that per capita income in each country is identical in the steady state, though GDP may differ across countries due to population differences). We evaluated the robustness of our solution procedure by using a nonlinear Newton-Raphson algorithm that does not rely on linearization around an initial steady state, and found that the results were nearly identical to those reported.

[16]Galí, López-Salido, and Vallés (2004) found in a New Keynesian sticky price model with HM agents that the equilibrium becomes indeterminate when the share of HM agents is substantial. By contrast, our model produces a determinate rational expectations solution even for much higher values of the share of HM households than the 50 percent assumed in our baseline.

Amato and Laubach (2003).[17] We set ξ_p and ξ_w to be consistent with four-quarter contracts (subject to full indexation).[18] The parameter $\xi_{p,x}$ is chosen to be consistent with two-quarter contracts.[19] We set the steady-state inflation rate π to yield an annual inflation rate of 4 percent.

The parameter ρ in the CES production function of the intermediate goods producers is set to -2, implying an elasticity of substitution between capital and labor of 1/2. Thus, capital and labor are less substitutable than the unitary elasticity case implied by the Cobb-Douglas specification. The quasi-capital share parameter ω_K is chosen to imply a steady-state investment to output ratio of 20 percent. The private consumption to output ratio is 70 percent, while government consumption is 10 percent of steady-state output. We set the cost of adjusting investment parameter $\phi_I = 3$, slightly below the value used by Christiano, Eichenbaum, and Evans (2005).

The parameter ω_C is chosen to match the estimated average share of imports in total U.S. consumption of about 9 percent (based on data from the U.S. Bureau of Economic Analysis), while the parameter ω_I is chosen to match the average share of imports in total U.S. investment of about 38 percent. Given that trade is balanced in steady state, this parameterization implies an import or export to GDP ratio for the home country (the United States) of about 13 percent. We choose the initial population levels ζ_0 and ζ_0^* so that the home country constitutes 25 percent of world output. This implies an import (or export) share of output of the foreign country of about 3 percent.

We assume that $\rho_C = \rho_I = 2$, consistent with a long-run price elasticity of demand for imported consumption and investment goods of 1.5. While this is higher than most empirical estimates using

[17] Rotemberg and Woodford (1999) found $\theta_p = 0.15$, while Amato and Laubach (2003) obtained $\theta_p = 0.19$ and $\theta_w = 0.13$. Given our assumption that there is perfect capital mobility across firms within a country, the parameter θ_p only affects steady-state relationships and does not otherwise appear in the dynamic equations of the log-linearized model.

[18] The inclusion of strategic complementarities along the lines suggested by Kimball (1995) and Woodford (2003) would allow our model to generate similar dynamic responses with shorter contract durations.

[19] The rapid adjustment of import prices is consistent with the evidence Campa and Goldberg (2004) derived from a panel of OECD countries, nothwithstanding their finding that long-run passthrough is generally well below 100 percent for OECD countries.

macrodata, we emphasize that the presence of adjustment costs translates into a much lower relative price sensitivity in the short to medium term. In particular, we set the adjustment cost parameters $\varphi_{M_C} = \varphi_{M_I} = 10$, implying a price elasticity near unity after four quarters. We choose a small value (0.001) for the financial intermediation cost ϕ_b, which is sufficient to ensure the model has a unique steady state.

We estimated the parameters of the monetary policy rule using U.S. data from 1983:1–2003:4.[20] Our estimates imply $\gamma_\pi = 0.6$, $\gamma_y = 0.28$, and $\gamma_i = 0.8$. For the tax rate reaction function, we choose $\nu_0 = 1$, $\nu_1 = 0.1$, $\nu_2 = 0.001$, and $b_G = 0$. We set the steady-state capital and labor tax rates equal to 0.3 and 0.2, respectively.

3. Comparisons of SIGMA and FRB/Global

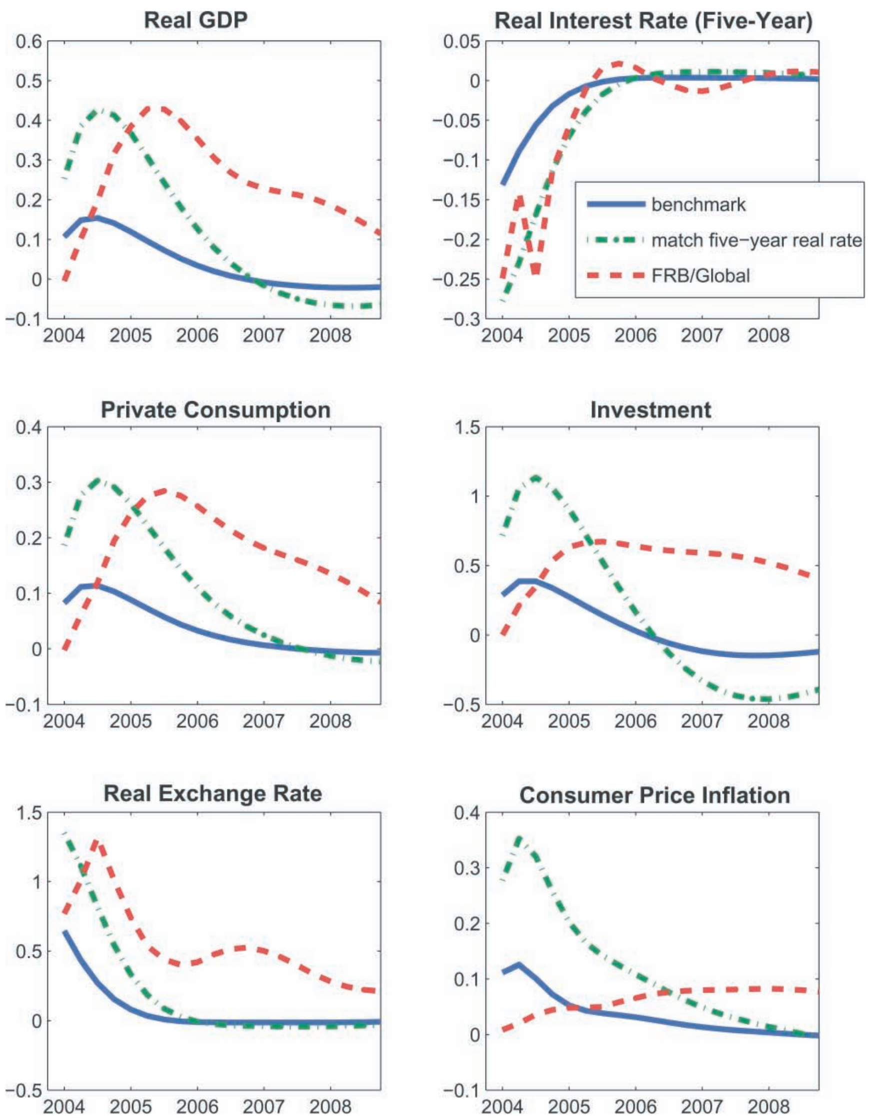
We now turn to assessing the short-run properties of our SIGMA model, comparing the implications of SIGMA with those of FRB/Global for an array of shocks often considered in policy analysis. To facilitate comparison across models, we specify the monetary policy rule in FRB/Global to be the same as in the benchmark version of SIGMA.

3.1 Loosening of Monetary Policy

We begin by examining the effects of a transient innovation to the monetary policy rule in SIGMA, i.e., a rise in ϵ_{it} in equation (31). The shock is scaled to induce an initial decline in the short-term nominal interest rate of about 75 basis points. As shown in figure 1, this policy shock raises output by slightly less than 0.2 percent after two to three quarters. The response of investment is roughly three times larger than the response of consumption, reflecting the higher interest sensitivity of the former. The fall in domestic real interest rates drives a modest depreciation of the real exchange rate (a rise in the figure), which pushes up import prices (not shown). The combination of a positive output gap and higher import prices causes consumer price inflation to rise.

[20] We estimated the rule using instrumental variables with lags of inflation and output growth as instruments.

Figure 1. Monetary Policy Loosening



While the figure shows that the qualitative effects of the shock are similar in FRB/Global, the peak effects on real GDP and the expenditure components are noticeably larger in FRB/Global than in SIGMA. This primarily reflects that long-term real interest rates (which are determined by a small-scale vector autoregression in FRB/Global) drop much more sharply and persistently than in SIGMA (in which long-term real rates are effectively determined by the expectations hypothesis). Given this disparity in the long-term real interest rate responses, it is useful to control for differences in the simulation results that are attributable to alternative term structure equations. Accordingly, the dash-dotted line in the figure shows an alternative calibration of the shock in SIGMA that implies a response of the five-year real interest rate that comes very close to matching that in FRB/Global (this calibration assumes that the persistence of the monetary policy error ϵ_{it} in the interest rate reaction function is 0.6 rather than 0 as in our benchmark). In this case, the peak responses of GDP and the real expenditure components are very close across the two models; the notable difference is in the inflation response, which is much larger in SIGMA.

These results indicate that the responses of the two models to a monetary policy shock that has similar effects on long-term real interest rates is quite similar across the two models (at least for real variables). However, we caution that the similarity in responses appears sensitive to the persistence of the underlying shock. Thus, the results should not be interpreted more broadly as indicating that the interest sensitivity of aggregate demand is similar across the two models, irrespective of persistence of the shock's effect on real interest rates. As we show below, SIGMA appears to exhibit a somewhat greater interest elasticity to shocks that exert highly persistent effects on long-term real interest rates.[21]

[21] The results from our SIGMA model also appear broadly consistent with the implications of the empirical vector autoregression (VAR) literature that estimates responses to a monetary policy innovation. For example, the results of Christiano, Eichenbaum, and Evans (2005) indicate that a monetary policy innovation that reduces the nominal interest rate by about 75 basis points induces output to rise about 0.4 percent, which is very close to our model for the calibration in which the monetary policy error follows an AR(1) with a persistence parameter of 0.6. However, direct comparisons of the quantitative effects are somewhat difficult because model results are fairly sensitive to the form of the monetary policy rule, which differs from that imposed in the VAR.

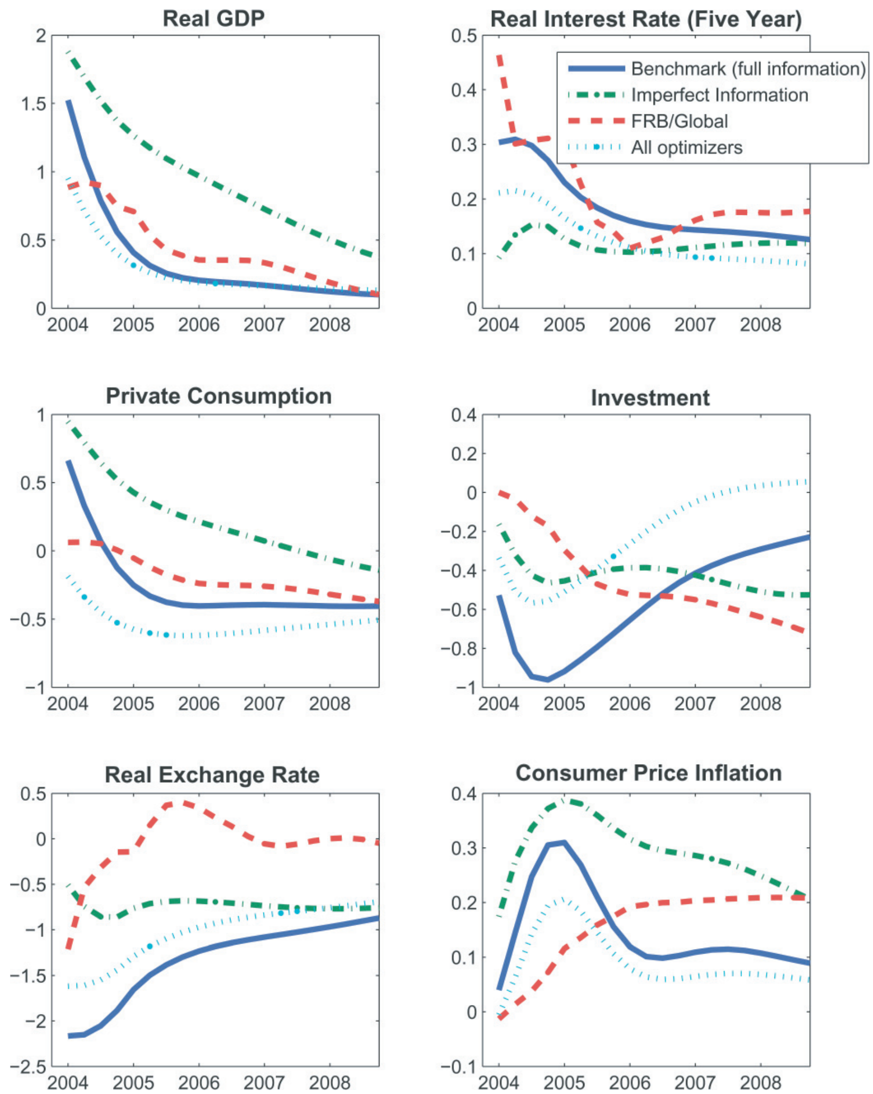
3.2 Rise in Government Spending

Figure 2 shows the effects of an exogenous rise in the U.S. government spending share of GDP of 1 percentage point relative to baseline. The shock is highly persistent ($\rho_g = 0.975$), so that the government spending share remains about 0.6 percentage point above baseline after five years.

The rise in government spending induces an immediate expansion of output. The government spending multiplier exceeds unity in the impact period due to the sharp rise in consumption of the HM households (as implied by the dotted line, the impact multiplier would be below unity if all households were optimizers, reflecting a sharp immediate fall in aggregate consumption). However, rising real interest rates quickly crowd out private investment and the consumption of the interest-sensitive optimizing households, which is depressed further due to the negative wealth effect of higher government spending. In fact, the fall in consumption for the optimizing households is exacerbated by the presence of the HM households, as real interest rates rise even more than would occur if all households were optimizers. Thus, overall private consumption falls below baseline after only a few quarters, and most of the output expansion is reversed. The small but more persistent component of the output increase reflects a rise in labor supply that is induced by the negative wealth effect.

Figure 2 also shows impulse responses derived from the FRB/Global model for a similar-sized rise in government spending. Clearly, the output response is noticeably more persistent in FRB/Global, as output remains nearly 0.4 percent above baseline three years after the shock occurs. The greater persistence of the output response reflects that the crowding out of private investment and consumption spending occurs much more gradually. This partly reflects that negative wealth effects play a less prominent role in depressing consumption in FRB/Global. In addition, private absorption (especially investment) is considerably less sensitive to persistent changes in the long-term real interest rate in FRB/Global, especially in the short run. Thus, while the responses in SIGMA and FRB/Global are similar qualitatively, the responses in the latter tend to exhibit noticeably greater persistence. The empirical vector autoregression evidence of Blanchard and Perotti (2002) also suggests that government

Figure 2. Rise in Government Spending



spending shocks exert highly persistent effects on output and on the expenditure components that appear more consistent with the FRB/Global responses.

A plausible channel for inducing greater persistence in the impulse responses of the SIGMA model is to allow for imperfect information, as also shown in the figure (dash-dotted lines).[22] Because agents initially perceive that the increase in government spending is temporary under imperfect information, there is a larger increase in GDP in this case than in the benchmark. Furthermore, the rise in GDP is more persistent, as agents only slowly update their beliefs about the persistence of the shock and are continually surprised by the higher-than-expected levels of government spending. Given both a less-pronounced rise in real longer-term interest rates and a smaller negative wealth effect (since agents think the government spending rise will be temporary), there is much less crowding out of consumption and investment spending—with consumption even remaining above baseline for a sustained duration. Thus, the inclusion of both non-Ricardian agents and information frictions would seem to provide a tractable channel for increasing the flexibility of SDGE models to account for responses that are similar to large-scale policy models, and to encompass the range of responses derived from empirical vector autoregressions.

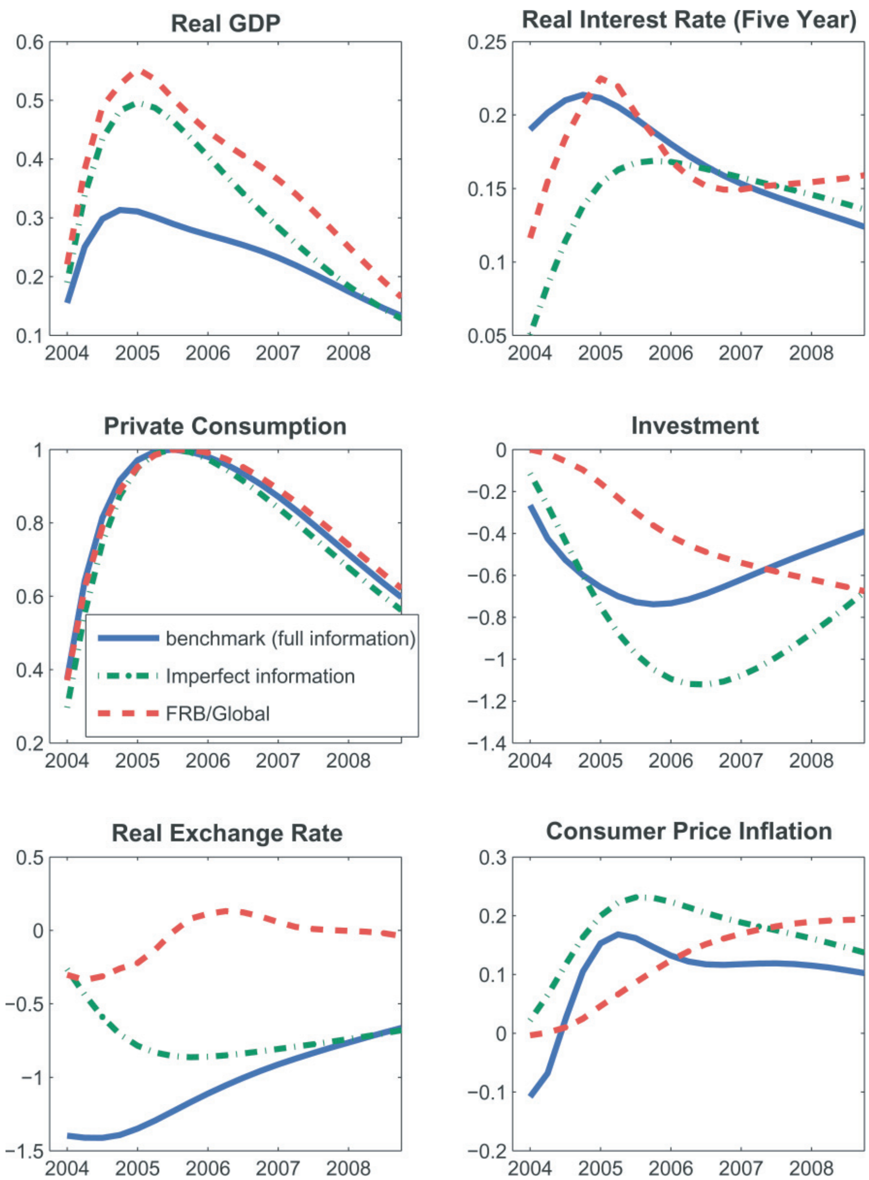
3.3 Rise in Home Consumption Demand

Figure 3 assesses the effects of a taste shock ν_{ct} to equation (22) that raises the marginal utility of consumption. This shock may be regarded as tantamount to the "autonomous shift in consumption demand" that is often considered in policy simulations. The shock is scaled so that it induces private consumption to rise by 1 percent above baseline at peak impact, and has a persistence of 0.975.

The taste shock exerts a highly persistent positive effect on real interest rates, accounting for the immediate rise in the five-year real interest rate shown in the figure. As a result, the stimulative effects of the rise in consumption demand on output are partly offset by a

[22]The temporary shock follows an AR(1) with a persistence parameter of 0.5. The innovation variances imply a Kalman gain parameter on the permanent component of 0.07. We use a similar calibration for the home consumption taste shock and foreign investment demand shocks considered below.

Figure 3. Rise in Home Consumption Demand



contraction in investment demand and by a reduction in real net exports. The latter occurs because higher real interest rates generate an appreciation of the real exchange rate. The exchange rate appreciation causes consumer price inflation to fall in the near term, although higher demand pressures eventually push up domestic goods prices by enough to cause consumer prices to rise.

These effects are qualitatively similar in the case of an autonomous shock to consumption demand in FRB/Global, i.e., a shock to the statistical residual in the consumption equation that is scaled to have the same effect on consumption. But from a quantitative perspective, the output effects are considerably larger in FRB/Global, because there is less crowding out of investment and a smaller decline in real net exports in that model (not shown). The smaller investment decline in FRB/Global reflects that investment is less sensitive to the long-term real interest rate (noting that long-term rates rise by a roughly commensurate amount in each model). The smaller decline in real net exports in FRB/Global reflects several factors, including less passthrough of the exchange rate to import prices, modestly lower trade price elasticities, and somewhat smaller real appreciation.

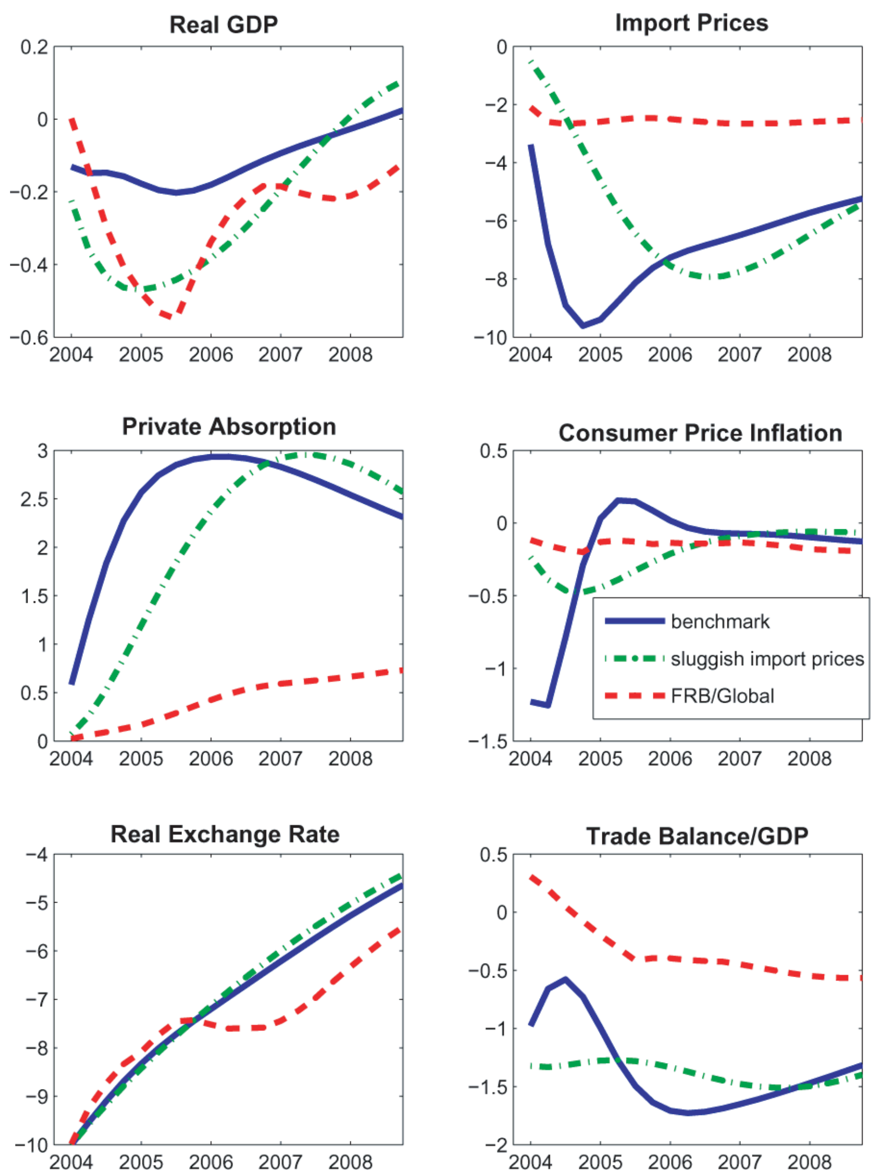
As in the case of the government spending rise, allowing for imperfect information about the persistence of the taste shock allows SIGMA to imply a larger and more persistent output response that is very similar to that in FRB/Global. The more gradual rise in long-term interest rates reduces the magnitude of real exchange rate appreciation, leading to a smaller export decline.

3.4 Fall in Home-Currency Risk Premium

Figure 4 shows the effects on the home country of a decline in the risk premium (ν_{bt}) on home-currency-denominated assets. As in McCallum and Nelson (1999) and in Kollman (2001), in this simulation we shock the exogenous component of the risk premium in the uncovered interest parity condition implied by the log-linearized model. The shock is scaled so that it induces an initial real appreciation of 10 percent, and the persistence of the shock is $\rho_b = 0.95$.

This shock reduces the required real return on all home-currency-denominated assets relative to the return on foreign assets. The lower required real return on home-currency assets occurs through a

Figure 4. Fall in Home-Currency Risk Premium



combination of persistently lower real interest rates, and through expected real currency depreciation. Thus, long-term real interest rates fall (not shown), and given that the shock has no long-run effect on the real exchange rate, the exchange rate is required to appreciate sharply in the impact period.

The appreciation of the real exchange rate depresses real exports and raises imports—and thus exerts a contractionary effect on real GDP. However, private domestic absorption is stimulated by lower real interest rates and lower import prices. As a result, real GDP shows only a modest contraction in the near term, and actually rises above baseline after a few years as higher investment spending leads to progressive capital deepening. The combination of stronger domestic demand and real exchange rate appreciation induces a significant deterioration of the nominal trade balance exceeding 1½ percentage points of GDP. Finally, PCE inflation shows a sizable but transient drop due to declining import prices.

The qualitative effects in FRB/Global are very similar, with the exchange rate appreciation leading to an initial output decline, lower real interest rates, some decline in inflation, and a trade balance deterioration. But while the response of GDP is fairly similar across models, there is considerable quantitative disparity in the responses of the expenditure components: private absorption rises much more in SIGMA than in FRB/Global, and real net exports correspondingly exhibit a larger contraction (not shown, though suggested by the larger trade balance response). A key factor accounting for these differences is that the passthrough of exchange rate changes to import prices is effectively 100 percent after a couple of quarters in SIGMA, whereas it is only about 30 percent in FRB/Global. The higher passthrough directly accounts for the larger effects on real imports and exports in the former model, and contributes significantly to driving the sharper swings in PCE price inflation and private absorption. As noted in the discussion of the government spending shock, the substantial response of private absorption in SIGMA also reflects a highly elastic response to the persistent decline in long-term real interest rates (not shown).

SIGMA's implication of very high exchange rate passthrough to import prices seems at odds with empirical evidence for the

United States that estimates long-run passthrough in the range of 20–30 percent.[23] Importantly, we emphasize that any model in which the desired markup is fixed (as in SIGMA) would appear unable to match the empirical pattern of import price response incorporated into FRB/Global, in which import prices react fairly quickly to exchange rate changes, but then adjust little subsequently. For example, the dash-dotted lines marked "sluggish import prices" use a calibration of SIGMA in which the mean duration of export prices is eight quarters (rather than two quarters as in the baseline). Even in this case of long-lived local currency pricing, exchange rate passthrough is nearly complete after two years, and there is much larger trade adjustment than in FRB/Global.

Given the central role of exchange rate passthrough in affecting the transmission of open economy shocks, we believe that it is of crucial importance to develop a theoretical framework that has the flexibility to account for the empirical features of passthrough evident in U.S. data. While some microfounded models can account for long-run passthrough that is below unity, including models with a distribution sector for retail goods as in Corsetti and Dedola (2003), and Corsetti, Dedola, and Leduc (2005), such models do not allow enough variation in desired markups to come close to matching the low level of passthrough in U.S. data.[24] Without much variation in the desired markup, models such as SIGMA imply implausibly large expenditure-switching effects in response to exchange rate movements that may limit their usefulness in addressing some important policy questions. Moreover, such restrictions may lead to large biases in the estimates of structural parameters.

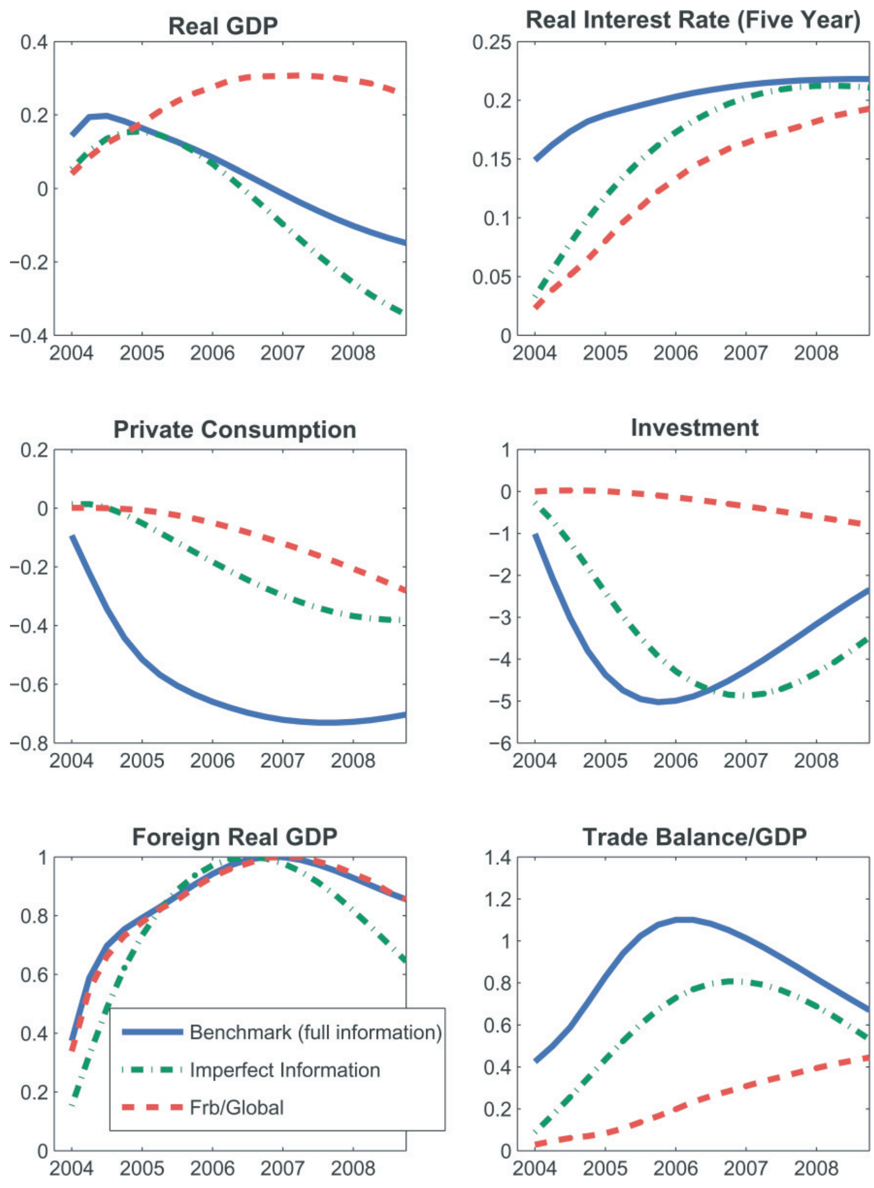
3.5 Alternative Foreign Demand Shocks

Figure 5 shows the effect on the home country of a rise in foreign investment demand. Specifically, investment in the foreign country

[23]The low long-run passthrough in FRB/Global is consistent with recent empirical estimates for the United States; see Marazzi, Sheets, and Vigfusson (2005).

[24]For example, Corsetti, Dedola, and Leduc (2005) find that their inclusion of a distribution sector can account for a reduction in long-run passthrough from unity to around 0.9. Other frameworks, such as Bergin and Feenstra (2001) and Gust and Sheets (2006) incorporate strategic complementarities in price setting à la Kimball (1995) that allow for greater variation in desired markups. These appear to be promising avenues to reduce long-run passthrough.

Figure 5. Rise in Foreign Investment Demand



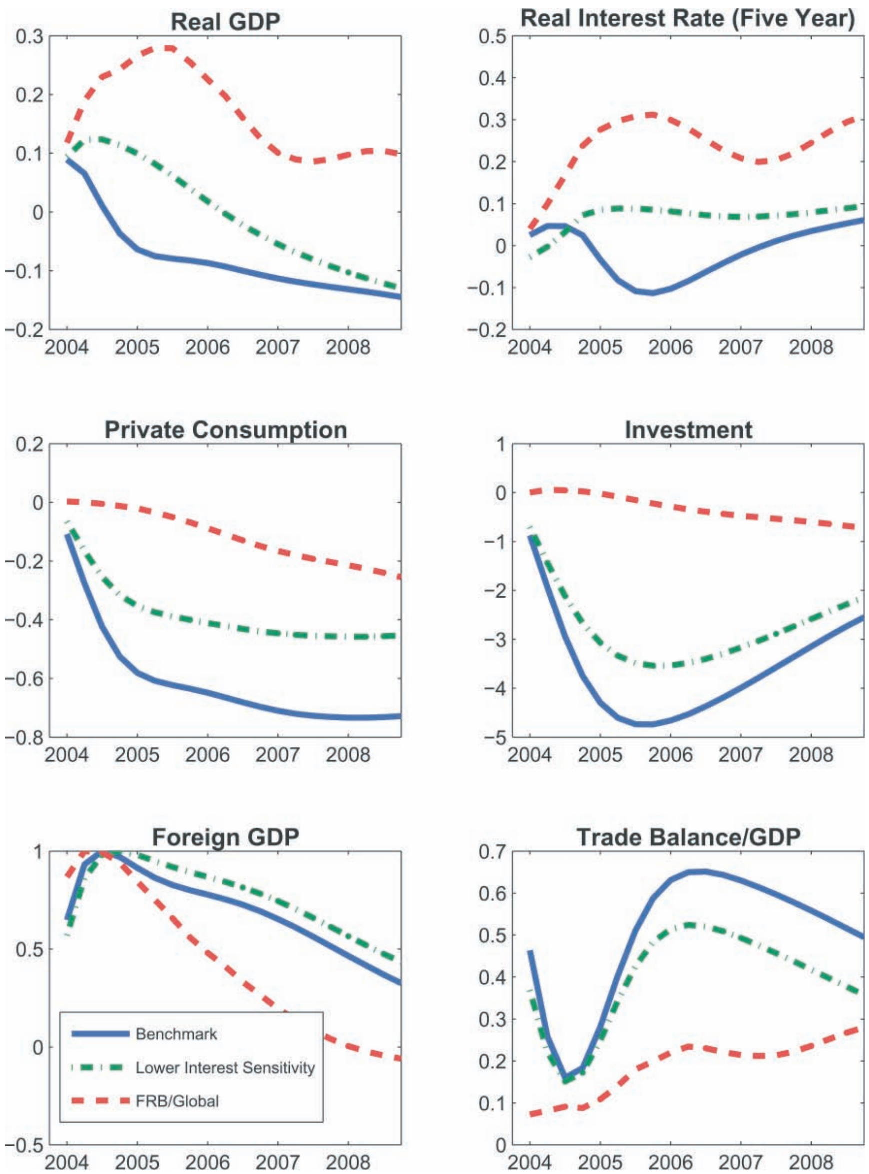
increases due to a highly persistent decline in its capital income tax rate; but it is useful to interpret the shock more broadly as reflecting changes in the investment climate abroad that boost the perceived return to capital. The shock is scaled so that foreign output eventually rises by 1 percent relative to baseline.

This foreign investment demand shock operates through similar channels both in SIGMA and in FRB/Global, and has qualitatively similar effects in each. Thus, the rise in foreign demand stimulates home real net exports, both directly because of the rise in foreign absorption, and indirectly through a depreciation of the home country's real exchange rate (not shown). This raises home output, though the stimulative effect on GDP arising from higher net exports is partly offset by declining private absorption (as domestic interest rates rise). The rise in real net exports generates an improvement in the nominal trade balance, while consumer prices rise due to higher import prices and stronger activity.

From a quantitative perspective, the "spillover effects" of the foreign demand increase on the home country are broadly similar for the first two years following the shock across the two models. This suggests that SDGE models may account for substantial spillover effects through trade channels in response to certain types of shocks, although we caution that the fact that the shock affects foreign investment spending—which is heavily import intensive—plays an important role in accounting for the relatively large effects. In addition, the spillover effects would decline if SIGMA incorporated lower exchange rate passthrough, since this would diminish the magnitude of the home country's export improvement. It is also evident that SIGMA implies much greater volatility in the expenditure components than FRB/Global. The trade balance exhibits a much larger improvement, and the components of private absorption fall much more sharply than in FRB/Global. As might be expected, the inclusion of imperfect information can markedly damp the volatility of the expenditure components in SIGMA by generating a smaller and more gradual response of long-term real interest rates, and by damping the impact on the real exchange rate.

Figure 6 illustrates that spillover effects on the home economy would be much smaller if the foreign output expansion were instead attributable to higher consumption spending. In particular, this alternative simulation assumes that foreign consumption demand rises

Figure 6. Rise in Foreign Consumption Demand



due to a taste shock, with the shock again scaled so that foreign output peaks at 1 percent above baseline. Given that foreign investment is crowded out by the shock, the shock has much smaller effects on the home country's real exports, and induces only a small and transient rise in output in our benchmark calibration of SIGMA. While it is interesting economically that SIGMA implies that the source of the foreign demand shock may be quite important in determining its effects on the domestic economy (whereas the domestic output increase in FRB/Global is broadly similar across the foreign demand shocks), it is plausible that certain features of SIGMA may unduly restrict the ability of the model to generate substantial spillover effects. These include both the high interest sensitivity of private absorption to the long-term real interest rate, and the omission of other potentially important financial channels. Recent work by Gilchrist, Hairault, and Kempf (2002) suggests that a financial accelerator may serve to enhance the ability of open economy SDGE models to account for larger spillover effects.

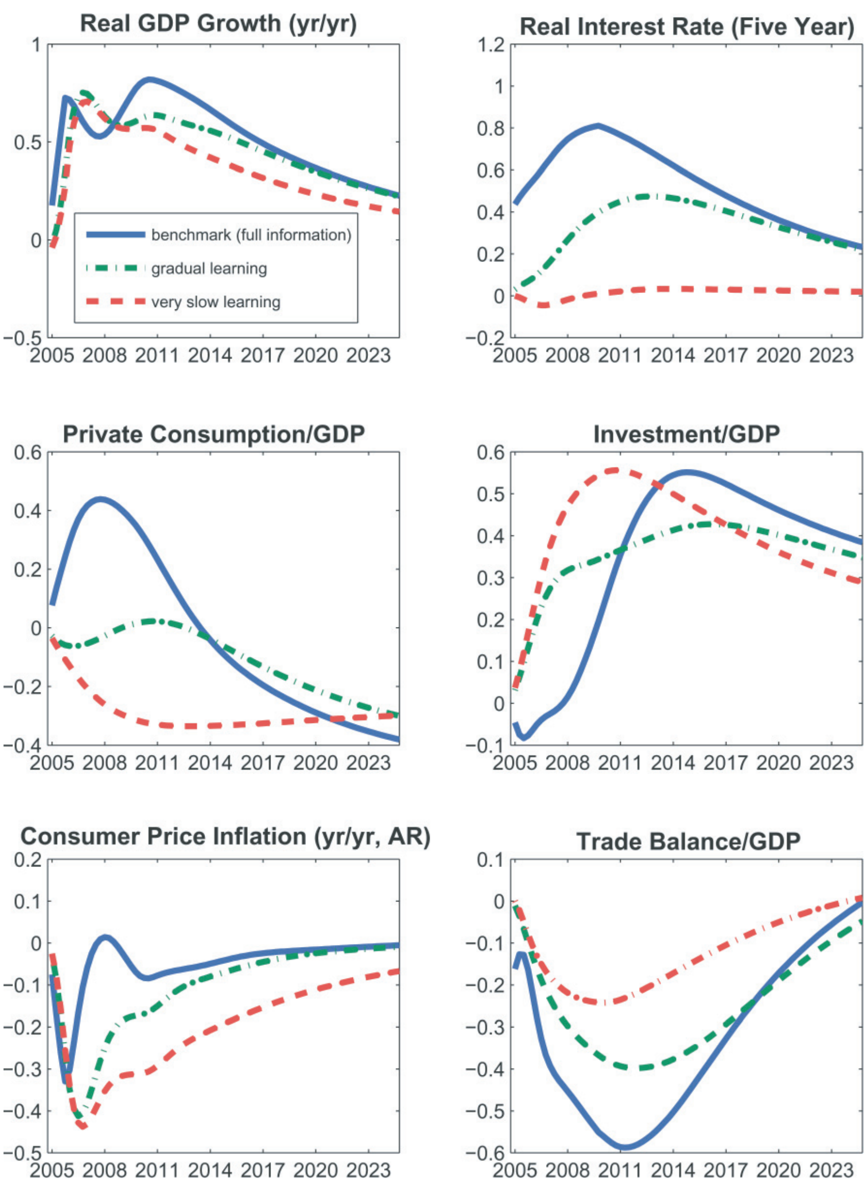
4. Long-Run Simulations in SIGMA

We now examine our model's implications for several supply-side shocks. Each shock exerts a highly persistent effect on output, in part because capital accumulation is very gradual. In addition to illustrating the model's long-run implications for the path of adjustment to each of these shocks, our analysis highlights the endogenous channels through which this adjustment occurs.

4.1 Persistent Rise in Productivity Growth

Figure 7 shows the responses of key variables to a productivity growth rate shock under alternative assumptions about the information structure. The shock raises technology growth by 1 percentage point per year over the the first five years of the simulation horizon, then decays slowly following an AR(1) process with an autocorrelation parameter of $\rho_p = .975$. The magnitude of the technology growth shock is similar to that experienced in the United States between 1996 and 2000; given the decay rate, it is consistent with agents projecting GDP growth five years ahead to rise immediately by about 3/4 percentage point above baseline.

Figure 7. Rise in Home Productivity Growth Rate



We begin by analyzing the effects of the shock under the assumption that agents have full information, and hence correctly ascertain that the shock will have highly persistent effects on the future growth rate of productivity (see the solid lines in the figure). Households project a much sharper rise in their future income than in the preshock baseline. This immediately stimulates consumption demand and depresses the saving rate. The increase in the expected marginal product of capital induces investment to rise (after a short delay). The expansion of domestic demand leads to higher real interest rates, putting upward pressure on the real exchange rate (not shown) and inducing a prolonged deterioration of the trade balance.

A hallmark feature of microfounded models such as SIGMA is that they completely articulate the longer-term economic forces that operate to correct any "imbalances," including in trade, its components, and the real exchange rate. The imposition of intertemporal budget constraints (and a debt-elastic risk premium) play a key role in generating these endogenous adjustments. In the case of the productivity acceleration, several aspects of the adjustment process account for the eventual movement of the trade balance into surplus. First, there is a long-run depreciation of the real exchange rate due to an increase in the supply of U.S. goods, which stimulates exports and reduces imports. Second, while the saving rate declines initially, it eventually increases as current income converges toward permanent income. Finally, after peaking about 10 years after the initial shock, the investment rate declines as capital approaches its new long-run level.

While these simulation results are instructive in understanding the forces that bring about long-run adjustment, the assumption that agents immediately recognize a productivity acceleration as permanent may be somewhat implausible. For example, in the U.S. experience of the late 1990s, forecasts of long-term output growth (i.e., five to ten years ahead) did not change noticeably until several years after the initial rise in productivity growth. This provides some empirical motivation for considering a "gradual learning" case in which agents do not immediately recognize that the productivity growth shock is highly persistent: given our calibration of the steady-state Kalman gain parameter, they initially believe the shock is mainly transitory.[25] In this case, agents project a much less pronounced rise in their income profile, so that consumption jumps much less than in the full information case; correspondingly, the figure shows that the consumption share of output remains nearly flat. Given reduced aggregate demand pressure relative to the full information case, real interest rates rise much less abruptly, which accounts for the somewhat larger rise in the investment share. While the trade deficit expands, the deterioration is also somewhat muted relative to the case of full information. This gradual adjustment of both domestic demand and external variables in response to the shock is more in line with the U.S. experience in the late 1990s, including with the relative constancy of the saving rate during that period.

4.2 Reduction in Labor Tax Rate

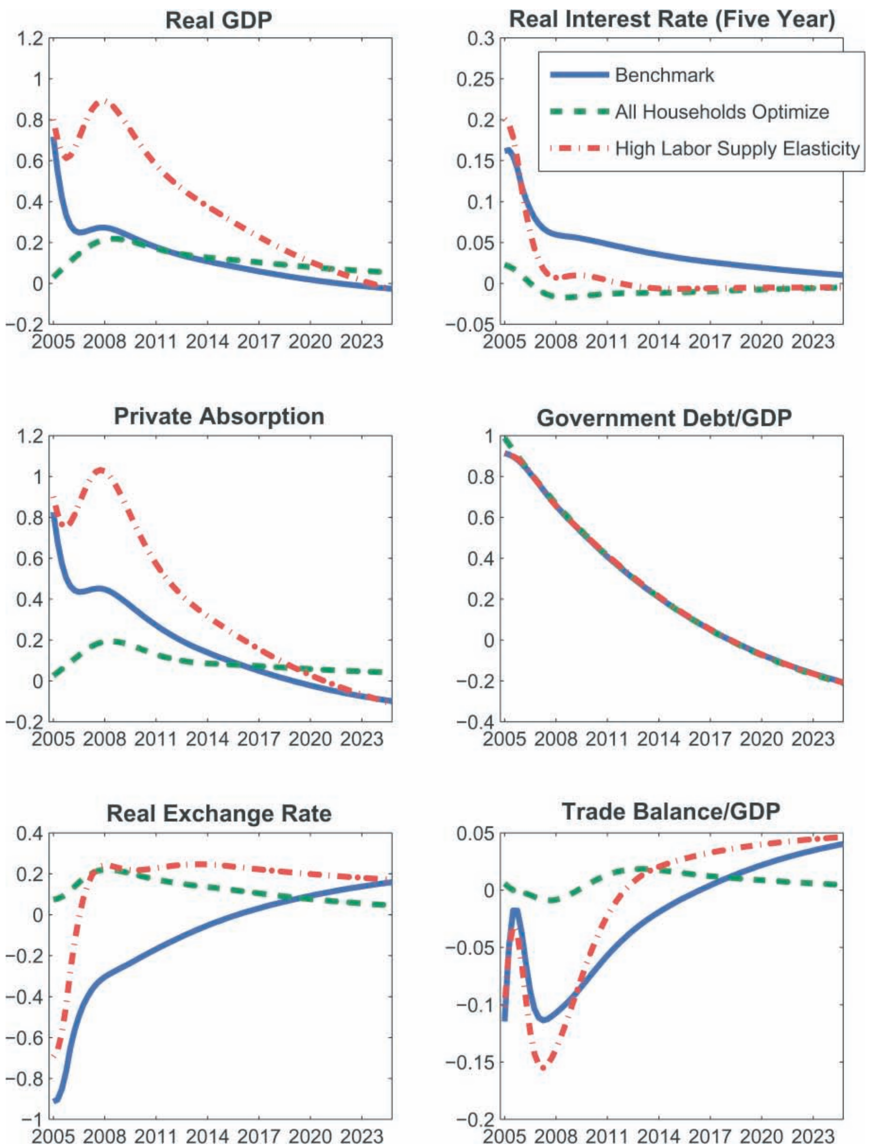
As another illustration of the forces that ensure long-run adjustment in the SIGMA model, figure 8 shows the effects of a cut in the labor tax rate. The tax cut is scaled so that government's labor tax revenue would fall by 1 percentage point of GDP if pretax labor income and output were unaffected. The shock is assumed to be highly persistent, with the autoregressive parameter $\rho_{\tau N}$ set to 0.975. This labor tax cut induces the fiscal deficit to rise initially by about 1 percent of GDP and to decay slowly thereafter.[26]

The cut in labor taxes induces a sharp initial rise in output. The shock exerts a strong stimulative effect on aggregate demand in the short run, as the HM households immediately expand their consumption in response to the increase in their after-tax income. The high level of persistence of aggregate consumption reflects that the consumption of the HM households remains high for an extended duration (given that the cut to labor taxes is very persistent and lump-sum taxes adjust slowly). Output declines from its initial peak as higher real interest rates crowd out investment spending and the consumption of optimizing households. However, output remains persistently above its preshock level even in the longer term because the shock exerts substantial supply-side effects: lower tax rates

[25] The temporary shock in this case is i.i.d., and the Kalman gain parameter on the permanent component is 0.10.

[26] We provide a more detailed discussion of the effects on the trade deficit in Erceg, Guerrieri, and Gust (2005).

Figure 8. Fall in Home Labor Tax Rate



induce households to work more by raising the cost of leisure, and the rise in labor supply in turn encourages capital accumulation.

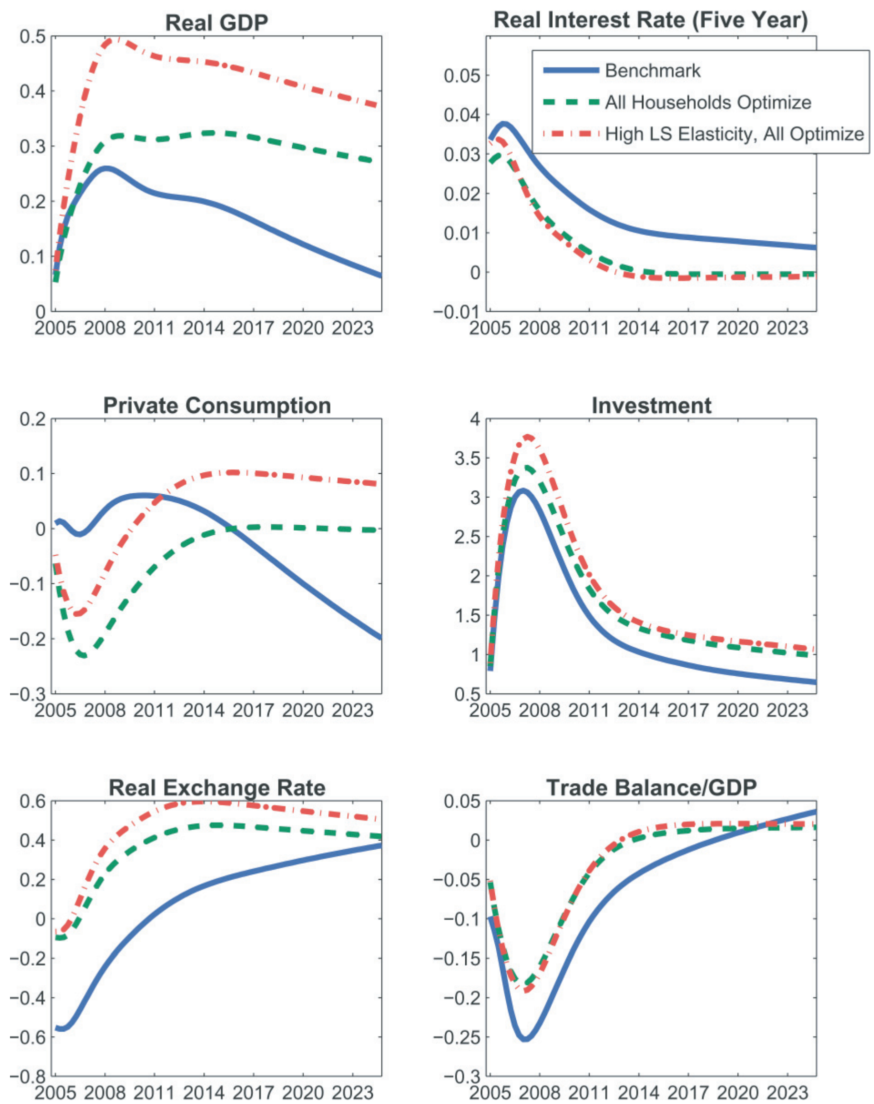
A clear advantage of our framework is that it is well suited to explore sensitivity to various structural characteristics determining consumption and labor supply behavior. In the context of assessing the effects of tax cuts, it is often of interest to policymakers to ascertain how results depend on the extent to which households are "Ricardian," or on their labor supply elasticity. We explore each of these alternatives in the figures. The dashed lines show the case in which all households are optimizers (rather than assuming 50 percent are optimizers as in the baseline). Because all households internalize the future tax increases necessary to satisfy the government intertemporal budget constraint, consumption shows a much smaller initial increase. In the longer term, the responses are similar to the benchmark, though there is somewhat greater capital deepening in this case. The dash-dotted line shows the case of a much higher Frisch labor supply elasticity (equal to unity, rather than 0.2 in our baseline). The larger labor supply response induces a much higher level of capital accumulation; accordingly, the longer-run output and consumption rise is greatly accentuated relative to the benchmark calibration.

The implications for the real exchange rate and trade balance under the alternative calibrations are shown in the lower panels. In the benchmark, the sharp rise in the real interest rate induces an initial appreciation in the real exchange rate, which drives the trade balance to deteriorate by around 0.1 percentage point of GDP. In the longer term, the supply-side effects dominate. Thus, the real exchange rate depreciates due to the higher supply of U.S. goods, while the trade balance shifts into surplus for the same reasons as in the case of the productivity acceleration. The trade balance exhibits a slightly larger deterioration in the calibration with a high labor supply elasticity (given the large stimulative effect on investment), while real exchange rates and trade show little reaction in the case in which all agents optimize.

4.3 Reduction in the Capital Tax Rate

Figure 9 shows the effects of a cut in the capital income tax rate that is scaled so that capital tax revenue would fall by 1 percentage point

Figure 9. Fall in Home Capital Tax Rate



of GDP if pretax labor income and output were unaffected. The shock is assumed to be highly persistent, with the autoregressive parameter set to 0.95. The capital tax cut induces the fiscal deficit to rise initially by about 1 percent of GDP and to decay slowly thereafter.

Given that the HM agents do not pay capital taxes, the capital tax rate reduction does not impart the same sort of short-run aggregate demand stimulus evident in the case of the labor tax cut. Accordingly, output increases slowly in line with the gradual rise in the capital stock. Higher real interest rates encourage somewhat greater saving by optimizing agents, even though the aggregate saving rate responds somewhat less due to the presence of the HM agents. The sharp rise in investment and high import content of investment goods encourages a substantial rise in imports, and this pressure on the trade balance is reinforced by a short-run real exchange rate appreciation: as a result, the trade balance experiences a peak deterioration of roughly 1/4 percentage point of GDP. In the long run, real exchange rate depreciation, a fall in the investment rate, and some rise in the saving rate (in part due to higher taxes on HM agents) induce the trade balance to move into persistent surplus.

Figure 9 also shows that output and investment would exhibit larger responses to a capital tax cut if all agents were optimizers (dotted lines). This reflects that the larger response of domestic saving in the latter case reduces pressure on real interest rates. The output response would be markedly accentuated if all agents were optimizers and had a much higher Frisch elasticity of labor supply of unity (five times higher than in the baseline case; see the dash-dotted lines), since the larger labor supply response would encourage higher capital accumulation.

5. Conclusion

The recent surge in interest in developing SDGE models for policy analysis seems warranted. The SDGE framework possesses some key advantages over that of existing large-scale econometric models by providing a clear linkage between structural features of the economy and its responses to shocks. Moreover, it offers a theoretically consistent framework for analyzing both short- and long-run responses that is helpful in assessing how the economy returns to its balanced growth path following disturbances.

While estimation remains an important future research objective, we have argued that an essential prerequisite involves identifying theoretical constraints of the particular SDGE framework that may preclude fitting the data on key dimensions. In this vein, even though it is encouraging that SIGMA implies responses to policy shocks that are generally similar to FRB/Global, there are at least two salient differences that seem attributable to restrictive aspects of our framework: namely, SIGMA implies much larger responses of both trade flows and the domestic expenditure components in response to shocks, and smaller and more transient spillover effects in response to foreign disturbances. Importantly, SIGMA's implications along these dimensions are likely to characterize a broad class of current SDGE models that imply fixed (or nearly fixed) desired markups, and that adopt a fairly standard framework for modeling investment and consumption behavior.

Given the importance of the magnitude of expenditure-switching effects for a wide range of open economy questions, we believe that it will be important in future work to develop mechanisms that can account for much lower long-term passthrough than implied by our model. Moreover, it will also be desirable to incorporate features that can potentially account for larger responses to foreign disturbances, at least under some conditions. It seems plausible that allowing for sectoral attachments of factors of production or the inclusion of a financial accelerator may allow SDGE models greater flexibility on this dimension.

Finally, while comparisons with FRB/Global are useful in evaluating the flexibility of SIGMA to fit responses similar to that of a data-oriented econometric model, we intend to adopt an estimation strategy that would allow significant departures from the responses of FRB/Global if the data provide a strong enough rationale. In particular, FRB/Global responses could be helpful in setting priors over certain estimated parameters in SIGMA in the context of a Bayesian approach, though sizable differences could emerge between the responses of the two models if the posterior distribution over the parameters diverged significantly from the prior.

References

Amato, Jeffery D., and Thomas Laubach. 2003. "Estimation and Control of an Optimization-Based Model with Sticky Prices and Wages." *Journal of Economic Dynamics and Control* 27 (7): 1181–1215.

Anderson, Gary. 1997. "A Reliable and Computationally Efficient Algorithm for Imposing the Saddle Point Property in Dynamic Models." Occasional Staff's Studies 4, Board of Governors of the Federal Reserve System.

Anderson, Gary, and George Moore. 1985. "A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models." *Economic Letters* 17 (3): 247–52.

Bergin, Paul R., and Robert C. Feenstra. 2001. "Pricing-to-Market, Staggered Contracts, and Real Exchange Rate Persistence." *Journal of International Economics* 54 (2): 333–59.

Betts, Caroline, and Michael B. Devereux. 1996. "The Exchange Rate in a Model of Pricing-to-Market." *European Economic Review* 40 (3–5): 1007–21.

Blanchard, Olivier J., and C. M. Kahn. 1980. "The Solution of Linear Difference Models under Rational Expectations." *Econometrica* 48 (5): 1305–12.

Blanchard, Olivier J., and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Quarterly Journal of Economics* 117 (4): 1329–68.

Brayton, Flint, Andrew Levin, Ralph Tryon, and John C. Williams. 1997. "The Evolution of Macro Models at the Federal Reserve Board." Finance and Economics Discussion Series, No. 1997-29, Board of Governors of the Federal Reserve System.

Brayton, Flint, and Peter Tinsley. 1996. "A Guide to FRB/US." Finance and Economics Discussion Series, No. 1996-42, Board of Governors of the Federal Reserve System.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3): 383–98.

Campa, José, and Linda Goldberg. 2004. "Exchange Rate Pass-Through into Import Prices." CEPR Discussion Paper No. 4391.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113 (1): 1–45.

Corsetti, Giancarlo, and Luca Dedola. 2003. "Macroeconomics of International Price Discrimination." CEPR Discussion Paper No. 3710.

Corsetti, Giancarlo, Luca Dedola, and Sylvain Leduc. 2005. "DSGE Models of High Exchange-Rate Volatility and Low Pass-Through." International Finance Discussion Paper No. 845, Board of Governors of the Federal Reserve System.

Devereux, Michael B., and Charles Engel. 2002. "Exchange Rate Pass-Through, Exchange Rate Volatility, and Exchange Rate Disconnect." *Journal of Monetary Economics* 49 (5): 913–40.

Elmendorf, Douglas W., and David L. Reifschneider. 2002. "Short-Run Effects of Fiscal Policy with Forward-Looking Financial Markets." *National Tax Journal* 55 (3): 357–87.

Erceg, Christopher, Luca Guerrieri, and Christopher Gust. 2005. "Expansionary Fiscal Shocks and the U.S. Trade Deficit." *International Finance* 8 (3): 363–97.

Erceg, Christopher J., Dale W. Henderson, and Andrew T. Levin. 2000. "Optimal Monetary Policy with Staggered Wage and Price Contracts." *Journal of Monetary Economics* 46 (2): 281–313.

Erceg, Christopher J., and Andrew T. Levin. 2003. "Imperfect Credibility and Inflation Persistence." *Journal of Monetary Economics* 50 (4): 915–44.

Fatás, Antonio, and Ilian Mihov. 2001. "The Effects of Fiscal Policy on Consumption and Employment." CEPR Discussion Paper No. 2760.

Galí, Jordi, David López-Salido, and Javier Vallés. 2004. "Rule-of-Thumb Consumers and the Design of Interest Rate Rules." *Journal of Money, Credit, and Banking* 36 (4): 739–63.

Gilchrist, Simon, Jean-Olivier Hairault, and Hubert Kempf. 2002. "Monetary Policy and the Financial Accelerator in a Monetary Union." International Finance Discussion Paper No. 750, Board of Governors of the Federal Reserve System.

Guerrieri, Luca. 2005. "Oil Shocks and the Global Economy." Manuscript, Board of Governors of the Federal Reserve System.

Gust, Christopher, and Nathan Sheets. 2006. "The Adjustment of Global External Imbalances: Does Partial Exchange Rate Pass-Through to Trade Prices Matter?" International Finance

Discussion Paper No. 850, Board of Governors of the Federal Reserve System.

Harrison, Richard, Kalin Nikolov, Meghan Quinn, Gareth Ramsay, Alasdair Scott, and Ryland Thomas. 2005. *The Bank of England Quarterly Model.* London: Bank of England Publications.

Kiley, Michael. 2001. "Business Investment in the Federal Reserve Board's U.S. Model (FRB/US): Specifications and Implications." Manuscript, Board of Governors of the Federal Reserve System.

Kimball, Miles S. 1995. "The Quantitative Analytics of the Neomonetarist Model." *Journal of Money, Credit, and Banking* 27 (4): 1241–77.

Kollman, Roberto. 2001. "The Exchange Rate in a Dynamic-Optimizing Business Cycle Model with Nominal Rigidities: A Quantitative Investigation." *Journal of International Economics* 55 (2): 243–62.

Laxton, Douglas, and Paolo Pesenti. 2003. "Monetary Policy Rules for Small, Open, Emerging Economies." *Journal of Monetary Economics* 50 (5): 1109–46.

Mankiw, N. Gregory. 2000. "The Savers-Spenders Theory of Fiscal Policy." NBER Working Paper No. 7571.

Marazzi, Mario, Nathan Sheets, and Robert Vigfusson. 2005. "Exchange Rate Pass-Through to U.S. Import Prices: Some New Evidence." International Finance Discussion Paper No. 833, Board of Governors of the Federal Reserve System.

McCallum, Bennett T., and Edward Nelson. 1999. "Nominal Income Targeting in an Open-Economy Optimizing Model." *Journal of Monetary Economics* 43 (3): 533–79.

McDaniel, Christine, and Edward Balistreri. 2003. "A Review of Armington Trade Substitution Elasticities." *Integration and Trade Journal* 7 (18): 161–73.

Obstfeld, Maurice, and Kenneth Rogoff. 1995. "Exchange Rate Dynamics Redux." *Journal of Political Economy* 103:624–60.

Orphanides, Athanasios, and Volker Wieland. 1998. "Price Stability and Monetary Policy Ineffectiveness when Nominal Interest Rates are Bounded at Zero." Finance and Economics Discussion Series, No. 98-35, Board of Governors of the Federal Reserve System.

Rotemberg, Julio J., and Michael Woodford. 1999. "Interest Rate Rules in an Estimated Sticky Price Model." In *Monetary Policy*

Rules, ed. John B. Taylor, 57–119. Chicago: University of Chicago Press.

Smets, Frank, and Raf Wouters. 2003. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association* 1 (5): 1124–75.

Turnovsky, Stephen J. 1985. "Domestic and Foreign Disturbances in an Optimizing Model of Exchange-Rate Determination." *Journal of International Money and Finance* 4 (1): 151–71.

Woodford, Michael. 2003. *Interest and Prices.* Princeton, NJ: Princeton University Press.

Yun, Tack. 1996. "Nominal Price Rigidity, Money Supply Endogeneity, and Business Cycles." *Journal of Monetary Economics* 37 (2): 345–70.

The Persistence of Inflation in OECD Countries: A Fractionally Integrated Approach*

María Dolores Gadea[a] and Laura Mayoral[b]
[a]Department of Applied Economics, University of Zaragoza
[b]Department of Economics and Business, Universitat Pompeu Fabra

The statistical properties of inflation and, in particular, its degree of persistence and stability over time is a subject of intense debate, and no consensus has been achieved yet. The goal of this paper is to analyze this controversy using a general approach, with the aim of providing a plausible explanation for the existing contradictory results. We consider the inflation rates of twenty-one OECD countries which are modeled as fractionally integrated (FI) processes. First, we show analytically that FI can appear in inflation rates after aggregating individual prices from firms that face different costs of adjusting their prices. Then, we provide robust empirical evidence supporting the FI hypothesis using both classical and Bayesian techniques. Next, we estimate impulse response functions and other scalar measures of persistence, achieving an accurate picture of this property and its variation across countries. It is shown that the application of some popular tools for measuring persistence, such as the sum of the AR coefficients, could lead to erroneous conclusions if fractional integration is present. Finally, we explore the existence of changes in inflation inertia using a novel approach. We conclude that the persistence of inflation is very high (although nonpermanent) in most postindustrial countries and that it has remained basically unchanged over the last four decades.

JEL Codes: C22, E31.

*We are grateful to Jordi Galí, Christian Haefke, and Eduardo Ley for their valuable comments. The authors also would like to thank two anonymous referees for helpful comments and the editor for encouragement. We acknowledge financial support from the Spanish Ministry of Education through grants SEC 2003/061006, SEC 2003-04429, and SEC 2003-04476 and also from the Barcelona Economics Program of CREA.

The study of the statistical properties of inflation has attracted a great deal of attention because this variable plays a central role in the design of monetary policy and has important implications for the behavior of private agents. Moreover, new interest in the subject has arisen in the last few years and, as a consequence, a large number of empirical and theoretical papers have appeared recently. Two reasons motivate this upsurge. Firstly, the international monetary context has experienced important changes such as the adoption of inflation-targeting regimes by some countries, the arrival of monetary union in Europe, and a general deflationist process in industrial economies. Secondly, the recent advances in the statistical treatment of time-series data have improved the tools of analysis.

In spite of the great effort, no consensus has been achieved yet about the most appropriate way to model the inflation rate, and various questions remain open. Two fundamental issues emerge in this macroeconomic debate: how to measure the persistence of inflation rates accurately and whether this persistence has changed recently. On the one hand, the degree of inflation persistence is a key element in the monetary transmission mechanism and a determinant of the success of monetary policy in maintaining a stable level of output and inflation simultaneously.[1] On the other hand, detecting whether persistence has fallen recently is crucial in determining the probability of recidivism by the monetary authority (see Sargent 1999) since, as Taylor (1998) and Hall (1999) have pointed out, tests in the spirit of Solow (1968) and Tobin (1968) will tend to reject the hypothesis of monetary neutrality if persistence estimates are revised downward. Thus, understanding the dynamics of inflation is a crucial issue with very important policy implications.

Various economic mechanisms have been put forward to characterize the price formation process, the sticky price models à la Taylor (1979, 1980) and Calvo (1983) being the dominant theoretical background in monetary policy. These models are not completely successful in capturing the observed inflation inertia, so

[1]The need to coordinate monetary policy with the degree of inflation persistence has given rise to numerous articles. For instance, Coenen (2003) and Angeloni, Coenen, and Smets (2003) study the robustness of monetary policy when there is uncertainty about the correct persistence of inflation and conclude that it would be preferable to design the monetary target assuming a high inflation inertia.

subsequent modifications have been designed to enhance their empirical performance (e.g., Fuhrer and Moore [1995], Fuhrer [1997], Galí and Gertler [1999], Christiano, Eichenbaum, and Evans [2001], Galí, Gertler, and López-Salido [2001], Roberts [2001], Driscoll and Holden [2004], Coenen and Wieland [2005], etc.). Nevertheless, from a more applied perspective, there is still a lot of controversy about the degree and stability of inflation persistence. On the one hand, there is abundant empirical evidence that postwar inflation exhibits high persistence in industrial countries. The papers of Pivetta and Reis (2004) for the United States and O'Reilly and Whelan (2004) in the euro zone are some examples. On the other hand, it has been argued that the above-mentioned results are very sensitive to the statistical techniques employed and that the observed persistence may be due to the existence of unaccounted structural changes, probably stemming from modifications in the inflation targets of monetary authorities, different exchange rate regimes, or shocks to key prices (see Levin and Piger 2003).[2] A similar lack of consensus is found in the analysis of persistence stability. Some authors have found evidence of a decrease in inflation inertia in recent years (see Taylor [2000], Cogley and Sargent [2001], and Kim, Nelson, and Piger [2004]) while others, employing different econometric techniques, give support to the opposite conclusion that inflation persistence is better described as unchanged over the last decades (see Batini [2002], Stock [2001], Levin and Piger [2003], O'Reilly and Whelan [2004], and Pivetta and Reis [2004]).

The goal of this article is to shed further light on this controversy by considering a wider statistical framework. Typically, the papers above only consider I(1) or I(0) processes (allowing sometimes for parameter instability) in order to fit these data. Although both formulations can deliver similar short-term predictions if appropriate parameters are chosen, their medium- and long-term implications are drastically different (see Diebold and Senhadji 1996). Processes containing a unit root are characterized by a flat sample autocorrelation function, revealing the fact that the impact of shocks to the series is permanent. In contrast, correlations in I(0) processes decay to zero at an exponential rate, implying that all shocks have a

[2]It is well known that the existence of changes of regime that are not explicitly taken into account may lead to the detection of spurious persistence (see Perron 1989).

short-lasting effect on the process. It is easy to find situations where this framework can be too restrictive, as there are both economic foundations and empirical evidence suggesting that many macroeconomic and financial variables react to shocks in a different fashion. This is the case, for instance, of variables whose shocks are non-permanent but vanish very slowly (with correlations, if they exist, decaying at a hyperbolic rather than at an exponential rate), resulting in series that may or may not be stationary, in spite of displaying mean reversion.[3] To overcome this limitation, a more flexible model has been introduced which is capable of encompassing the I(1)/I(0) paradigm as well as a richer class of persistence behaviors. The autoregressive fractionally integrated moving average (ARFIMA) models are similar to the ARIMA models, but the order of integration, d, is allowed to be any real number instead of only integer ones. It turns out that the former models are very convenient for analyzing the persistence properties of inflation since they are able to account for a wide variety of persistence features very parsimoniously.

In this paper, we demonstrate that fractionally integrated (FI) behavior can appear in the inflation rate as a result of aggregating prices from firms that are heterogeneous in their price adjustment costs, and we test this conjecture on a large data set containing twenty-one OECD countries.[4] In order to do so, FI models are estimated and tested against other popular specifications (such as different ARMA and ARIMA models, possibly affected by parameter instability) using both classical and Bayesian techniques.

We have found strong support for our conjecture, which is robust across the different countries, the various competing models, and the set of employed techniques. According to these results, it is shown that if ARIMA models are used to measure persistence, they will

[3]Evidence of these features has been found in variables such as GNP (Diebold and Rudebusch [1989] and Sowell [1992b]); asset price and exchange rate volatility (Andersen and Bollerslev [1997], Andersen et al. [1999], Ding, Granger, and Engle [1993], and Breidt, Crato, and Lima [1998]); political opinion data (Byers, Davidson, and Peel [1997]); and many others. See Henry and Zaffaroni (2002) for other significant references.

[4]FI models have already been employed in the literature to model inflation data, but, to the best of our knowledge, no economic justification for the presence of FI has been provided. See Baillie, Chung, and Tieslau (1992, 1996), Hassler and Wolters (1995), Franses and Ooms (1997), Barkoulas, Baum, and Oguz (1998), Bos, Franses, and Ooms (1999, 2002), Delgado and Robinson (1994), Baum, Barkoulas, and Caglayan (1999), and Ooms and Doornik (1999).

tend to overestimate this property. Furthermore, we show that the usual procedure of fitting an AR(k) process to the data and identifying a value of the sum of the AR coefficients close to one with the existence of an (integer) unit root can easily lead to persistence overestimation. This is so because any FI model with a fractional integration order strictly greater than zero admits an AR(∞) representation that verifies that the sum of the corresponding coefficients ($\rho(1)$) is equal to one.[5] When fitting an AR model to an FI process, any sensible information criterion chooses a finite and relatively small value of k, but the sum of the estimated coefficients is still close to one in most cases. Therefore, prudence recommends to interpret $\rho(1) \approx 1$ not as a signal of an integer unit root but just as an indication of some type of integration, possibly fractional, in the data. The implications in term of persistence of the former or the latter interpretation are drastically different.[6]

The main results that we have obtained can be summarized as follows. Once fractional integration is allowed for, both the I(0) and the I(1) specifications are clearly rejected. Furthermore, for most countries the FI specification is also preferred to the alternative of I(0) processes suffering from parameter instability, which could be an alternative explanation of the observed persistence.[7] Inflation rates are estimated using different techniques, and it is shown that they are best characterized as FI models with a memory parameter, d, around 0.6–0.8. This implies that they are very persistent, nonstationary; however, as opposed to I(1) variables, shocks have a nonpermanent character, so the series are mean reverting. We provide various persistence measures that permit an adequate comparison of inflation inertia across countries and their evolution over time. We

[5]This is true for the same reasons as in the I(1) case: the polynomial of the AR expansion contains the factor $(1-L)^d$, where L is the lag operator and d is a real number representing the order of integration. Clearly, $L = 1$ is a root of this polynomial if $d > 0$ which, in turn, implies that the sum of the AR coefficients associated with lagged values of the process has to be equal to one. See section 4 for a more technical explanation.

[6]As it will be shown in section 2, the class of FI models with an integration order, d, strictly greater than zero is very large, containing both stationary and nonstationary processes that, in the latter case, may or may not be mean reverting.

[7]It is well known that FI models and I(0) processes with structural changes may look very similar (see section 3). The possibility of directly testing these hypotheses is also a major novelty of this paper.

find important differences across countries. According to the half-life measure (HL), U.S. inflation is the most persistent and inflation of Central and Nordic European countries presents the lowest degree of inertia. We also provide persistence estimates computed from ARIMA specifications and show that the permanent-shock restriction introduced by the unit-root hypothesis leads to persistence overestimation. Finally, we have also explored the possibility of a change in persistence, but for most countries we find no evidence of any such change. Throughout the article, our results are compared with those of previous works, and explanations of the divergence are provided. We also describe some potential pitfalls deriving from the use of some popular persistence tools when the DGP is FI but this property is not taken into account.

The rest of the paper is structured as follows. Section 1 presents a standard preliminary analysis of inflation. Section 2 describes the concept and the main characteristics of fractionally integrated processes and provides an economic explanation of the existence of these features in inflation data. Section 3 reports the results of fitting ARFIMA models to this data set by using both classical and Bayesian methods and tests the FI(d) hypothesis against various alternatives such as I(1), I(0), and I(0) with a structural break in the mean. Impulse response functions and other scalar measures of persistence are provided in section 4. Section 5 analyzes the hypothesis of a change in inflation persistence. Finally, section 6 gives some concluding remarks.

1. Data Description and Preliminary Tests

We consider the quarterly consumer price index in the period running from the first quarter of 1957 to the last quarter of 2003 for twenty-one OECD countries. The data have been obtained from the International Financial Statistics database of the International Monetary Fund. The countries included in the study are Australia (AU), Austria (AUS), Belgium (BE), Canada (CA), Denmark (DK), Finland (FI), France (FR), Germany (GE), Greece (GR), Italy (IT), Japan (JP), Luxembourg (LX), Netherlands (NL), New Zealand (NZ), Norway (NO), Portugal (PO), Spain (SP), Sweden (SWE), Switzerland (SWI), United Kingdom (UK), and the United States (USA).

In order to construct the inflation rates, we have proceeded as follows. Firstly, the price series for each country has been seasonally adjusted using the X12 quarterly seasonal adjustment method of the U.S. Census Bureau. Secondly, inflation rates are computed as $\pi_t^i = \ln P_t^i - \ln P_{t-1}^i$ and, finally, an outlier analysis has been carried out and the additive outliers (AO) that clashed with methodological changes in the price indices have been removed. This has been the case of Austria (1957:3), Belgium (1967:1, 1971:1), Finland (1972:1), France (1980:1), Germany (1991:1), Greece (1959:1, 1970:1), Italy (1967:1), Netherlands (1960:1, 1961:1, 1981:1, 1984:2), New Zealand (1970:1), and Sweden (1980:1).

The evolution of the inflation series is shown in figures 5 to 7 (see the appendix). The well-known trends of postwar inflation in developed countries can be easily identified in these graphs. Starting from low levels in the 1960s, around 3 percent for most countries, prices rose dramatically in the 1970s after the oil crisis (inflation figures almost tripled) and this sharp increase was accompanied by high volatility. In the 1980s, inflation was moderately reduced by the application of tight monetary policies, but high levels of volatility were still observed. Finally, the 1990s were characterized by a generalized decrease in the mean and in the variance of inflation.

The preliminary analysis proceeds as follows. Firstly, standard unit-root tests have been computed on the inflation series and the results are presented in table 1. To be precise, the ADF test of Dickey and Fuller (1981), the PP of Phillips and Perron (1988), the MZ-GLS of Ng and Perron (2001), and the KPSS of Kwiatkowski et al. (1992) have been employed. Columns 1–3 of table 1 take the I(1) model as the null hypothesis, whereas the fourth column considers the I(0). The latter hypothesis is clearly rejected for all countries at the 1 percent significance level (column 4), whereas the I(1) is rejected for sixteen out of the twenty-one countries by at least two tests (columns 1–3). Four countries (IT, SP, PO, and USA) present rejection in one of the tests, and for only one country (BE) is it not possible to reject the I(1) conjecture with any of these tests. Since unit-root tests are known to lack power in many relevant situations, the results above cast serious doubts about the existence of a unit root in inflation rates. This finding is relevant because some tests (like the monetary neutrality tests) start by assuming a unit root in inflation rates and are not valid outside this framework.

Table 1. Unit-Root and Stationarity Tests

	ADF	PP	MZ_t-GLS	KPSS
AU	−2.39 (2)	−4.46** (8)	−2.16* (2)	0.88** (10)
AUS	−4.71** (2)	−5.60** (8)	−0.35 (2)	0.94** (9)
BE	−2.26 (3)	−2.77 (10)	−1.85 (3)	0.79** (10)
CA	−3.01* (1)	−3.89** (3)	−2.77** (1)	0.93** (10)
DK	−3.49** (2)	−4.94** (8)	−1.62 (2)	1.38** (9)
FI	−3.32* (1)	−4.11** (5)	−3.06** (1)	1.34** (10)
FR	−3.69** (1)	−3.49** (6)	−3.22** (1)	1.46** (10)
GE	−3.01* (2)	−4.75** (7)	−2.77** (2)	0.70** (10)
GR	−3.23* (1)	−3.71** (2)	−2.80** (1)	1.32** (10)
IT	−1.50 (5)	−3.56** (10)	−0.91 (5)	0.90** (10)
JP	−2.76 (2)	−4.60** (6)	−2.50* (2)	1.78** (10)
LX	−3.11* (7)	−4.38** (4)	−3.62** (7)	0.72** (10)
NL	−3.81** (3)	−5.20** (7)	−3.32** (3)	1.01** (9)
NZ	−4.14** (1)	−4.42** (6)	−3.42** (1)	1.00** (10)
NO	−3.42* (1)	−2.77 (1)	3.16** (1)	0.99** (10)
PO	−2.02 (4)	−3.74** (2)	−1.42 (4)	1.01** (10)
SP	−2.19 (4)	−5.08** (3)	−1.94 (4)	1.13** (10)
SWE	−3.00* (2)	−5.38** (7)	−2.28* (2)	1.04** (10)
SWI	−3.08* (2)	−5.09** (4)	−2.84** (2)	0.82** (10)
UK	−3.22* (1)	−3.26* (3)	−2.89** (1)	0.85** (10)
USA	−2.61 (3)	−2.63 (4)	−2.70** (3)	0.75** (10)

Notes: **, * denote significance at the 1 percent and 5 percent level, respectively. Figures in parentheses correspond to the number of lags and the bandwidth for the ADF and MZ_t-GLS and the PP and KPSS, respectively. Lag length was chosen according to the SBIC criterion. Bartlett's window was used as a kernel estimator in the PP and KPSS (bandwidth was chosen according to Newey and West 1994).

To sum up, since for most countries both the I(0) and the I(1) hypotheses are rejected, it seems that the ARIMA framework does not provide a good characterization of this data set. This result has been interpreted in the literature as an indicator of a behavior midway between the I(0) and the I(1) formulations.[8] If a process is I(1), all shocks have a permanent effect, whereas they disappear exponentially when the process is I(0). An alternative to both formulations that has been widely explored in the literature is the existence of structural breaks. This amounts to considering that only a few shocks, such as stock market crashes, oil crises, wars, etc., have a permanent effect on the series while all the others vanish rapidly. Perron (1989) showed that standard unit-root tests are not able to reject the I(1) hypothesis if a trend stationary process suffers from occasional breaks in the parameters that describe the trend and/or the level.

To explore the existence of breaks in the mean, we employ the method proposed by Bai and Perron (1998, 2003a, 2003b), henceforth BP, for multiple structural breaks. BP propose three types of tests. The $\sup F_T(\mathrm{k})$ test considers the null hypothesis of no breaks against the alternative of k breaks. The $\sup F_T(l+1/l)$ test takes the existence of l breaks, with $l = 0, 1, \ldots$, as H_0 against the alternative of $l + 1$ changes. Finally, the so-called "double maximum" tests, UDmax and WDmax, test the null of absence of structural breaks versus the existence of an unknown number of breaks. Bai and Perron (2003a) suggest beginning with the sequential test $\sup F_T(l+1/l)$. If no break is detected, they recommend checking this result with the UDmax and WDmax tests to see if at least one break exists. When this is the case, they recommend continuing with a sequential application of the $\sup F_T(l + 1/l)$ test, with $l = 1, \ldots$. This strategy has been followed to obtain the figures in table 2.

To test the changes in the level of the series, the following representation has been considered:

$$\pi_t^i = \varphi + \varsigma_t^i,$$

[8]It is well known that standard unit roots still have power when the DGP is not the one postulated under the alternative hypothesis. This is the case, for instance, of fractionally integrated processes (see Diebold and Rudebusch [1991] and Lee and Schmidt [1996] for the DF and KPSS tests, respectively) or some types of structural breaks (see Perron 1989).

Table 2. Breaks in the Mean

	Number of Breaks	Dates of the Breaks
AU	2	1970:4, 1991:1
AUS	3	1970:1, 1983:3, 1995:4
BE	2	1971:4, 1985:3
CA	4	1965:1, 1972:3, 1983:1, 1990:4
DK	3	1972:4, 1985:2, 1992:1
FI	3	1971:1, 1982:3, 1991:2
FR	3	1973:2, 1985:3, 1992:3
GE	2	1970:1, 1983:1
GR	2	1973:1, 1993:3
IT	3	1972:2 1983:3, 1995:3
JP	2	1981:3, 1993:4
LX	2	1970:1, 1985:3
NL	2	1963:4, 1985:4
NZ	2	1970:1, 1988:3
NO	2	1970:4, 1990:3
PO	4	1963:4, 1971:2, 1983:5, 1992:3
SP	4	1973:2, 1980:1, 1986:4, 1995:3
SWE	2	1970:1, 1992:1
SWI	1	1993:3
UK	3	1970:1, 1991:1, 1982:1
USA	2	1967:3, 1982:4

Note: The consistent covariance matrix is constructed using a quadratic kernel following Andrews (1991).

where φ is a constant capturing the level of the series and ς_t^i is a (short-memory) linear process. Following Perron (1989), attention is focused on sharp changes of the level, φ. A maximum number of five breaks has been considered, which, in accordance with the sample size $T = 186$, supposes a trimming $\varepsilon = 0.15$. The process ς_t^i is allowed to present autocorrelation and heteroskedasticity. A nonparametric correction has been employed to take account of these effects.

The results of applying the multiple-break tests to changes in the level of the inflation rates are presented in table 2. For most countries two or three breaks in the level are detected. The first break usually takes place at the beginning of the 1970s, whereas the second is located in the middle of the 1980s. The third, if it exists, occurs at the beginning of the 1990s. Thus, the chronology of the break points is in agreement with the general features of inflation discussed above.

The preliminary analysis of the inflation processes of OECD countries highlights the difficulties of modeling these series. On the one hand, there is evidence against both short-memory stationarity (I(0)) and unit-root behavior, which are the most common formulations employed to model these series. An alternative to both settings is to consider a model containing structural breaks in some parameters, and evidence supporting this hypothesis has been found. If the latter were true, it would mean that the persistence often found in these series is likely to be spurious. This is the conclusion put forward by Levin and Piger (2003). They analyze the inflation rates of twelve industrial countries and find evidence of breaks in the intercept of the inflation rate. They claim that conditional on these breaks, many countries do not show strong persistence.

Nevertheless, the existence of structural breaks is not the only alternative to the I(0)/I(1) framework. Fractionally integrated models can also bridge the gap between these two formulations. Moreover, it is well-known that FI and structural breaks can be easily confused. Since both types of models have very different implications in terms of persistence, it is crucial to determine which of the two phenomena is more likely to be present in the data. Sections 2 and 3 will deal with this issue.

2. Fractional Integration in Inflation Data

The previous results cast serious doubts on the adequacy of either the I(1) or the I(0) models to fit inflation series. When one is interested in analyzing the long-run impact of contemporaneous shocks, the above categories represent two extreme possibilities. Models containing a unit root are characterized by shocks that have a permanent effect, while innovations of I(0) processes disappear so fast that correlations decay at an exponential rate. Nevertheless, it has been shown that this framework could be too narrow in many instances, as there is ample empirical evidence suggesting that shocks of many macroeconomic and financial series behave differently. A class that embeds both the I(1) and the I(0) models and, at the same time, is able to account for richer persistence types is given by the so-called fractionally integrated (FI) models. Among this class, the most popular parametric model is the ARFIMA one, independently introduced by Granger and Joyeux (1980) and Hosking (1981). The main advantage of this formulation with respect to the ARIMA one is the introduction of a new parameter, d, that models the "memory" of the process, that is, the medium- and long-run impact of shocks on the process. More specifically, y_t is an ARFIMA(p, d, q) if it can be written as

$$\Phi(L)(1 - L)^d y_t = \Theta(L)\varepsilon_t, \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2),$$

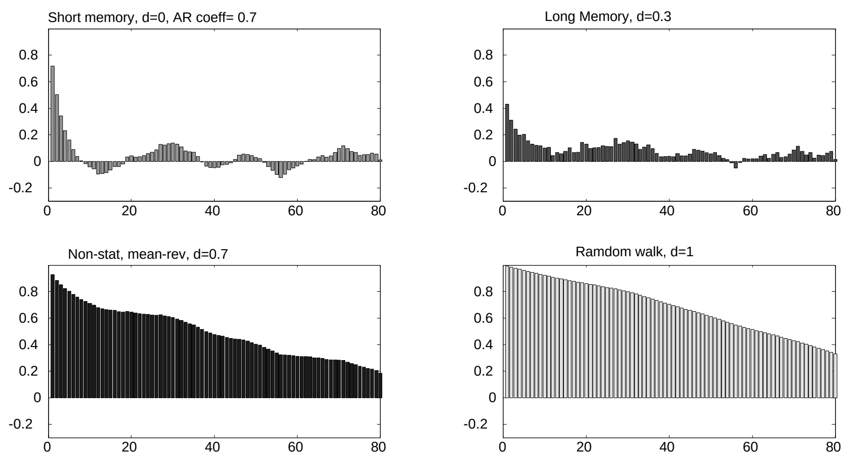
where the so-called *memory parameter*, d, determines the integration order of the series and is allowed to take values in the real, as opposed to the integer, set of numbers.[9] The terms $\Phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ and $\Theta(L) = 1 - \theta_1 L - \ldots - \theta_q L^q$ represent the autoregressive and moving average polynomials, respectively, with all their roots lying outside the unit circle. While d captures the medium- and long-run behavior of the process, $\Phi(L)$ and $\Theta(L)$ model the short-run dynamics. As Diebold and Rudebusch (1989) notice, this provides for "parsimonious and flexible modeling of low frequency variation."[10]

The bigger the value of d, the more persistent the process is. Stationarity and invertibility require $|d| < 1/2$, which can always be achieved by taking a suitable number of (integer) differences. *Short memory* is implied by a value of $d = 0$, where the process is characterized by absolutely summable correlations decaying at an exponential rate. By contrast, *long memory* occurs whenever d belongs to the (0,0.5) interval. Hosking (1981) showed that the correlation function in this case is proportional to k^{2d-1} as $k \to \infty$, that is, it decays at a hyperbolic rather than at an exponential rate. These processes are also characterized by an unbounded spectral density at frequency zero. These facts reflect the slower decay of shocks with

[9] ARIMA models are a particular case, where $d = 0, 1, 2, \ldots$. Notice that, in contrast to the ARIMA case, in the ARFIMA framework, d is a parameter that requires estimation.

[10] Furthermore, the fact of having two sets of parameters modeling the long- and short-run dynamics separately avoids some estimation problems that might affect the ARMA processes. As Sowell (1992b) points out, maximum likelihood estimation of ARMA models may sacrifice the long-run fit to obtain a better fit of the short-run behavior.

Figure 1. Sample Autocorrelation Function of Several Processes



respect to the I(0) case. A particularly interesting region for macroeconomic applications is the interval $d \in [0.5, 1)$. In this range, shocks are transitory, but the impulse response to shocks vanishes so slowly that the variance is not bounded and, therefore, the process is nonstationary in spite of being mean reverting (as shocks eventually disappear). Shocks have a permanent effect whenever $d \geq 1$.

Figure 1 illustrates the differences described above. The main diagonal contains the sample correlation function up to lag 80 of an I(0) and an I(1) process, respectively, whereas the other diagonal represents the same function for two FI processes. It can be seen that, after a few lags, the I(0) and the I(1) characterizations are drastically different, while the FI ones are able to fill the gap between the former models. The upper left graph depicts the sample autocorrelation function of an AR(1) process with an autoregressive coefficient equal to 0.7. Although this process is highly correlated at first lags, autocorrelations decay to zero very fast and become nonsignificant after a few lags. The behavior changes drastically whenever d is allowed to take strictly positive values. The long-memory case is illustrated in the upper right graph that contains the sample correlation function of an ARFIMA$(0, 0.3, 0)$. It is characterized by a slow decay of correlations, which remain significantly different from zero even at distant horizons. The two bottom graphs represent an

ARFIMA(0,0.7,0) and an I(1) process. Both are nonstationary, very persistent, but correlations for the former decay faster, revealing the fact that the process is eventually mean reverting. The graph on the lower right corresponds to a random walk where all shocks have a permanent effect.

The success of these models in economics may be attributed to the development of a rationale for the presence of FI in macro-level economic and financial systems. Robinson (1978) and Granger (1980) showed that FI behavior could appear in the aggregate produced from a large number of heterogeneous I(0) processes describing the microeconomic dynamics of each unit. This result has been incorporated in different economic settings to show analytically that some relevant variables can display FI[11] and is also the approach that we exploit to justify the existence of FI behavior in the inflation rate. Another way of obtaining FI behavior was proposed by Parke (1999). He considers the cumulation of a sequence of shocks that switch to zero after a random delay. If the probability that a shock survives for k periods, p_k, decreases with k at the rate $p_k = k^{2d-2}$ for $d \in (0, 1]$, Parke demonstrates that the error duration model generates a process with the same autocovariance structure as an I(d) process. He also shows how this mechanism can be applied to generate FI in aggregate employment and asset price volatility. From an empirical point of view, evidence supporting FI in financial and macroeconomic data is very large. See Henry and Zaffaroni (2002) for a detailed list of references.

Operationally, a binomial expansion of the operator $(1 - L)^d$ is used in order to fractionally differentiate a time series:

$$(1 - L)^d = \sum_{i=0}^{\infty} \pi_i(d) L^i, \tag{1}$$

where

$$\pi_i = \Gamma(i - d)/\Gamma(-d)\Gamma(i + 1) \tag{2}$$

and $\Gamma(\cdot)$ denotes the gamma function. When $d = 1$, (1) is just the usual first-differencing filter. For noninteger d, the operator $(1-L)^d$ is an infinite-order lag-operator polynomial with coefficients that decay

[11]Some examples are Michelacci and Zaffaroni (2000), Abadir and Talmain (2002), Haubrich and Lo (2001), Byers, Davidson, and Peel (1997), etc.

very slowly. Since the expansion is infinite, a truncation is needed in order to fractionally differentiate a series in practice (see Dolado, Gonzalo, and Mayoral [2002] for details on the consequences of the truncation).

2.1 The Sources of Fractional Integration in Inflation Data

Before testing for the presence of the above-described features in inflation series, it would be enlightening to have some plausible explanations for their existence in the data.

Why can inflation be fractionally integrated? One plausible mechanism for generating long-run dependence in inflation could stem from the fact that some economically important shocks have long memory. Evidence of this behavior in geophysical and meteorological variables is well documented (see, among others, Mandelbrot and Wallis 1969). Some authors have argued that the prices of some goods (in particular, raw materials) could inherit this property which, in turn, they transmit to other related goods (see Haubrich and Lo 2001). It seems difficult, however, to assess the extent of this effect in a price index and, therefore, we will not pursue this explanation here.

A more satisfactory explanation of the FI behavior, however, is provided by models that produce strong dependence despite white noise shocks. By applying the aggregation results on heterogeneous agents, it is easy to show that FI could appear in inflation data. Let us consider a model of sticky prices as in Rotemberg (1987), where it is assumed that each firm faces a quadratic cost of changing its price.[12] It is well known that when this is the case, the dynamics of prices are given by

$$p_t^i = \vartheta p_{t-1}^i + (1-\vartheta) p_t^{i*}, \tag{3}$$

where p and p^* represent the actual and optimal level of prices of firm i, and ϑ is a parameter that captures the extent to which imbalances are remedied in each period. Equation (3) can also be written as

$$\Delta p_t^i = \vartheta \Delta p_{t-1}^i + \nu_t^i, \tag{4}$$

[12]Quadratic costs of changing prices are equivalent, up to a first-order approximation, as far as aggregates are concerned, to a model such as Calvo (1983) where firms have a constant hazard of adjusting prices.

with $\nu_t^i = (1 - \vartheta)\Delta p_t^{i*}$. The parameter ϑ is a function of the adjustment costs and describes the speed of the adjustment, while $\vartheta/(1-\vartheta)$ is the expected time of adjustment. Since costs may differ across firms, it is natural to consider the case where ϑ may also depend on i. Then,

$$\Delta p_t^i = \vartheta^i \Delta p_{t-1}^i + \nu_t^i. \tag{5}$$

To build a price index, aggregation over a huge number of individual prices has to be considered (for instance, prices for the goods and services used to calculate the CPI are collected in eighty-seven urban areas throughout the United States and from about 23,000 retail and service establishments). Let us define the change in the price index Δp_t that verifies

$$\Delta p_t = \sum_{i=1}^{N} \Delta p_t^i.$$

Provided the distribution of ϑ^i verifies some (mild) semiparametric restrictions, Δp_t will display an FI behavior. Zaffaroni (2004) provides a full discussion of these restrictions. We will assume that ϑ belongs to a family $\Im$ of continuous distributions on [0,1) with density

$$\Im(\vartheta, d) \sim c\vartheta^{-d} \text{ as } \vartheta \to 0^+, \tag{6}$$

with $c \in (0, \infty)$. This is a very mild semiparametric specification of the cross-sectional distribution of ϑ. Zaffaroni (2004) shows that if ϑ is distributed according to (6), then the aggregated series will be FI(d). The bigger the proportion of agents having values of ϑ^i close to one, the higher the memory of the process. In other words, if an important proportion of agents correct the imbalances between the actual and the optimal level of prices only by a very small amount each period, the inertia in the inflation rate will be very high since the main factor driving the dynamics will be past values of prices.

It is interesting to notice that the behavior of $\Im(\vartheta, d)$ within any interval $[0, \gamma]$ is completely unspecified. Many parametric specifications verify the restriction in (6), for instance, the uniform and the Beta distributions. Zaffaroni's results imply that if the value of the memory parameter d is known (or can be estimated), then it is possible to infer a precise indication of the shape of the cross-sectional distribution of the $\vartheta^{i\prime}s$ near one. This implies that it is possible to infer on certain aspects of the microenvironment using aggregate information only.

3. Evidence of FI Behavior in Inflation Data

In this section we analyze the evidence of FI behavior in inflation data through a series of steps. Subsection 3.1 reports the results of applying several estimation techniques that explicitly allow for FI. In order to obtain more robust results, both classical and Bayesian methods are employed. For all countries and across the different techniques, fractional values of d, distant from both $\{0, 1\}$, are found. Next, we perform different tests of integer versus fractional integration, and the results are reported in subsection 3.2. Finally, the possibility of having detected spurious long memory as a consequence of the existence of an unknown number of structural changes in the data has been analyzed in subsection 3.3.

3.1 Estimation Results

In order to obtain robust estimates of the parameters of interest, we have considered several of the most popular estimation techniques, namely, the Geweke and Porter-Hudak (1983) (GPH) semiparametric method and three parametric ones: exact maximum likelihood (EML; see Sowell 1992a), nonlinear least squares (NLS; Beran 1994), and a minimum distance estimator (MD; Mayoral 2004a).[13] The estimated values of the memory parameter d are presented in table 3.

Several conclusions can be drawn from the inspection of this table. Firstly, the finding of fractional values of d, distant from the unit root, is robust across countries and across estimation methods. Most countries display values of d in the nonstationary ($d \geq 0.5$) but mean-reverting ($d < 1$) range, implying that, although very persistent, shocks are transitory. The semiparametric GPH method usually delivers slightly higher values of d than the other parametric techniques. This can be explained on the grounds that short-run

[13]NLS and EML have been computed with the ARFIMA package 1.0 for OX (Doornik and Ooms 2001), while MD has been implemented in MATLAB. Parametric models have been chosen according to the AIC information criteria.

Table 3. Estimation of FI(d) Models

	GPH	NLS	EML	MD
AU	0.78 (0.20)	0.79 (0.10)	0.69 (0.06)	0.74 (0.06)
AUS	0.78 (0.19)	0.69 (0.13)	0.80 (0.10)	0.73 (0.10)
BE	0.83 (0.21)	0.58 (0.10)	0.56 (0.06)	0.611 (0.08)
CA	0.76 (0.17)	0.69 (0.10)	0.73 (0.07)	0.69 (0.09)
DK	0.66 (0.16)	0.67 (0.11)	0.63 (0.07)	0.66 (0.07)
FI	0.74 (0.14)	0.59 (0.08)	0.60 (0.15)	0.62 (0.10)
FR	0.75 (0.21)	0.89 (0.21)	0.65 (0.06)	0.72 (0.08)
GE	0.94 (0.27)	0.58 (0.27)	0.61 (0.09)	0.68 (0.09)
GR	0.64 (0.30)	0.66 (0.10)	0.62 (0.05)	0.60 (0.06)
IT	1.19 (0.27)	0.72 (0.42)	0.66 (0.05)	0.69 (0.08)
JP	0.62 (0.09)	0.59 (0.16)	0.75 (0.10)	0.63 (0.10)
LX	0.74 (0.29)	0.69 (0.18)	0.68 (0.11)	0.65 (0.13)
NL	0.86 (0.20)	0.67 (0.14)	0.72 (0.12)	0.70 (0.11)
NZ	0.52 (0.41)	0.62 (0.14)	0.57 (0.08)	0.63 (0.10)
NO	0.64 (0.26)	0.66 (0.13)	0.55 (0.26)	0.64 (0.15)
PO	0.80 (0.22)	0.63 (0.10)	0.63 (0.07)	0.59 (0.10)
SP	0.90 (0.16)	0.61 (0.15)	0.60 (0.07)	0.65 (0.11)
SWE	0.58 (0.16)	0.59 (0.14)	0.52 (0.09)	0.59 (0.10)
SWI	0.56 (0.18)	0.62 (0.11)	0.59 (0.12)	0.61 (0.11)
UK	0.78 (0.20)	0.69 (0.22)	0.64 (0.10)	0.62 (0.10)
USA	0.66 (0.14)	0.68 (0.32)	0.72 (0.20)	0.69 (0.16)

Note: Standard deviation is shown in parentheses.

correlation may bias the estimator upward (see Agiakloglou, Newbold, and Wohar 1992). The parametric methods present very similar values, and for most countries estimated values of d around 0.6–0.7 are found.

A problem often associated with parametric estimators of d is that they are very sensitive to the selection of the specific parametric model, so estimated values can vary greatly across different specifications. To overcome this problem, we have also computed some Bayesian estimates of d in order to take the model uncertainty into account. We follow Koop et al. (1997) and consider the sixteen possible combinations of ARFIMA models with $p, q \leq 3$. A uniform density for d in the interval $[0, 1.5]$ has been assumed. So, the method puts $2/3$ of the prior mass on values of d implying nonpermanent shocks ($d < 1$) and $1/3$ on values that correspond to permanent shocks ($d > 1$).

The outcome of the Bayesian estimation is reported in table 4. The mean and the standard deviation of d is provided for both the "best model" (the one with the highest posterior probability) and the "overall model," which weights the sixteen ARFIMA models according to their posterior probabilities.[14] Since the method computes the density function of d for each model, the probability that inflation is mean reverting ($P(d_i < 1)$) can be easily obtained and is also displayed in this table.

The results reported in table 4 suggest that there is a high variability associated with the estimation of d. In general, the Bayesian approach offers higher values of the memory parameter than the classical methods, although in almost all cases the estimated values remain below one. Moreover, the posterior probability of nonpermanent shocks ($d < 1$) is bigger than $2/3$ (the a priori probability) for eighteen out of the twenty-one countries considered.

Summing up, the Bayesian analysis, in accordance with the classical approach, confirms the very persistent but mean-reverting behavior of inflation data.

[14] See Koop et al. (1997) for details on the estimation procedure. Computations have been carried out using the Fortram code provided by them.

Table 4. Bayesian Estimation of ARFIMA Models

	BEST ARFIMA		OVERALL ARFIMAS	
	Mean(d)	$P(d<1/\text{data})$	Mean(d)	$P(d<1/\text{data})$
AU	0.88 (0.19)	0.75	0.82 (0.20)	0.82
AUS	0.34 (0.06)	1	0.34 (0.06)	1
BE	0.86 (0.14)	0.90	0.87 (0.15)	0.76
CA	0.99 (0.26)	0.55	0.85 (0.21)	0.74
DK	0.85 (0.21)	0.71	0.87 (0.23)	0.63
FI	0.62 (0.06)	1	0.67 (0.15)	0.95
FR	0.66 (0.07)	1	0.68 (0.14)	0.93
GE	0.78 (0.33)	0.86	0.83 (0.26)	0.76
GR	0.64 (0.06)	1	0.78 (0.17)	0.82
IT	0.73 (0.18)	0.92	0.66 (0.13)	0.96
JP	0.64 (0.10)	0.99	0.62 (0.21)	0.91
LX	0.98 (0.31)	0.65	0.83 (0.22)	0.78
NL	0.91 (0.28)	0.54	0.79 (0.25)	0.76
NZ	0.91 (0.31)	0.60	0.85 (0.22)	0.66
NO	0.57 (0.06)	1	0.71 (0.19)	0.86
PO	1.33 (0.12)	0.03	1.14 (0.18)	0.25
SP	1.30 (0.30)	0.30	1.07 (0.31)	0.52
SWE	0.42 (0.05)	1	0.80 (0.24)	0.74
SWI	0.60 (0.06)	1	0.65 (0.17)	0.94
UK	0.60 (0.06)	1	0.80 (0.15)	0.75
USA	0.58 (0.19)	0.97	0.64 (0.22)	0.86

Note: Standard deviation in shown in parentheses.

3.2 Testing Fractional versus Integer Integration

Tables 3 and 4 support our initial hypothesis of the fractionally integrated behavior of inflation data and that the order of integration is, in general, far from both zero and one. But one could argue that this could be the case even if the series has an integer degree of integration since it would be very unlikely to obtain an exact integer value for d. In this section, we will formally test these hypotheses.

Several authors have found evidence in favor of the existence of a unit root in inflation (see, for instance, Pivetta and Reis 2004). Other authors, such as Cogley and Sargent (2001), postulate an I(0) representation for inflation on the basis that nonstationary ones are not plausible since they would imply an infinite asymptotic variance of inflation. They argue that this could never be optimal if the central bank's loss function includes the aforementioned variance. We will show below that when the possibility of fractional integration is considered, both the I(0) and the I(1) representations are rejected in our data set.

The simplest test is to build confidence intervals around the estimated values of d reported in table 3. Although simple, this approach has an important drawback: usually intervals are too wide and most hypotheses cannot be rejected (see Sowell 1992a). Fortunately, other simple and more powerful methods are available in the literature. To test the unit root versus the FI hypothesis, the Fractional Dickey-Fuller (FDF) test (see Dolado, Gonzalo, and Mayoral 2002, 2003) has been employed. This test generalizes the traditional Dickey-Fuller test of I(1) against I(0) to the more general framework of I(1) versus FI(d). It is based upon the t-ratio associated with the coefficient of $(1-L)^d y_{t-1}$ in a regression of $(1-L)y_t$ on $(1-L)^d y_{t-1}$ and, possibly, some lags of $(1-L)y_t$ to account for the short-run autocorrelation of the process and/or some deterministic components if the series displays a trending behavior or initial conditions different from zero.[15] Table 5 presents the results of applying the FDF test to

[15] The FDF invariant regression that has been run is equal to $\Delta y_t = \alpha_1 \tau_{t-1}(d) + \phi \Delta^d y_{t-1} + \sum_{j=1}^{k} \psi_j \Delta y_{t-j} + a_t$, and a number of lags of Δy_t equal to two was chosen according to the BIC criterion. The coefficient α_1 is associated to the deterministic components (a constant; see Dolado, Gonzalo, and Mayoral 2003). The term $\tau_t(d)$ is defined as $\tau_t(d) = \sum_{i=0}^{t-1} \pi_i(d)$, where the coefficients $\pi_i(\delta)$ come from the expansion of $(1-L)^\delta$ as defined in equation (2).

Table 5. FDF Test (I(1) versus FI(d)). $H_0: d_0 = 1$; $H_1: d = d_1$

H_1 :	$d_1 = 0.6$	$d_1 = 0.7$	$d_1 = 0.8$	$d_1 = 0.9$
AU	-8.76^{**}	-4.65^{**}	-4.68^{**}	-4.69^{**}
AUS	-8.56^{**}	-8.54^{**}	-8.47^{**}	-8.36^{**}
BE	-7.39^{**}	-7.53^{**}	-7.62^{**}	-7.69^{**}
CA	-5.92^{**}	-5.66^{**}	-3.73^{**}	-3.70^{**}
DK	-6.14^{**}	-6.05^{**}	-5.94^{**}	-5.81^{**}
FI	-5.45^{**}	-5.19^{**}	-4.90^{**}	-3.20^{**}
FR	-4.34^{**}	-4.12^{**}	-3.27^{**}	-3.26^{**}
GE	-6.77^{**}	-6.79^{**}	-6.77^{**}	-6.72^{**}
GR	-5.79^{**}	-5.62^{**}	-5.43^{**}	-5.24^{**}
IT	-4.82^{**}	-2.87^{**}	0.01	0.17
JP	-8.73^{**}	-4.52^{**}	-4.51^{**}	-4.50^{**}
LX	-7.32^{**}	-4.55^{**}	-4.60^{**}	-4.65^{**}
NL	-6.86^{**}	-6.68^{**}	-6.49^{**}	-5.89^{**}
NZ	-9.31^{**}	-4.70^{**}	-4.56^{**}	-4.41^{**}
NO	-6.77^{**}	-6.50^{**}	-6.22^{**}	-3.12^{**}
PO	-8.04^{**}	-4.40^{**}	-4.31^{**}	-4.20^{**}
SP	-7.88^{**}	-7.65^{**}	-3.80^{**}	-3.89^{**}
SWE	-6.07^{**}	-6.03^{**}	-5.79^{**}	-5.78^{**}
SWI	-5.86^{**}	-5.58^{**}	-3.73^{**}	-3.68^{**}
UK	-6.07^{**}	-5.84^{**}	-5.58^{**}	-5.32^{**}
USA	-2.27^{*}	-2.18^{*}	-2.11^{*}	-2.04^{*}

Note: *,** denote rejection at the 5 percent and the 1 percent level, respectively. Critical values: $N(0, 1)$.

this data set. Several alternative hypotheses have been considered (d = 0.6, 0.7, 0.8, and 0.9). The conclusion of this table is clear: the unit-root model is clearly rejected (usually at the 1 percent significance level) against fractionally integrated alternatives in all countries.

Next, we test for FI versus short memory (I(0)). To this end, a point-optimal test recently proposed by Mayoral (2004b) has been implemented and the results are presented in table 6. The test works as follows. Given the characteristics of the inflation data, the following DGP has been considered:

$$\begin{aligned} y_t &= \mu + x_t \\ \Delta^{d_i} x_t &= u_t, i = \{0, 1\}, \end{aligned}$$

where μ is a constant, u_t is a linear $I(0)$ process, and $d = d_0$ and $d_1 = 0$ are, respectively, the integration orders under H_0 and H_1. Under the Neyman-Pearson lemma, the most powerful test will reject the null hypothesis of $d = d_0$ for small values of $L(d, \sigma)|_{H_1} - L(d, \sigma)|_{H_0}$, where L is the log-likelihood function. After some manipulation, the critical region of the most powerful test for these hypotheses is given by

$$\frac{\sum(y_t - \mu)^2}{\sum(\Delta^{d_0}(y_t - \mu))^2} < k_T. \tag{7}$$

The asymptotic distribution of this statistic (scaled by T^{1-2d}) is not standard, and critical values can be found in Mayoral (2004b) for the case where u_t is i.i.d. When u_t is a general linear short-memory process, a nonparametric correction should be introduced using any of the standard techniques available in the literature (see Mayoral 2004b).

To interpret the figures reported in table 6, it is important to notice that the test is consistent (rejects the null hypothesis of FI(d_0) for large T) if the true integration order, d^*, is smaller than the integration order used as the null hypothesis, d_0. Consequently, whenever $d_0 > d^*$, the test will reject the FI(d_0) hypothesis. For example, if the true integration order is $d^* = 0.7$ but $d_0 = 0.9$ is taken as H_0, the test will tend to reject the hypothesis of $d_0 = 0.9$.

The results in table 6 are very homogeneous across countries. For moderate values of d, around 0.6–0.7 and even 0.8 for most countries, the null hypothesis of FI cannot be rejected. Nevertheless, for higher values of d_0 ($d_0 = 0.9$), the same null is rejected. This result confirms the outcome of the estimation methods in table 3 since, according to this table, the true integration orders are around 0.7. Therefore, taking into account the properties of the test, when higher $d_0's$ are

Table 6. Test of FI(d) versus I(0)

	R^c Test (Mayoral 2004)			
H_0 :	$d = 0.6$	$d = 0.7$	$d = 0.8$	$d = 0.9$
AU	1.136	0.456	0.175	0.064*
AUS	0.592	0.257	0.071*	0.024*
BE	0.550	0.196*	0.069*	0.024*
CA	1.315	0.547	0.217	0.083*
DK	0.899	0.339	0.124	0.044*
FI	1.054	0.438	0.174	0.067*
FR	0.939	0.397	0.162	0.064*
GE	0.839	0.327	0.123	0.044*
GR	0.737	0.273	0.098*	0.035*
IT	1.434	0.614	0.251	0.099
JP	1.013	0.408	0.158	0.059*
LX	1.125	0.466	0.184	0.070*
NL	0.513	0.282	0.063*	0.022*
NZ	0.817	0.314	0.117*	0.042*
NO	1.006	0.400	0.154	0.057*
PO	1.218	0.483	0.184	0.068*
SP	1.079	0.448	0.178	0.068*
SWE	1.019	0.405	0.155	0.058*
SWI	0.840	0.347	0.138	0.053*
UK	1.014	0.412	0.161	0.061*
USA	1.225	0.535	0.225	0.091*
Crit. Values (5% *S.L.*)	0.502	0.241	0.122	0.092

employed, the test should reject $H_0 : d = d_0$, as it actually does. Thus, the test supports the hypothesis of FI behavior with a degree of integration close to 0.7.

3.3 Testing Fractional Integration versus Structural Breaks

It is well known that it is very difficult to provide an unambiguous answer as to whether a process is fractionally integrated or is short memory plus some deterministic components perturbed by sudden changes. Several authors have pointed out that many standard techniques for detecting persistence can spuriously find this property in short-memory processes when there is parameter instability (e.g., Bhattacharya, Gupta, and Waymire [1983], Künsh [1986], Perron [1989], Teverosky and Taqqu [1997], Giraitis, Kokoszka, and Leipus [2001], Mikosch and Starica [2004], Perron and Zhu [2005], and many others). Other authors have studied the opposite effect, that is, how conventional procedures for detecting and dating structural changes tend to find spurious breaks, usually in the middle of the sample, when in fact there is only fractional integration (see Nunes, Kuan, and Newbold [1995], Krämer and Sibbertsen [2002], and Hsu [2000]). Therefore, although there is a general consensus on the fact that most economic series are nonstationary, it is often difficult to be sure about the source of the nonstationarity, that is, whether it comes from a high degree of persistence or from the existence of parameter changes.

In view of these results, it is not surprising that evidence supporting both the existence of breaks in the mean (section 1) and strong persistence (subsections 3.1 and 3.2) is found for the same data set. For the purposes of this article, distinguishing between these two models is crucial since they have very different implications in terms of the degree of persistence. Thus, we now explore the possibility that the existence of different regimes in the mean in an otherwise short-memory process could be generating spurious memory in the inflation rate. To do so, an extension of the test described in subsection 3.2 has been employed. The aim of the test is to determine if the persistence observed in the data is real or is an artifact of other phenomena such as the existence of breaks. More specifically, the hypotheses of FI(d_0) versus I(0) with a break in the level are considered. The test works as follows: let T_B be the time when the break occurs and $\omega = T_B/T$ the parameter that describes the location of the break point in the sample. To allow for breaks in the level, the dummy variable $DC_t(\omega) = 1$ if $t > T_B$ and 0 otherwise is defined. Since the date where the break occurs is unknown, the test has a critical region given by

$$\min_{\omega} \frac{\min_{\alpha_1,\alpha_2} \sum (y_t - \alpha_1 - (\alpha_2 - \alpha_1) DC_t(\omega))^2}{\min_{\alpha_0} \sum (\Delta^{d_0}(y_t - \alpha_0))^2} \leq k_T, \qquad (8)$$

where the minimization is carried out in $\omega \in \Omega$, where, following Andrews (1993), $\Omega = [0.15, 0.85]$. The distribution of the statistic in

Table 7. Tests of FI(d) versus Breaks

H_0 :	0.6	0.7	0.8	0.9
AU	0.4284	0.1132*	0.0304*	0.0084*
AUS	0.1997*	0.0602*	0.0184*	0.0051*
BE	0.4121	0.0817*	0.0091*	0.0021*
CA	0.9027	0.2466	0.0678*	0.0181*
DK	0.1953*	0.0539*	0.0152*	0.0043*
FI	0.8050	0.2284	0.0645*	0.0181*
FR	0.8796	0.3228	0.1309	0.0400*
GE	0.4001	0.0979*	0.0188*	0.0053*
GR	0.7318	0.2077	0.0534*	0.0032*
IT	1.7063	0.4857	0.1469	0.0401
JP	0.3638*	0.0987*	0.0269*	0.0070*
LX	0.6286	0.1688*	0.0456*	0.0121*
NL	0.1350*	0.0299*	0.0086*	0.0025*
NZ	0.3422*	0.0998*	0.0224*	0.0062*
NO	0.6561	0.1847	0.0520*	0.0141*
PO	0.5419	0.1476*	0.0409*	0.0114*
SP	0.6018	0.1658*	0.0460*	0.0129*
SWE	0.3004*	0.0823*	0.0229*	0.0061*
SWI	0.6391	0.1782*	0.0494*	0.0131*
UK	0.8408	0.2339	0.0648*	0.0180*
USA	1.4462	0.4098	0.1136	0.0313*
Crit. Values (5% *S.L.*)	0.399	0.175	0.0844	0.0404

(8), scaled by T^{2d-1}, is nonstandard, and critical values are provided in Mayoral (2004b). Again, since short-term structure is allowed, the test statistic has been corrected using standard nonparametric techniques (see Mayoral [2004b] for details).

Table 7 summarizes the output of the tests. For fifteen out of the twenty-one countries considered, the null hypothesis of fractional integration cannot be rejected.[16] The countries for which this

[16]Notice that the simulations reported in Mayoral (2004b) show that the employed techniques are very powerful against a wide variety of DGPs under the alternative hypothesis, with rejection rates ranging from 90 to 100 percent for this sample size. Then, we are confident that the nonrejection of the null hypothesis is not due to lack of power.

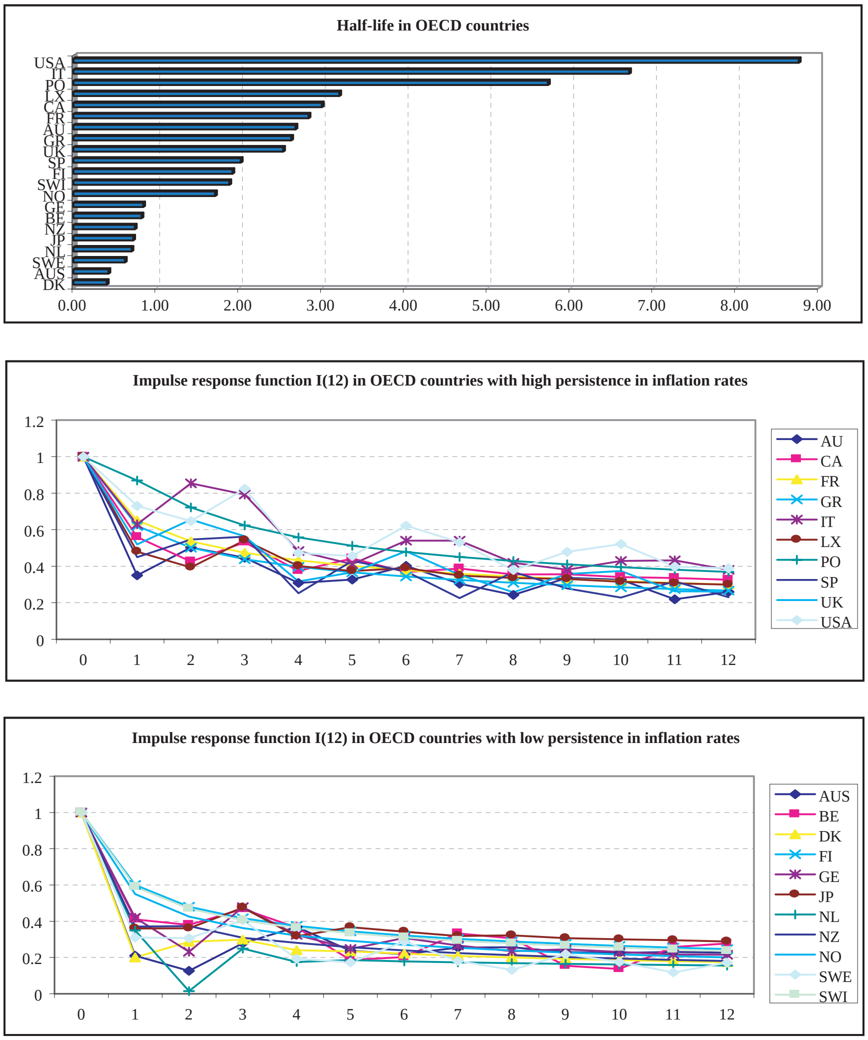
hypothesis is dismissed are Austria, Denmark, Japan, Netherlands, New Zealand, and Sweden. Two more countries, Belgium and Germany, are on the border between rejection and nonrejection. For these eight countries, the hypothesis of $d > 0.5$ versus I(0)+ breaks has also been tested, and the null was only rejected for four of them (NL, DK, AUS, and SWE). To understand this finding, it is interesting to look at the first graph of figure 2, which depicts the half-life measure of persistence, and figure 3, which shows the IRF(h) of these countries. Notice that, in figure 2, the latter four countries appear at the very bottom of the graph, implying that they are the least persistent. Right above those four, JP, NZ, BE, and GE are found. Therefore, it seems that at least some of the persistence that has been found in these series is spurious and derives from the existence of some breaks in the average level of inflation.

4. Measuring Persistence

In sections 2 and 3 we have presented an economic explanation and some robust empirical evidence supporting the hypothesis of fractionally integrated behavior in inflation data. Bearing in mind these results, we turn now to the main goal of the paper, the measurement of inflation persistence. In the following, by *persistence* we mean the long-term effect of a shock to the series.

In this section we provide various persistence measures that permit an adequate comparison of inflation inertia across countries and their evolution over time. The relevance of explicitly considering FI alternatives will become clear now. Our results demonstrate that, although in the short run the estimated persistence from the ARIMA and ARFIMA specifications is similar, the medium- and long-run implications are very different. This is due to the fact that, in order to model nonstationarity, ARIMA models necessarily impose the restriction of permanent shocks, while the more flexible ARFIMA formulations are able to characterize nonstationarity without imposing such a restriction. We show that some scalar measures of persistence, such as the sum of the AR coefficients (or its equivalent, the *cumulative impulse response* [see Andrews and Chen 1994]) are not suitable for measuring persistence in this context since they deliver exactly the same value for all FI(d) processes with $d > 0$ (equal to one for the former and to ∞ for the latter), despite the fact that processes

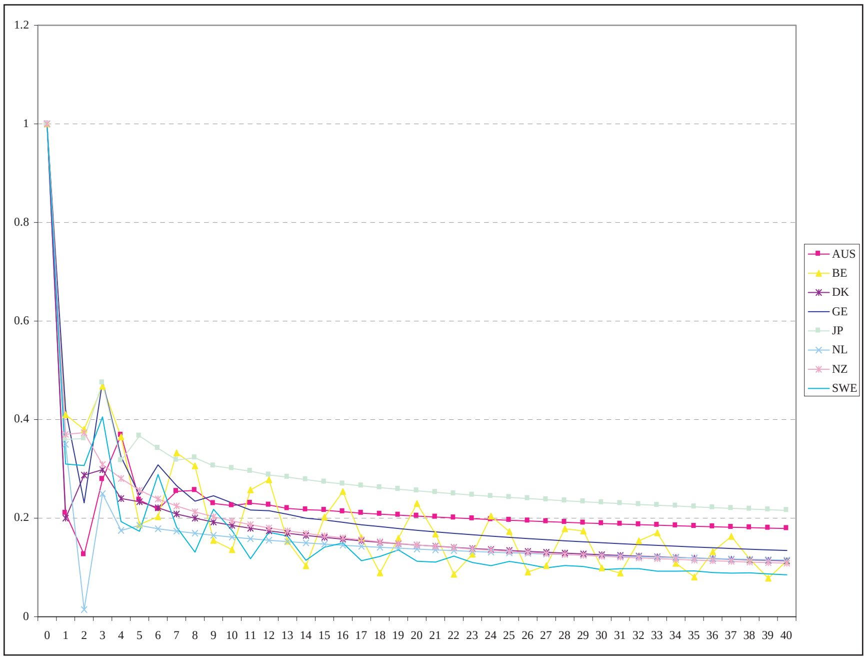
Figure 2. Half-Life and Impulse Response Functions in the Middle Run



in this group are of a very different character. In relation to this behavior, we also discuss some potential pitfalls that these techniques may present when used in applied work.

There are several ways to measure persistence, each with its virtues and faults. In the next subsection, we describe the tools that

Figure 3. Impulse Response Functions of Countries with Inflation Process I(0) with a Break in the Mean



will be used in this analysis. In order to have an accurate picture of this important property, we consider the estimation under both the classical and the Bayesian approach. Subsections 4.2 and 4.3 report the corresponding results.

4.1 Measuring Persistence with FI Processes

We consider three different tools in order to evaluate persistence—firstly, the *impulse response function* (*IRF*), which measures "the effect of a change in the innovation ε_t by a unit quantity on the current and subsequent values of y_t" (see Andrews and Chen 1994, 189). This measure is problematic because it is a vector, not a scalar, and, therefore, could be more difficult to interpret. For this reason, we also consider two scalar measures that will be described below.

For stationary series, the impulse responses are the coefficients of their Wold decomposition. For I(1) processes, the IRF(h) is usually computed[17] as the sum from zero to h of the impulse response coefficients of the first differences of the original series.[18] The above-mentioned expressions are embedded in the general formulation of the IRF(h) of an ARFIMA(p, d, q) process. This is defined as the h-th coefficient of $A(L) = (1 - L)^{-d}\Phi(L)^{-1}\Theta(L)$, where $\Phi(L)$ and $\Theta(L)$ are the AR and MA polynomials, respectively. The corresponding coefficients can be computed according to the following formula (see Koop et al. [1997] for details):

$$IRF(h) = \sum_{i=0}^{h} \pi_i(-d) J(h - i), \tag{9}$$

where each $\pi_i(-d)$ comes from the binomial expansion of $(1 - L)^{-d}$ and is defined in (2), and $J(\cdot)$ is the standard ARMA(p, q) impulse response, given by

$$J(i) = \sum_{j=0}^{q} \theta_j f_{i+1-j},$$

with $\theta_0 = 1$, $f_h = 0$ for $h \leq 0$, $f_1 = 1$, and

$$f_h = -(\phi_1 f_{h-1} + \ldots + \phi_p f_{h-p}), \qquad \text{for } h \geq 2.$$

Notice that if $d = 1$, $\pi_i(-1) = 1$ for all i and, therefore, the traditional IRF for I(1) processes is recovered, i.e., IRF$(h) = \sum_{i=0}^{h} J(h-i)$ (see Campbell and Mankiw 1987, 861). The limit behavior of the IRF(h) when $h \to \infty$ depends upon the value of d and verifies

$$IRF(\infty) = \begin{cases} 0, \text{ if } d < 1, \\ \Phi(1)^{-1}\Theta(1), \text{if } d = 1, \\ \infty \text{ if } d > 1. \end{cases} \tag{10}$$

[17] A different approach, which will not be pursued in this article, is to compute impulse responses based on estimating local projections at each period of interest (see Jordà 2005).

[18] Since the IRF(h) function in the I(1) case is computed by accumulating the individual I(0) impulse responses, it is often called *cumulative impulse response function* (see, for instance, Diebold and Rudebusch 1989). However, we will not use this terminology here in order to avoid confusion with other measures that share a similar name. This is the case of the *cumulative impulse response* (see Andrews and Chen 1994).

Expression (10) means that the effect of a shock is transitory for $d < 1$, as the long-term impact of any shock is equal to zero. By contrast, shocks are permanent for any $d \geq 1$. If the process contains a unit root ($d = 1$), the long-run effect of the shock is bounded away from zero and finite and is given by the sum of the Wold coefficients of its stationary transformation (or alternatively, by $\Phi(1)^{-1}\Theta(1)$ if it admits an ARMA representation). Finally, for any $d > 1$ the effect of any shock is magnified and the final impact is not bounded. Based on this behavior, Hauser, Pötscher, and Reschenhofer (1999) have criticized the use of ARFIMA models for measuring persistence. They argue that, although the ARIMA class is nested within the more general ARFIMA formulation, it would not be wise to use these models if the true DGP is in fact ARIMA. This is so because if $d = 1$, it would be extremely unlikely to obtain exactly this estimated value in finite samples. Thus, since the IRF(∞) is highly discontinuous, this would be equivalent to imposing an a priori value of this function either equal to zero (if $\hat{d} < 1$) or to infinity (if $\hat{d} > 1$). According to their view, imposing these long-term restrictions would also adversely affect the estimation of the IRF(h) for finite values of h (see the simulations provided in Hauser, Pötscher, and Reschenhofer (1999, Table 1).

We agree with them that, for the purpose of persistence estimation, it is important to treat the ARFIMA and the ARIMA classes as two different groups of models, despite the fact that one contains the other. This is one of the reasons that led us to apply an ample battery of tests to distinguish between these formulations in our data set. But, in our opinion, it does not follow from here that the use of ARFIMA processes is inadequate to measure persistence. There are several ways in which the criticisms in Hauser, Pötscher, and Reschenhofer (1999) can be answered. The most obvious is that their misspecification argument can be easily reversed, that is, if the DGP is FI(d) but an ARIMA model (with integer d) is fitted to the data to compute the impulses, the (wrong) long-term restrictions imposed by the ARIMA specification might bias the estimates as a result of the misspecification. Since the empirical evidence found in the previous section supports the better fit of the ARFIMA over the ARIMA model, the use of the former is well justified. The estimated values of d obtained for our data set are, in general, less than one, which means that the IRF(∞) associated with these processes is zero. This restriction reflects the main finding of section 3: the inflation rate is best characterized as a nonstationary but mean-reverting process. If this condition is true, imposing a unit root to compute the impulses will result in higher estimated persistence, since the permanent shock restriction will upwardly bias the estimates. This fact is illustrated in table 9, shown in subsection 4.3.

Finally, we are aware that it is not possible to be certain about the true nature of the DGP. So, in order to avoid possible biases in our estimates stemming from imposing a possibly incorrect long-term restriction, in subsection 4.3 we estimate the impulse responses using a Bayesian approach that explicitly acknowledges model uncertainty. By allowing for a strictly positive probability mass on the I(1) model, we will be able to obtain a continuous impulse response function with a strictly positive and bounded value at infinity. To do so, we will follow the approach of Koop et al. (1997).

In addition to the IRF, two scalar measures of persistence are also reported: the half life (HL), defined as the number of periods that a shock needs to vanish by 50 percent, and ρ_{40}, which is given by

$$\rho_{40} = 1 - 1/\sum_{h=0}^{40} IRF(h).$$

This quantity can be interpreted as a truncated version of the sum of the AR coefficients (see Andrews and Chen 1994), defined as

$$\rho(1) = 1 - 1/\sum_{h=0}^{\infty} IRF(h),$$

and is introduced here in order to overcome the problems that this measure presents in this context. It turns out that $\rho(1) = 1$ for any integrated process with an integration order strictly greater than zero. This is so because any invertible FI(d) process admits an AR(∞) representation, given by

$$(1 - L)^d C(L)^{-1} y_t = \varepsilon_t,$$

where the innovations $\{\varepsilon_t\}_{-\infty}^{\infty}$ are white noise and $C(L)$ is the polynomial of the Wold representation of the I(0) variable $(1-L)^d y_t$. For any $d > 0$, $L = 1$ is a root of the polynomial $(1-L)^d C(L)^{-1}$. Calling

$\Lambda(L) = (1-L)^d C(L)^{-1} = 1 - \sum_{i=1}^{\infty} \lambda_i L^i$ and noticing that $L = 1$ is a root of $\Lambda(L)$, it follows that $1 - \sum_{i=1}^{\infty} \lambda_i 1^i = 0$, which implies that $\rho(1) = \sum_{i=1}^{\infty} \lambda_i = 1$. An equivalent way of looking at this result is by considering the *cumulative impulse response* (*CIR*), given by

$$CIR = 1/(1-\rho(1)) = \sum_{i=0}^{\infty} IRF(h).$$

This measure is proportional to the spectral density at frequency zero (see Andrews and Chen 1994). Since the spectral density of any FI(d) process with $d > 0$ is unbounded at frequency zero, it follows that CIR $= \infty$ for any FI(d) process with $d > 0$. Since the degree of persistence varies a great deal across the different values of d in this range, it follows that $\rho(1)$ cannot be taken as a good measure of persistence in this case. To overcome this problem, we consider a truncated version of it, ρ_{40}, which, instead of considering the sum of the IRF(h) for $h = 1, \ldots, \infty$, only considers the first forty coefficients (which we identify with the long run). Interestingly, this measure can be considerably far from 1 for moderate values of d (for instance, in an FI(i) process with $i = \{0.1, 0.2, 0.3\}$, it would be around 0.35, 0.59, and 0.74, respectively).

4.2 Classical Estimation

We now report the estimated values of the three tools presented above, obtained using classical techniques. Table 8 presents the IRF(h) at different time horizons h, namely, $h = 4$, 12, and 40, representing the short, middle, and long run, respectively. In addition, columns 4 and 5 report the values of the HL and ρ_{40}, respectively. The full path of IRF(h) for different time horizons is displayed in figure 4.

The information in table 8 can be summarized as follows. For the twenty-one industrial countries, the IRF decreases in the middle- and long-run horizons, although the remaining effect of shocks differs considerably across countries, ranging from 38 percent for the United States versus 17 percent for Sweden in the middle-run horizon, and 30 percent versus 8 percent for the same countries in the long term. The ρ_{40} measure oscillates within the interval [0.90,0.96], confirming the high persistence of the series. It is interesting to compare this result with the one obtained in Pivetta and Reis (2004).

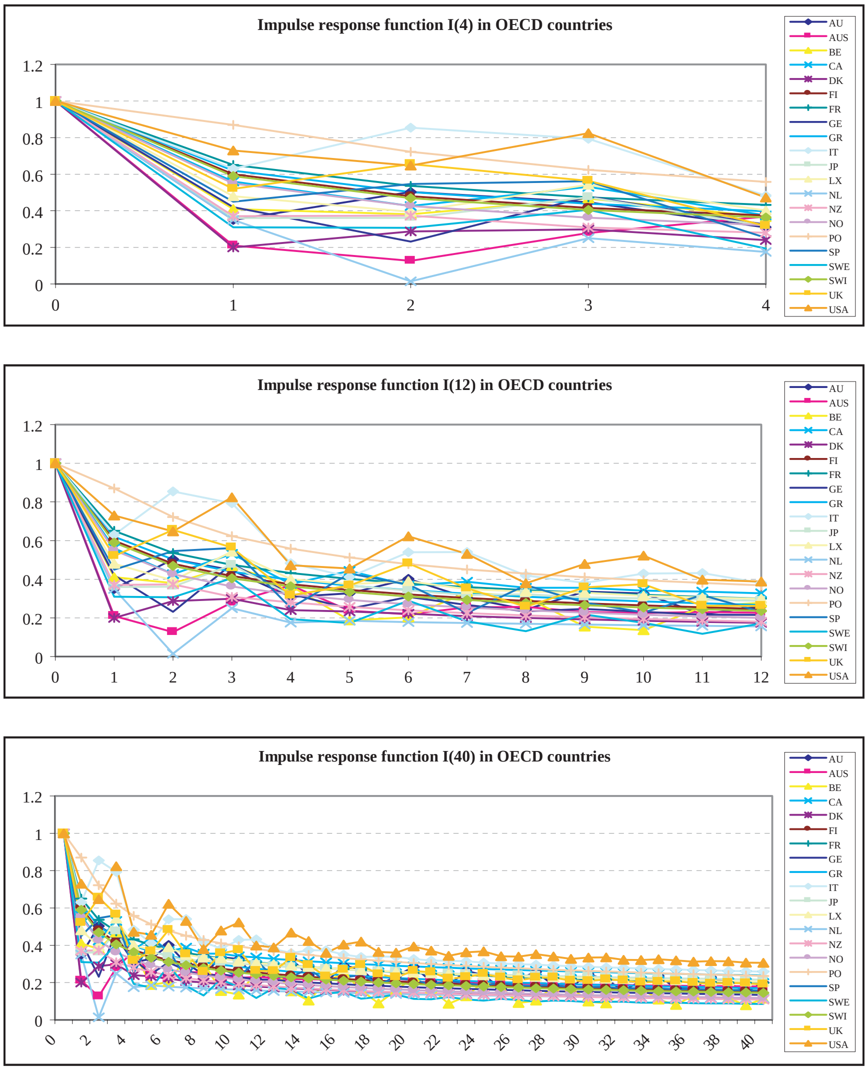
Table 8. IRF and Scalar Measures of Persistence

	IRF(4)	IRF(12)	IRF(40)	HL	ρ_{40}
AU	0.3087	0.2601	0.2002	2.68	0.94
AUS	0.3684	0.2266	0.1789	0.42	0.94
BE	0.3650	0.2779	0.1135	0.82	0.92
CA	0.3749	0.3267	0.2386	3.00	0.95
DK	0.2399	0.1742	0.1134	0.40	0.91
FI	0.3744	0.2461	0.1531	1.92	0.90
FR	0.4324	0.3000	0.1980	2.84	0.92
GE	0.3246	0.2152	0.1344	0.84	0.93
GR	0.3969	0.2666	0.1698	2.63	0.94
IT	0.4829	0.3805	0.2585	6.71	0.96
JP	0.3168	0.2879	0.2158	0.72	0.95
LX	0.4101	0.2991	0.2036	3.21	0.95
NL	0.1751	0.1553	0.1149	0.70	0.91
NZ	0.2804	0.1803	0.1087	0.74	0.91
NO	0.3316	0.2002	0.1173	1.71	0.92
PO	0.5581	0.3689	0.2360	5.73	0.96
SP	0.2521	0.2307	0.1573	2.02	0.91
SWE	0.1930	0.1713	0.0853	0.62	0.90
SWI	0.3634	0.2363	0.1453	1.88	0.93
UK	0.3166	0.2615	0.1902	2.53	0.95
USA	0.4726	0.3883	0.3058	8.76	0.96

Notes: IRF(h), h = 4, 12, 40 denote the impulse response function. HL is the half life defined as the number of periods that a shock needs to vanish by 50 percent. ρ_{40} is computed as $1 - 1/\sum_{h=1}^{40}$ IRF(h).

They estimate $\rho(1)$ for the U.S. inflation rate from an AR(p) specification, where $p = 3$ is chosen according to the Bayesian information criterion (BIC). They obtain estimates of this quantity around 0.95 and they conclude that inflation has a unit root and, therefore, that shocks to inflation are permanent. Nevertheless, as has been shown above, a value of $\rho(1)$ close to one does not imply an integer unit root but only a fractional one. Thus, one cannot say much about inflation

Figure 4. Impulse Response Functions



persistence just by looking at this quantity since very different types of integrated processes share this property.

In order to illustrate this, we have carried out a small Monte Carlo experiment: we have generated 5,000 ARFIMA$(0, d, 0)$ processes with a value of $d = 0.7$ (which is approximately the estimated value for U.S. inflation obtained in section 3; see table 3). Then we have fitted an AR(p) process using the BIC as in Pivetta and Reis (2004). Although the DGP is AR(∞), any sensitive information criteria will select a much shorter lag length. In fact, we have found that, on average, the chosen lag length is $p = 3$ and the mean (median) of $\rho(1)$ is 0.89 (0.90) with a standard deviation equal to 0.26. This example shows that the traditional interpretation that identifies $\rho(1) \approx 1$ with the existence of an integer unit root is clearly unfounded and could lead to persistence overestimation if one concludes from here that shocks are permanent.

A related problem can be found in Cogley and Sargent (2001). These authors assume that inflation is stationary. In order to impose this assumption, they truncate the parameter space so that the largest autoregressive root (LAR) is strictly less than one. Thus, they are imposing not only stationarity (which is compatible with an LAR equal to one in a fractional model with $d < 0.5$) but short memory (bounded spectral density). As Pivetta and Reis (2004) point out, this truncation could strongly bias the results toward lower values of persistence.

Figure 2, shown previously in subsection 3.3, ranks the different countries in accordance with their HL value and shows that its behavior varies a lot across them. Broadly speaking, two groups can be distinguished: the *low inflation persistence* group, exhibiting an HL of less than two periods (equivalent to six months), and the *high inflation persistence* group, with an HL superior to two periods. In the first group, the Scandinavian countries SWE, FI, and NO, together with JP, NZ, and SWI, can be found. All of them show a low inflation rate in most of the period with a mean around 4 percent. Other countries such as AUS, DK, NL, BE, and GE are also included in this group and are characterized by a tight monetary discipline and an implicit commitment with the German currency, whether they belonged to the European Monetary System or not. The members of the second group are AU, CA, FR, GR, LX, IT, PO, SP, UK, and USA with an inflation mean around 6 percent. The United States is the country with the highest HL, with a value around two years. However, this quantity is considerably smaller than that obtained by Pivetta and Reis (2004), who present figures of the HL of more than five years. This important difference in magnitude is a consequence of the use of the I(1) (permanent shocks) specification instead of the FI(d) one with $d < 1$ (mean-reverting shocks) employed in this article.

4.3 Bayesian Estimation

We now turn to the Bayesian estimation of inflation persistence. Although we have found abundant evidence against integer values of d, in this subsection we acknowledge our uncertainty by considering different combinations of ARIMA and ARFIMA models. The main motivation for undertaking this analysis is to overcome the criticism presented by Hauser, Pötscher, and Reschenhofer (1999). They argued that ARFIMA models may not be appropriate for measuring persistence because they imply a limit behavior of IRF(∞) which is either zero (if $d < 1$) or infinity (if $d > 1$). Nevertheless, using Bayesian techniques, it is possible to achieve a continuous distribution of the IRF(∞) in the interval $[0, \infty)$ if a strictly positive prior probability is assumed for the integer values of d. Following Koop et al. (1997), we have considered sixteen ARIMA models (where $d = 1$ is imposed) and sixteen ARFIMA models, corresponding to the different combinations of ARMA parameters, with $p, q \leq 3$ in both cases. In order to determine the prior probabilities assigned to both groups of models, we will use the posterior probabilities of $d_i < 1$ that were obtained in subsection 3.1. It is clear that $P(d \geq 1) = 1 - P(d < 1)$ and, therefore, we can use this expression as an upper bound for the probability of $P(d = 1)$. This quantity will be used as the prior probability for the ARIMA models. Table 9 reports the IRF evaluated at different time horizons for the best ARIMA and ARFIMA models (the ones with highest posterior probability) and also for the OVERALL model, constructed as a sum of the thirty-two models weighted by their posterior probabilities.

Bayesian IRFs present slightly higher values than those obtained under the classical paradigm, but in general, the nonpermanent character of shocks and the classification among countries is maintained. It is also interesting to compare the results obtained from the ARFIMA and ARIMA models. Both deliver very similar values in the short run, but they are very different in the medium and long run. Therefore, if only ARIMA alternatives are considered, it is very easy to conclude that shocks are much more persistent than they actually are.

Table 9. Bayesian Estimation of IRF

	IRF(4)			IRF(12)			IRF(40)		
	B-FI	B-I	All	B-FI	B-I	All	B-FI	B-I	All
AU	0.35 (0.06)	0.38 (0.06)	0.38 (0.07)	0.31 (0.09)	0.38 (0.06)	0.34 (0.09)	0.29 (0.13)	0.38 (0.06)	0.30 (0.13)
AUS	0.15 (0.04)	0.22 (0.05)	0.15 (0.04)	0.08 (0.03)	0.22 (0.05)	0.08 (0.03)	0.04 (0.02)	0.22 (0.05)	0.04 (0.02)
BE	0.53 (0.10)	0.57 (0.03)	0.52 (0.08)	0.34 (0.11)	0.42 (0.03)	0.37 (0.08)	0.30 (0.14)	0.41 (0.03)	0.32 (0.11)
CA	0.46 (0.08)	0.50 (0.07)	0.44 (0.08)	0.44 (0.14)	0.50 (0.07)	0.39 (0.11)	0.47 (0.24)	0.50 (0.07)	0.36 (0.15)
DK	0.22 (0.05)	0.22 (0.05)	0.21 (0.05)	0.19 (0.05)	0.22 (0.05)	0.19 (0.06)	0.17 (0.07)	0.22 (0.05)	0.17 (0.07)
FI	0.40 (0.07)	0.58 (0.08)	0.42 (0.12)	0.27 (0.07)	0.58 (0.08)	0.32 (0.10)	0.18 (0.06)	0.58 (0.08)	0.24 (0.12)
FR	0.45 (0.08)	0.44 (0.10)	0.43 (0.09)	0.32 (0.08)	0.40 (0.11)	0.32 (0.10)	0.22 (0.08)	0.40 (0.11)	0.24 (0.11)
GE	0.34 (0.09)	0.34 (0.06)	0.34 (0.08)	0.41 (0.17)	0.36 (0.10)	0.33 (0.12)	0.39 (0.26)	0.25 (0.17)	0.29 (0.16)
GR	0.42 (0.07)	0.35 (0.08)	0.41 (0.08)	0.29 (0.07)	0.36 (0.09)	0.32 (0.099)	0.19 (0.06)	0.40 (0.12)	0.27 (0.10)
IT	0.56 (0.07)	0.87 (0.13)	0.51 (0.11)	0.46 (0.09)	0.90 (0.14)	0.42 (0.14)	0.35 (0.15)	0.92 (0.18)	0.32 (0.17)
JP	0.08 (0.03)	0.30 (0.13)	0.32 (0.09)	0.08 (0.02)	0.28 (0.09)	0.26 (0.10)	0.05 (0.02)	0.26 (0.10)	0.18 (0.10)
LX	0.41 0.09)	0.47 (0.06)	0.42 (0.09)	0.47 (0.21)	0.47 (0.06)	0.42 (0.13)	0.53 (0.41)	0.47 (0.06)	0.39 (0.19)
NL	0.18 (0.06)	0.18 (0.06)	0.19 (0.06)	0.18 (0.06)	0.20 (0.05)	0.17 (0.06)	0.17 (0.09)	0.21 (0.05)	0.16 (0.07)
NZ	0.30 (0.07)	0.28 (0.07)	0.29 (0.08)	0.26 (0.09)	0.26 (0.09)	0.24 (0.08)	0.24 (0.14)	0.24 (0.14)	0.21 (0.09)
NO	0.34 (0.06)	0.33 (0.06)	0.36 (0.08)	0.22 (0.05)	0.26 (0.05)	0.27 (0.08)	0.13 (0.04)	0.15 (0.11)	0.21 (0.10)
PO	0.21 (0.03)	0.36 (0.07)	0.32 (0.07)	0.31 (0.04)	0.36 (0.07)	0.33 (0.08)	0.32 (0.06)	0.36 (0.07)	0.32 (0.08)
SP	0.21 (0.09)	0.31 (0.07)	0.31 (0.08)	0.25 (0.10)	0.38 (0.12)	0.36 (0.14)	0.24 (0.19)	0.30 (0.18)	0.32 (0.12)
SWE	0.21 (0.04)	0.23 (0.08)	0.27 (0.06)	0.11 (0.03)	0.23 (0.08)	0.23 (0.07)	0.06 (0.02)	0.23 (0.08)	0.20 (0.09)
SWI	0.38 (0.07)	0.36 (0.07)	0.41 (0.08)	0.25 (0.06)	0.26 (0.12)	0.30 (0.11)	0.16 (0.05)	0.16 (0.17)	0.22 (0.13)
UK	0.38 (0.07)	0.74 (0.08)	0.44 (0.08)	0.26 (0.06)	0.56 (0.08)	0.36 (0.10)	0.16 (0.05)	0.56 (0.08)	0.29 (0.11)
USA	0.68 (0.11)	0.62 (0.07)	0.66 (0.13)	0.51 (0.17)	0.42 (0.11)	0.53 (0.19)	0.32 (0.20)	0.22 (0.17)	0.42 (0.23)

Note: B-FI: best ARFIMA; B-I: best ARIMA; All: overall models. Standard deviation is shown in parentheses.

Summarizing, in agreement with previous findings, this section confirms the high degree of inflation inertia. The United States emerges as the country with the highest inflation persistence in contrast to the Nordic countries, which display the lowest rates. Interestingly, high inertia is compatible with mean-reverting shocks in the framework considered in this article, a feature that cannot be captured in the I(1) setup. This finding is relevant in many contexts, for instance, if one is interested in testing monetary neutrality.

5. Changes in Persistence

Another issue that has been widely studied recently is the stability of persistence over time. Changes in persistence may have a decisive impact on monetary strategy design. Some authors have pointed out that, if there is a decrease in inflation persistence, tests of the natural rate hypothesis in the spirit of Solow (1968) or Tobin (1968) may reject the null hypothesis of monetary neutrality as a consequence of this decrease. On the other hand, monetary policy is usually implemented in a more aggressive way in a context where inflation persistence increases. Furthermore, many macroeconomic models incorporate a measure of the persistence of inflation and, if persistence is not constant over time, Lucas's critique could apply.

The hypothesis of the stability of inflation persistence has been tested recently in various articles. Nevertheless, no consensus seems to have been reached. On the one hand, authors such as Taylor (2000), Cogley and Sargent (2001), and Kim, Nelson, and Piger (2004) have found that inflation inertia has decreased in recent years as a result of a general deflationist process, the implementation of target rules, and a more credible performance of central banks.[19] On the other hand, Stock (2001), Batini (2002), Levin and Piger (2003), O'Reilly and Whelan (2004), Hondroyiannis and Lazaretou (2004), and Pivetta and Reis (2004) have found little evidence of changes in persistence for different countries.

In many of the latter papers, the decrease in persistence has been tested by checking whether the sum of the AR coefficients has

[19] By using a more historical perspective, some authors have found changes in persistence linked to different monetary regimes (c.f. Barsky [1987], Alogoskoufis and Smith [1991], Alogoskoufis [1992], Bordo and Schwartz [1999], Kim [2000], and Benati [2002]).

changed from one to a value strictly smaller than one. But, as was pointed out in section 4, this procedure is not completely correct if FI is allowed for. If a process is FI(d), the sum of the AR coefficients is equal to one for any $d > 0$. So, a decrease in persistence, associated with a lower value of d, does not have any theoretical impact on this sum (whose value will remain equal to one) as long as the new d is larger than zero. Therefore, a test based on the aforementioned criteria is likely to have very low power.

In this section, we will explore the stability of inflation persistence using a different approach. We will directly test whether the memory parameter d has remained constant over time or not. In order to do so, a Lagrange multiplier (LM) test of the stability of d will be applied (see Mayoral [2005] for further details). The following DGP is considered:

$$\begin{aligned} y_t &= \mu + x_t \\ \Delta^{d+\theta D_t(\omega)} x_t &= \Phi(L)^{-1}\Theta(L)\varepsilon_t. \end{aligned}$$

The process y_t is the sum of a constant term, μ, and a fractionally integrated process x_t. The parameter $\omega = t_0/T$ describes the location of a change in the value of d in the sample that, if it occurs, happens at time t_0. $D_t(\omega)$ is a dummy variable that takes the value one if $\omega T < t$ and zero otherwise. The process ε_t is assumed to be i.i.d. and $\Phi(L)$, $\Theta(L)$ are the standard AR and MA polynomials, respectively. Under the null hypothesis, there is no change in persistence and, therefore, $\theta = 0$. Under H_1, a single break in d is allowed to take place so that θ can take both positive and negative values, indicating an increase or a decrease of persistence, respectively. The test is developed following Andrews (1993) and works as follows: assuming normality, the test statistic derived under the LM principle for any fixed ω is given by

$$LM_T(\omega) = S_T(\omega)' A^{-1} S_T(\omega),$$

where S_T is the score obtained by deriving the likelihood function with respect to θ,

$$S_T(\omega) = \frac{\partial L(d, \theta, \sigma^2, \beta, \omega)}{\partial \theta} = \omega T \sum_{i=1}^{\omega T} \frac{1}{k} \hat{\rho}_k,$$

where $\hat{\rho}_k$ is the k-th correlation associated with the residuals after (parametrically) estimating x_t. The matrix A contains the relevant terms of the expression $E_0[\frac{\partial L}{\partial \eta}\frac{\partial L}{\partial \eta'}]$. Its form depends upon the ARMA components. For instance, in the case where $\Phi(L) = \acute{\Theta}(L) = 1$, it becomes

$$A = t_0 \sum_{i=1}^{t_0-1} \frac{1}{i^2}\left(1 - \frac{1}{it_0}\right).$$

It can be easily shown that for any fixed ω, $LM(\omega) \xrightarrow{w} \chi_1^2$. But since ω is, in general, unknown, we adopt a common method used in this scenario and consider test statistics of the form $\sup_{\omega\in\Omega} LM(\omega)$. Critical values can be found in Mayoral (2005). To carry out the test on our data set, residuals are computed using Sowell's ML method.

The first column of table 10 presents the results of the test while the second column displays the date of the break for the cases where it turned out to be significant. It is noteworthy that the results are very homogeneous across countries: for eighteen out of the twenty-one countries no evidence of a change in persistence has been found. That conclusion is only reversed for Austria, Belgium, and Germany, for which some evidence of a break in persistence is found. For all three countries, the shock is found at the beginning of the 1960s. Nevertheless, we should remember that we are running twenty-one tests at the 5 percent significance level and, therefore, we should expect some rejections even if the null hypothesis is true.

In short, our results agree with the recent literature that finds little empirical evidence supporting a change in inflation persistence.

6. Conclusions

This paper explores the inflation rates of a group of OECD countries, focusing on their persistence properties. We propose modeling this data set using ARFIMA models, since they are very flexible, to represent the medium- and long-run properties of time series. An economic justification for the existence of fractionally integrated behavior in the data, as well as solid empirical evidence supporting this hypothesis, is provided. In agreement with previous works, we find that inflation rates are very persistent but, in contrast to most

Table 10. Changes in Persistence

	supLM	Break Date
AU	6.026	–
AUS	25.601**	1964:1
BE	13.373*	1966:1
CA	1.759	–
DK	2.382	–
FI	2.004	–
FR	5.670	–
GE	10.340*	1963:2
GR	3.294	–
IT	1.738	–
JP	0.000	–
LX	6.033	–
NL	3.787	–
NZ	0.761	–
NO	3.270	–
PO	1.451	–
SP	6.850	–
SWE	2.060	–
SWI	1.106	–
UK	3.691	–
USA	4.577	–
C.V.	(9.68,13.5)	–

of them, we believe that shocks do not have, in general, a permanent effect, implying that the series are mean reverting. The latter finding is very relevant since it implies that the I(1) characterization is not suitable for this data set. We have shown that some widely used tools to measure persistence and to test its stability, such as the sum of the AR coefficients (or its equivalent, the cumulative impulse response) are not suitable if the DGP is FI. Since there is always uncertainty about the true DGP, these conclusions should always be taken into account when computing these tools.

Our measures of persistence allow us to establish cross-country comparisons, and it is shown that important differences arise between the nations that we have considered, which may be related to the different monetary institutions present in each of them. Finally, for most countries, little evidence in favor of a change in inflation persistence has been found, in accordance with the recent literature in this area.

Appendix. Evolution of the Inflation Series

Figure 5. Evolution of Inflation Rates in OECD Countries

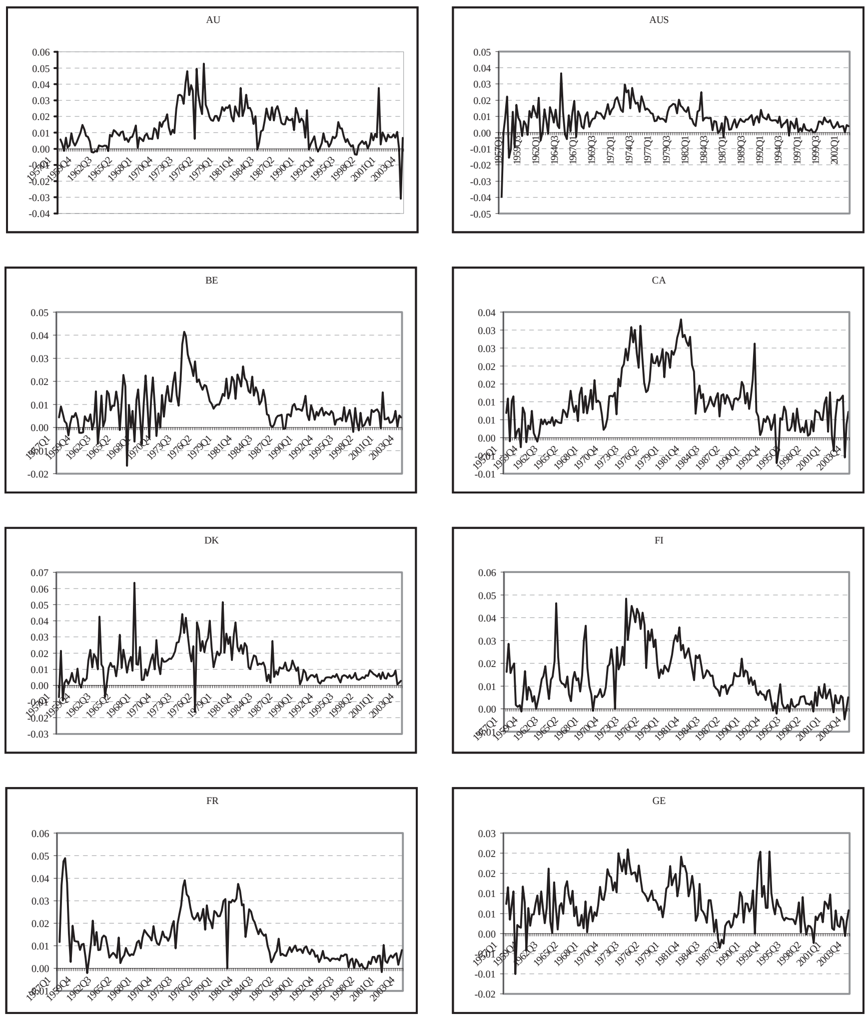


Figure 6. Evolution of Inflation Rates in OECD Countries

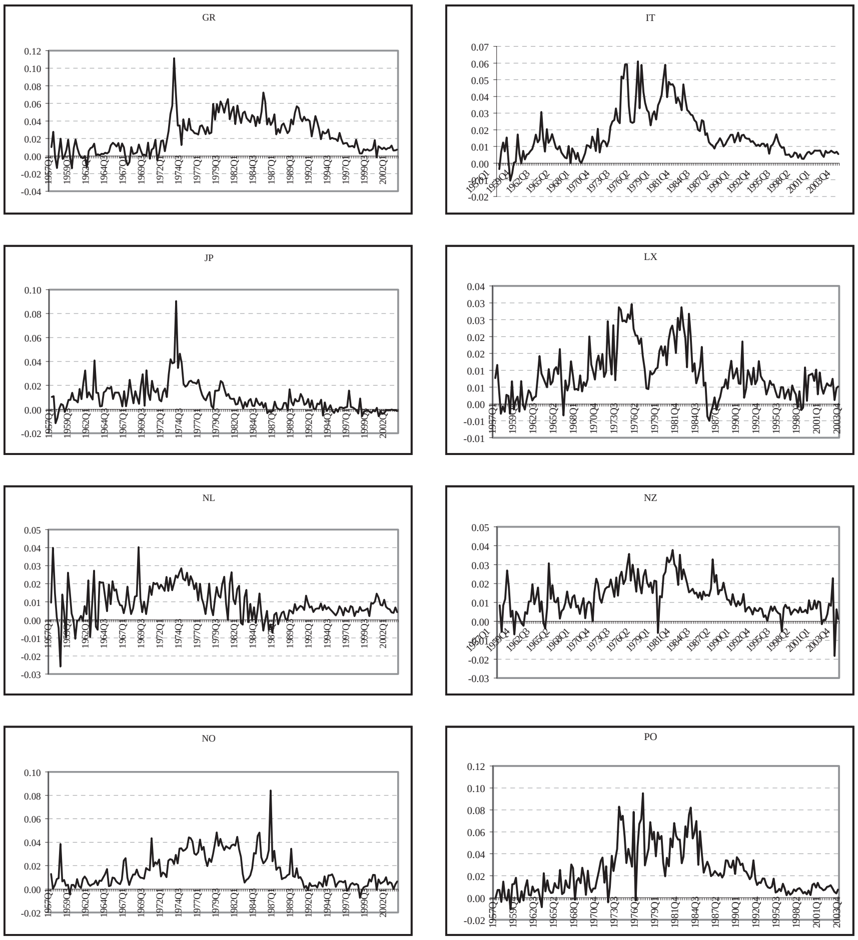
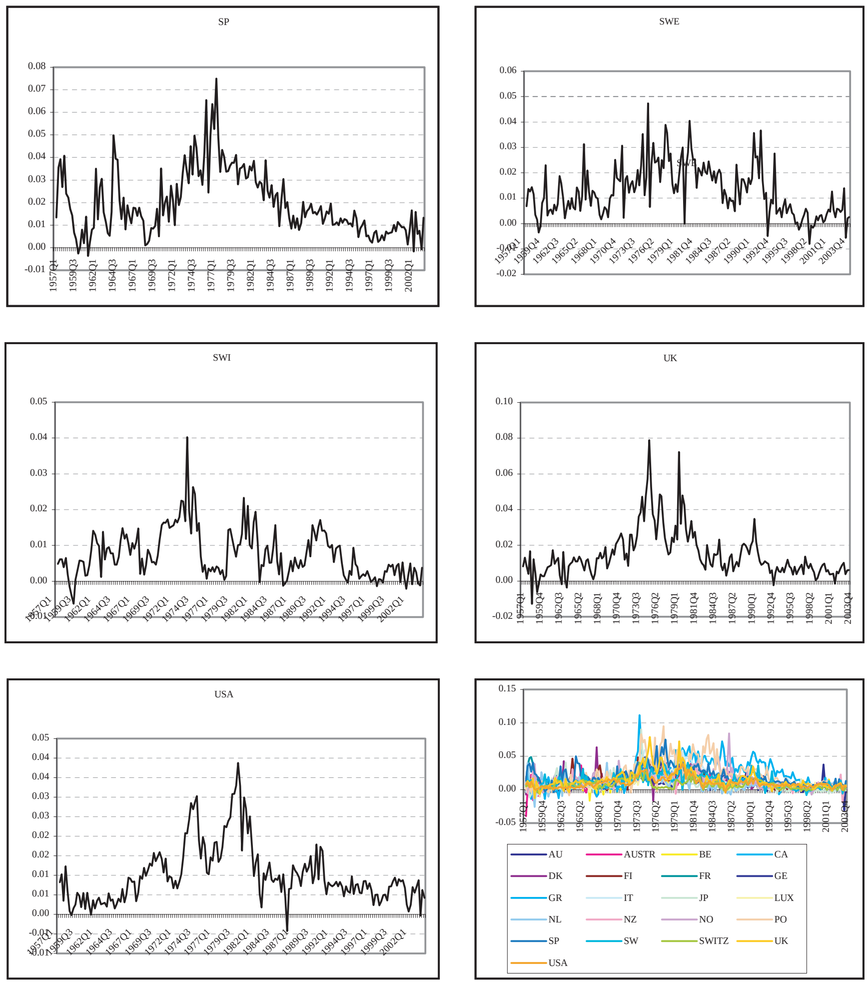


Figure 7. Evolution of Inflation Rates in OECD Countries



References

Abadir, Karim, and Gabriel Talmain. 2002. "Aggregation, Persistence and Volatility in a Macro Model." *Review of Economic Studies* 69 (4): 749–79.

Agiakloglou, Christos, Paul Newbold, and Mark Wohar. 1992. "Bias in an Estimator of the Fractional Difference Parameter." *Journal of Time Series Analysis* 14 (3): 235–46.

Alogoskoufis, George. 1992. “Monetary Accommodation, Exchange Rate Regimes and Inflation Persistence.” *The Economic Journal* 102 (410): 461–80.

Alogoskoufis, George, and Ronald Smith. 1991. “The Phillips Curve, the Persistence of Inflation, and the Lucas Critique: Evidence from Exchange-Rate Regimes.” *American Economic Review* 81 (5): 1254–75.

Andersen, Torben G., and Tim Bollerslev. 1997. “Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns.” *Journal of Finance* 52 (3): 975–1005.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 1999. “The Distribution of Exchange Rate Volatility.” NBER Working Paper No. 6961.

Andrews, Donald W.K. 1991. “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation.” *Econometrica* 59 (3): 817–58.

———. 1993. “Exactly Median-Unbiased Estimation of First Order Estimation of Autoregressive Unit Root Models.” *Econometrica* 61 (1): 139–65.

Andrews, Donald W.K., and Hong-Yuan Chen. 1994. “Approximately Median-Unbiased Estimation of Autoregressive Models.” *Journal of Business and Economic Statistics* 12:187–204.

Angeloni, Ignazio, Günter Coenen, and Frank Smets. 2003. “Persistence, the Transmission Mechanism and Robust Monetary Policy.” Working Paper No. 250, European Central Bank.

Bai, Jushan, and Pierre Perron. 1998. “Estimating and Testing Linear Models with Multiple Structural Changes.” *Econometrica* 66 (1): 47–78.

———. 2003a. “Computation and Analysis of Multiple Structural Change Models.” *Journal of Applied Econometrics* 18 (1): 1–22.

———. 2003b. “Critical Values for Multiple Structural Change Tests.” *The Econometrics Journal* 6 (1): 72–78.

Baillie Richard T., Ching-Fan Chung, and Margie A. Tieslau. 1992. “The Long Memory and Variability of Inflation: A Reappraisal of the Friedman Hypothesis.” Discussion Paper, Tilburg University.

———. 1996. “Analyzing Inflation by the Fractionally Integrated ARFIMA-GARCH Model.” *Journal of Applied Econometrics* 11 (1): 23–40.

Barkoulas, John T., Christopher Baum, and Gurkan S. Oguz. 1998. "Stochastic Long Memory in Traded Goods Prices." *Applied Economics Letters* 5 (3): 135–38.

Barsky, Robert B. 1987. "The Fisher Hypothesis and the Forecastability and Persistence of Inflation." *Journal of Monetary Economics* 19 (1): 3–24.

Batini, Nicoletta. 2002. "Euro Area Inflation Persistence." Working Paper No. 201, European Central Bank.

Baum, Christopher F., John T. Barkoulas, and Mustafa Caglayan. 1999. "Persistence in International Inflation Rates." *Southern Economic Journal* 65 (4): 900–14.

Benati, Luca. 2002. "Investigating Inflation Persistence across Monetary Regimes I: Empirical Evidence." Mimeo, Bank of England.

Beran, Jan. 1994. *Statistics for Long-Memory Processes*. New York: Chapman & Hall.

Bhattacharya, R. N., V. K. Gupta, and E. Waymire. 1983. "The Hurst Effect under Trends." *Journal of Applied Probability* 20: 649–62.

Bordo, Michael D., and Anna J. Schwartz. 1999. "Under What Circumstances, Past and Present, Have International Rescues of Countries in Financial Distress Been Successful?" *Journal of International Money and Finance* 18 (4): 683–708.

Bos, Charles, Philip Hans Franses, and Marius Ooms. 1999. "Long Memory and Level Shifts: Re-analyzing Inflation Rates." *Empirical Economics* 24 (3): 427–49.

———. 2002. "Inflation, Forecast Intervals and Long Memory Regression Models." *International Journal of Forecasting* 18 (2): 243–64.

Breidt, F. Jay, Nuno Crato, and Pedro de Lima. 1998. "The Detection and Estimation of Long Memory in Stochastic Volatility." *Journal of Econometrics* 83 (1–2): 325–48.

Byers, David, James Davidson, and David Peel. 1997. "Modelling Political Popularity: An Analysis of Long-Range Dependence in Opinion Poll Series." *Journal of the Royal Statistical Society,* Series A, 160 (3): 471–90.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3): 383–98.

Campbell, John Y., and N. Gregory Mankiw. 1987. "Are Output Fluctuations Transitory?" *Quarterly Journal of Economics* 102 (4): 857–80.

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans. 2001. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." Working Paper No. 2001-08, Federal Reserve Bank of Chicago.

Coenen, Günter. 2003. "Inflation Persistence and Robust Monetary Policy Design." Working Paper No. 290, European Central Bank.

Coenen, Günter, and Volker Wieland. 2005. "A Small Estimated Euro Area Model with Rational Expectations and Nominal Rigidities." *European Economic Review* 49 (5): 1081–1104.

Cogley, Timothy, and Thomas J. Sargent. 2001. "Evolving Post-World War II U.S. Inflation Dynamics." In *NBER Macroeconomics Annual*, ed. Ben S. Bernanke and Kenneth S. Rogoff. Cambridge, MA: MIT Press.

Delgado, M., and Peter Robinson. 1994. "New Methods for the Analysis of Long-Memory Time Series." *Journal of Forecasting* 13:97–107.

Dickey, David A., and Wayne A. Fuller. 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica* 49 (4): 1057–72.

Diebold, Francis X., and Glenn D. Rudebusch. 1989. "Long Memory and Persistence in Aggregate Output." *Journal of Monetary Economics* 24 (2): 189–209.

———. 1991. "On the Power of the Dickey-Fuller Tests against Fractional Alternatives." *Economic Letters* 35 (2): 155–60.

Diebold, Francis X., and Abdelhak S. Senhadji. 1996. "The Uncertain Unit Root in Real GNP: Comment." *American Economic Review* 86 (5): 1291–98.

Ding, Zhuanxin., Clive W.J. Granger, and Robert F. Engle. 1993. "A Long Memory Property of Stock Market Returns and a New Model." *Journal of Empirical Finance* 1 (1): 83–106.

Dolado Juan J., Jesús Gonzalo, and Laura Mayoral. 2002. "A Fractional Dickey-Fuller Test for Unit Roots." *Econometrica* 70 (5): 1963–2006.

———. 2003. "Testing for a Unit Root against Fractional Alternatives in the Presence of a Maintained Trend." Mimeo.

Doornik, Jurgen A., and Marius Ooms. 2001. "A Package for Estimating, Forecasting and Simulating ARFIMA Models: ARFIMA Package 1.01 for Ox." Mimeo, University of Rotterdam.

Driscoll, John, and Steiner Holden. 2004. "Fairness and Inflation Persistence." *Journal of the European Economic Association* 2 (2): 240–51.

Franses, Philip Hans, and Marius Ooms. 1997. "A Periodic Long-Memory Model for Quarterly UK Inflation." *International Journal of Forecasting* 13 (1): 117–26.

Fuhrer, Jeffrey C. 1997. "The (Un)Importance of Forward-Looking Behavior in Price Setting." *Journal of Money, Credit, and Banking* 29 (3): 338–50.

Fuhrer, Jeffrey, and George Moore. 1995. "Inflation Persistence." *Quarterly Journal of Economics* 110 (1): 127–59.

Galí, Jordi, and Mark Gertler. 1999. "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics* 44 (2): 195–222.

Galí, Jordi, Mark Gertler, and J. David López-Salido. 2001. "European Inflation Dynamics." *European Economic Review* 45 (7): 1237–70.

Geweke, John, and Susan Porter-Hudak. 1983. "The Estimation and Application of Long Memory Time Series Models." *Journal of Time Series Analysis* 4:221–38.

Giraitis, Liudas, Piotr Kokoszka, and Remigijus Leipus. 2001. "Testing for Long Memory in the Presence of a General Trend." *Journal of Applied Probability* 38 (4): 1033–54.

Granger, Clive W.J. 1980. "Long Memory Relationships and the Aggregation of Dynamic Models." *Journal of Econometrics* 14 (2): 227–38.

Granger, Clive W.J, and R. Joyeux. 1980. "An Introduction to Long Memory Series." *Journal of Time Series Analysis* 1:15–30.

Hall, Robert E. 1999. Comment on "Rethinking the Role of NAIRU in Monetary Policy: Implications of Model Formulation and Uncertainty," by Arturo Estrella and Frederic S. Mishkin. In *Monetary Policy Rules*, ed. John B. Taylor. Chicago: University of Chicago Press.

Hassler, Uwe, and Jurgen Wolters. 1995. "Long Memory in Inflation Rates: International Evidence." *Journal of Business and Economic Statistics* 13 (1): 37–45.

Haubrich, Joseph G., and Andrew W. Lo. 2001. "The Sources and Nature of Long-Term Memory in Aggregate Output." *Economic Review* (Q II):15–30, Federal Reserve Bank of Cleveland.

Hauser, Michael A., Benedikt M. Pötscher, and Erhard Reschenhofer. 1999. "Measuring Persistence in Aggregate Output: ARMA Models, Fractionally Integrated ARMA Models and Nonparametric Procedures." *Empirical Economics* 24 (2): 243–69.

Henry, Marc, and Paolo Zaffaroni. 2002. "The Long Range Dependence Paradigm for Macroeconomics and Finance." In *Theory and Applications of Long Range Dependence*, ed. Paul Doukhan, George Oppenheim, and Murad S. Taqqu. Boston: Birkhäuser.

Hondroyiannis, George, and Sophia Lazaretou. 2004. "Inflation Persistence during Periods of Structural Change: An Assessment using Greek Data." Working Paper No. 370, European Central Bank.

Hosking, Jonathan R.M. 1981. "Fractional Differencing." *Biometrika* 68 (1): 165–76.

Hsu, Chih-Chiang. 2000. "Long Memory or Structural Change: Testing Method and Empirical Examination." Paper contributed to the Econometric Society World Congress, Seattle, WA.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95 (1): 161–82.

Kim, Jae-Young. 2000. "Detection of Change in Persistence of a Linear Time Series." *Journal of Econometrics* 95 (1): 97–116.

Kim, Chang-Jin, Charles R. Nelson, and Jeremy Piger. 2004. "The Less-Volatile U.S. Economy: A Bayesian Investigation of Timing, Breadth, and Potential Explanations." *Journal of Business and Economic Statistics* 22 (1): 80–93.

Koop, Gary, Eduardo Ley, Jacek Osiewalski, and Mark F.J. Steel. 1997. "Bayesian Analysis of Long Memory and Persistence using ARFIMA Models." *Journal of Econometrics* 76 (1–2): 149–69.

Krämer, Walter, and Philipp Sibbertsen. 2002. "Testing for Structural Change in the Presence of Long Memory." *International Journal of Business and Economics* 1 (3): 235–43.

Künsh, H. 1986. "Discrimination between Deterministic Trends and Long Range Dependence." *Journal of Applied Probability* 23: 1025–30.

Kwiatkowski, Denis, Peter C.B. Phillips, Peter Schmidt, and Yongcheol Shin. 1992. "Testing the Null Hypothesis of

Stationarity against the Alternative of a Unit Root." *Journal of Econometrics* 54 (1–3): 159–78.

Lee, Dongin, and Peter Schmidt. 1996. "On the Power of the KPSS Test of Stationarity against Fractionally-Integrated Alternatives." *Journal of Econometrics* 73 (1): 285–302.

Levin, Andrew, and Jeremy M. Piger. 2003. "Is Inflation Persistence Intrinsic in Industrial Economies?" Working Paper No. 023E, Federal Reserve Bank of St. Louis.

Mandelbrot, Benoit, and James R. Wallis. 1969. "Robustness of the Rescaled Range R/S in the Measurement of Noncyclic Long-Run Statistical Dependence." *Water Resources Research* 5:967–88.

Mayoral, Laura. 2004a. "A New Minimum Distance Estimator for ARFIMA Processes." Mimeo.

———. 2004b. "Is the Observed Persistence Spurious or Real? A Test for Fractional Integration versus Structural Breaks." Mimeo.

———. 2005. "Lagrange Multiplier Tests for Long Memory and Breaks." Mimeo.

Michelacci, Claudio, and Paolo Zaffaroni. 2000. "(Fractional) Betaconvergence." *Journal of Monetary Economics* 45 (1): 129–53.

Mikosch, Thomas, and Catalin Starica. 2004. "Non-stationarities in Financial Time Series, the Long Range Dependence, and the IGARCH Effects." *Review of Economics and Statistic*s 86 (1): 378–90.

Newey, Whitney K., and Kenneth D. West. 1994. "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economics Studies* 61:631–53.

Ng, Serena, and Pierre Perron. 2001. "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power." *Econometrica* 69 (6): 1519–54.

Nunes, Luis C., Chung-Ming Kuan, and Paul Newbold. 1995. "Spurious Break." *Econometric Theory* 11 (4): 736–49.

Ooms, Marius, and Jurgen A. Doornik. 1999. "Inference and Forecasting for Fractional Autoregressive Integrated Moving Average Models; with an Application to US and UK Inflation." Econometric Institute Report 9947/A, Erasmus University Rotterdam.

O'Reilly, Gerard, and Karl Whelan. 2004. "Has Euro-Area Inflation Persistence Changed over Time?" Working Paper No. 335, European Central Bank.

Parke, William R. 1999. "What Is Fractional Integration?" *The Review of Economics and Statistics* 81 (4): 632–38.

Perron, Pierre. 1989. "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis." *Econometrica* 58:1361–1401.

Perron, Pierre, and Xiaokang Zhu. 2005. "Structural Breaks with Deterministic and Stochastic Trends." *Journal of Econometrics* 129 (1–2): 65–119.

Phillips, Peter C.B., and Pierre Perron. 1988. "Testing for a Unit Root in Time Series Regression." *Biometrika* 75 (2): 335–46.

Pivetta, Frederic, and Ricardo Reis. 2004. "The Persistence of Inflation in the United States." Mimeo, Harvard University.

Roberts, John M. 2001. "How Well Does the New Keynesian Sticky-Price Model Fit the Data?" Finance and Economics Discussion Series Paper No. 2001-13, Board of Governors of the Federal Reserve System.

Robinson, Peter M. 1978. "Statistical Inference for a Random Coefficient Autoregressive Model." *Scandinavian Journal of Statistics* 5:163–68.

Rotemberg, Julio. 1987. "The New Keynesian Microfoundations." *Macroeconomics Annual* 2:69–104.

Sargent, Thomas J. 1999. *The Conquest of American Inflation*. Princeton, NJ: Princeton University Press.

Solow, R. 1968. "Recent Controversy on the Theory of Inflation: An Eclectic View." In *Inflation: Its Causes, Consequences, and Control; Proceedings of a Symposium*, ed. Stephen W. Rousseas. New York: New York University.

Sowell, Fallaw. 1992a. "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series." *Journal of Econometrics* 53 (1–3): 165–88.

———. 1992b. "Modeling Long-Run Behavior with the Fractional ARIMA Model." *Journal of Monetary Economics* 29 (2): 277–302.

Stock, J. 2001. "Comment on Evolving Post World War II U.S. Inflation Dynamics." Mimeo, Harvard University.

Taylor, John B. 1979. "Staggered Wage Setting in a Macro Model." *American Economic Review* 69:108–13.

———. 1980. "Aggregate Dynamics and Staggered Contracts." *Journal of Political Economy* 88 (1): 1–23.

———. 1998. "Monetary Policy Guidelines for Unemployment and Inflation Stability." In *Inflation, Unemployment and Monetary Policy*, ed. Robert M. Solow and John B. Taylor. Cambridge, MA: MIT Press.

———. 2000. "Low Inflation, Pass-Through and the Pricing Power of Firms." *European Economic Review* 44 (7): 1389–1408.

Teverosky, V., and M. Taqqu. 1997. "Testing for Long-Range Dependence in the Presence of Shifting Means or a Slowly Declining Trend, Using a Variance-Type Estimator." *Journal of Time Series Analysis* 18 (3): 279–304.

Tobin, J. 1968. "Discussion." In *Inflation: Its Causes, Consequences, and Control; Proceedings of a Symposium*, ed. Stephen W. Rousseas. New York: New York University.

Zaffaroni, Paolo. 2004. "Contemporaneous Aggregation of Linear Dynamic Models in Large Economies." *Journal of Econometrics* 120 (1): 75–102.

The Bank of Japan's Monetary Policy and Bank Risk Premiums in the Money Market*

Naohiko Baba, Motoharu Nakashima, Yosuke Shigemi
Bank of Japan

Kazuo Ueda
University of Tokyo

Using the interest rates on negotiable certificates of deposit issued by individual banks, we first show that under the Bank of Japan's zero interest rate policy and quantitative monetary easing policy, not just the levels of money market rates but also the dispersion of rates across banks have fallen to near zero. We next show that the fall in the dispersion of the rates is not fully explained by a fall in the dispersion of credit ratings of the banks. We also present some evidence on the role of the Bank of Japan's monetary policy in reducing risk premiums.

JEL Codes: E43, E52.

This paper analyzes the effects of the Bank of Japan's (BOJ) monetary policy since the latter half of the 1990s, namely, the so-called zero interest rate policy (ZIRP) and quantitative

*The authors are grateful to Ben Bernanke for his detailed comments and suggestions on an earlier draft of the paper. We also greatly benefited from discussions with Nobu Kiyotaki and Ken Singleton, among others. Any remaining errors are solely our responsibility. Ueda thanks the Center for Advanced Research in Finance of the University of Tokyo for financial assistance. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Bank of Japan. Author contact: Baba: Senior Economist and Director, Financial Markets Department and Institute for Monetary and Economic Studies, Bank of Japan; e-mail: naohiko.baba@boj.or.jp. Nakashima: Economist, Financial Markets Department, Bank of Japan; e-mail: motoharu.nakashima@boj.or.jp. Shigemi: Director, Secretariat of the Policy Board, Bank of Japan; e-mail: yousuke.shigemi@boj.or.jp. Ueda: Professor, University of Tokyo (formerly, Member of the Policy Board, Bank of Japan); e-mail: ueda@e.u-tokyo.ac.jp.

monetary easing policy (QMEP), on credit risk premiums demanded of Japanese banks in the money market.

The ZIRP, the combination of a zero short-term interest rate and a commitment to maintain it until deflationary concerns are dispelled, was adopted by the BOJ between February 1999 and August 2000.[1] In March 2001, the BOJ introduced the QMEP framework whereby the operational target of policy was changed to the current account balances (CABs) held by the financial institutions at the BOJ from the overnight call market rate. At the same time, the BOJ promised to maintain the level of the balances well above required reserves until core CPI inflation becomes above zero on a sustainable basis. The QMEP can be thought of as a version of the ZIRP plus the provision of reserves well in excess of the levels necessary to achieve a zero short-term interest rate.[2]

There is growing literature on the effectiveness of monetary policy near the zero lower bound on interest rates.[3] The literature mostly focuses on the effects of monetary policy on interest rates on safe assets such as government bills and bonds. An often neglected, yet significant aspect of the ZIRP and/or QMEP has been their effects on the credit risk premiums financial institutions pay in the market. That is, the BOJ's policy has lowered such risk premiums to extremely low levels, especially in the money market. As a result, not just the levels but also the dispersion of money market interest rates among banks have been reduced to near zero. Such reductions in risk premiums have been significant in view of a sharp rise in risk premiums during the 1997–98 credit/liquidity crunch, which seriously affected the overall economy.

This paper attempts to document such reductions in the dispersion of risk premiums across banks. In doing so, we look at the market for negotiable certificates of deposit (NCDs) where issuance rates of

[1]More precisely, on February 12, 1999, the BOJ Policy Board determined to encourage the uncollateralized overnight call rate to move as low as possible. In April 1999, the BOJ promised to maintain a zero interest rate "until deflationary concerns are dispelled." The ZIRP was abandoned on August 11, 2000.

[2]The QMEP framework is still in place at the time of the writing of this paper.

[3]See, for example, Baba et al. (2005).

individual banks' NCDs are available on a weekly basis.[4] Using the data, we first show that the standard deviation of the NCD rates among the banks rose sharply toward the financial crisis of 1997–98, but has declined since then. In particular, it declined with the introduction of the ZIRP and declined further as the BOJ intensified its easy policy stance with the QMEP. We then show that the declines in risk premiums cannot be fully explained by recent improvements in the creditworthiness of the banks.

In order to further investigate the background behind the declines in risk premiums on NCD rates, we look at spreads and the dispersion of rates on a wider range of credit instruments. We find that spreads and the dispersion of rates have declined in many areas of financial markets since around 1999, including bonds issued by both banks and nonfinancial corporations. We also find, however, that the decline in spreads on NCD issuance rates has been more significant than spreads on longer-term liabilities. We then carry out a regression analysis to show that the BOJ's monetary policy, especially the commitment to maintain a zero interest rate until deflation ends under the ZIRP and QMEP, has significantly contributed to the decline in the dispersion of rates in the money market. While we do not find evidence in favor of the direct effect of the higher current account balances (CABs) under the QMEP on risk premiums, we find the possibility that longer-dated money market fund-supplying operations have affected risk premiums.

The rest of the paper is organized as follows. In section 1, we present a brief description of the NCD market in Japan. In section 2, we analyze the movements of the standard deviation of the NCD rates over time. In section 3, we look at the relationship between the risk premiums for individual banks and the banks' credit ratings. We find that the relationship has become looser, that is, risk premiums have declined further since the introduction of the ZIRP in 1999. In section 4, we analyze the relationship between the declines in risk premiums and the BOJ's monetary policy. Section 5 concludes the paper.

[4]The BOJ has collected the NCD issuance rates from the domestically licensed banks on a weekly basis and has released the average rates on its web site.

1. The Market for Negotiable Certificates of Deposit (NCDs)

1.1 The Size of the NCD Market

NCDs are debt instruments issued by banks including city, regional, trust, and foreign banks in Japan.[5] NCDs were the first-ever product with deregulated interest rates in Japan and have been issued since May 1979. The amount outstanding of NCDs issued by Japanese domestically licensed banks has been moving around 30 trillion yen in 2004. Of this total, about 80 percent is issued by major banks, namely city and trust banks. Major banks have recently raised around 30 percent of their total funding needs from markets by issuing NCDs. Thus, NCDs can be thought of as one of their principal instruments for raising operating funds.[6]

1.2 NCD Issuance by Maturity

Next, take a look at major banks' issuance of NCDs by maturity. The maturity of NCDs varies from a few weeks to several years. Issuances with maturities of less than 30 days account for about 60 percent of the total based on fiscal 2004 averages.[7] Therefore, market liquidity for NCDs with maturities of less than 30 days is likely to be the highest of all the maturity zones.

2. The Dispersion of Interest Rates on Newly Issued NCDs among Major Banks

Interest rates on major banks' newly issued NCDs had served as a main indicator for deregulated interest rates, although they had moved broadly in tandem across banks for some time since the first NCDs were issued in May 1979. That is, the interest rates had not reflected the differences in bank credit risks. Since the 1990s, however, the interest rates had started to reflect the credit risk of individual issuing banks, mostly due to the rising concern over the stability

[5] For a more detailed description of the Japanese NCD market, see chapter 7 of Totan Research (2002).

[6] For the size of the Japanese NCD market, see appendix figure 1 posted on www.ijcb.org.

[7] For more details, see appendix figure 2 at www.ijcb.org.

of the Japanese financial system.[8] Such concern heightened during the period from late 1997 to 1998. This is shown in figure 1 by substantial jumps in the degree of dispersion as measured by the standard deviation of the NCD interest rates in November 1997.[9] The standard deviations declined significantly, however, after the adoption of the ZIRP in February 1999 and have fallen further following the adoption of the QMEP in March 2001.[10] It is also worth noting that under the QMEP, the standard deviations of the interest rates on newly issued NCDs declined to or even below the levels observed before the period of financial instability.

Table 1 reports the result of statistical tests on the difference in the averages of the standard deviations between four subperiods of the sample: (i) the period before financial instability (up to October 1997), (ii) the period of financial instability between November 1997 and December 1998, (iii) the ZIRP period, and (iv) the QMEP period.[11]

First, the null hypothesis that the averages of the standard deviations were equal was rejected at the 1 percent significance level between the period of financial instability and the ZIRP period. Second, the same null hypothesis was rejected between the ZIRP and QMEP periods. Third, it was also rejected between the prefinancial instability years and the ZIRP period except for the maturity of less than 90 days. Finally, the average is significantly lower during the QMEP period than in prefinancial instability years at all maturities. Thus, we can statistically confirm the following observation:

---

[8] See chapter 7 of Totan Research (2002).

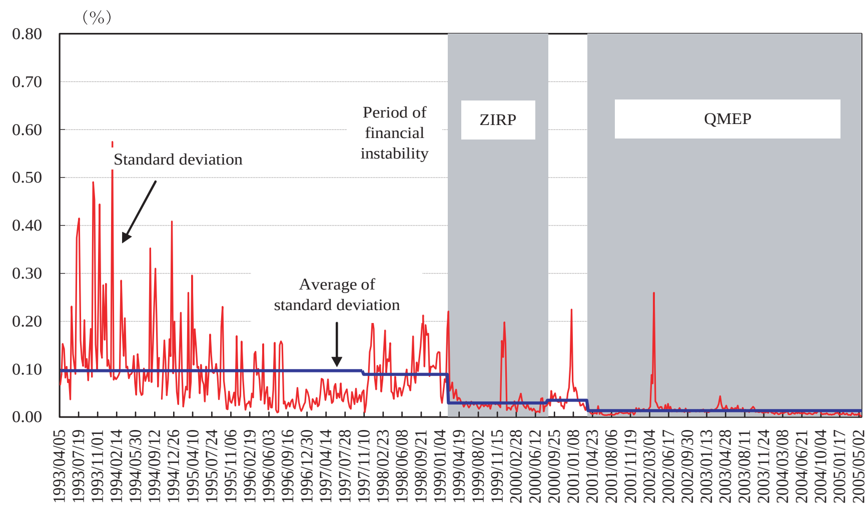
[9] In November 1997, concern over the financial stability heightened following a series of failures of four financial institutions: Sanyo Securities (November 3), Hokkaido Takushoku Bank (November 17), Yamaichi Securities (November 24), and Tokuyo City Bank (November 26). The concern over the financial instability subsided after the nationalization of Long-Term Credit Bank of Japan (October 23, 1998) and Nippon Credit Bank (December 13, 1998).

[10] As shown by appendix figure 3 at www.ijcb.org, a similar tendency is observed in fund-raising costs via deposits, defined as payment of deposit interest rates divided by the amount outstanding of deposits. The standard deviation of the deposit cost lags behind the standard deviation of NCD interest rates by about two years. This is mainly due to longer average maturity of deposits.

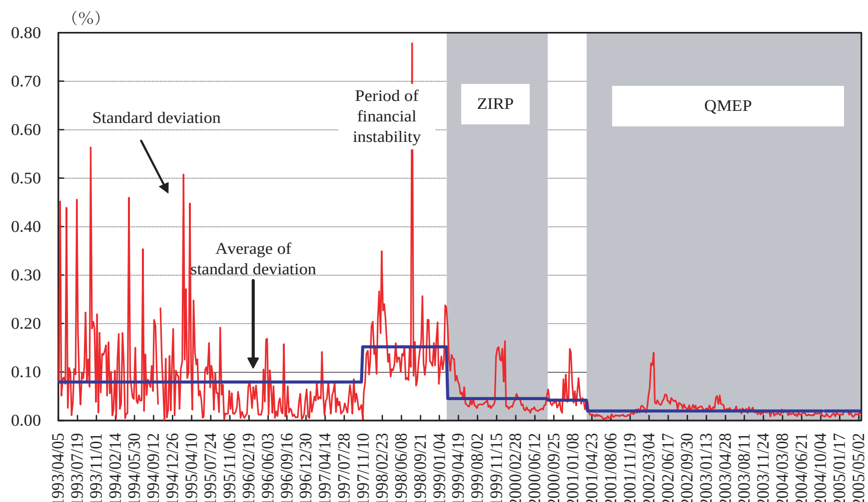
[11] Each period is defined as follows. The period before financial instability is April 5, 1993–October 27, 1997. The period of financial instability is November 3, 1997–December 28, 1998. The ZIRP period is February 15, 1999–August 14, 2000. The QMEP period is March 26, 2001–May 9, 2005.

Figure 1. Standard Deviation of Interest Rates on Newly Issued NCDs among Banks

A. Maturity of Less than 30 Days



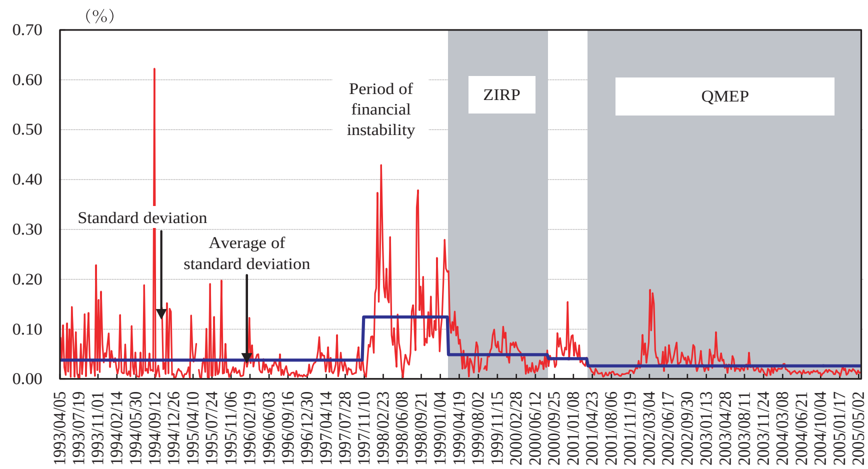
B. Maturity of Less than 60 Days



(*continued*)

Figure 1 (continued). Standard Deviation of Interest Rates on Newly Issued NCDs among Banks

C. Maturity of Less than 90 Days



Notes: Sample banks are the following banks for which weekly data are available throughout the above period (April 5, 1993–May 9, 2005): Sumitomo Mitsui Banking Corporation, the Bank of Tokyo-Mitsubishi, UFJ Bank, Resona Bank, Shinsei Bank, Aozora Bank, the Mitsubishi Trust and Banking Corporation, the Sumitomo Trust & Banking, Mizuho Trust & Banking, UFJ Trust Bank, and the Chuo Mitsui Trust and Banking Company. Fuji Bank and Mizuho Bank are excluded, as a large portion of their NCDs were issued to local governments. For Sumitomo Mitsui Banking Corporation prior to its merger in April 2001, data for the former Sumitomo Bank are used.
We regarded the following dates as "event dates" and those were excluded in calculating the average of standard deviation: (i) the end of 1999 (Y2K problem); (ii) the end of 2000 (preparation for the introduction of RTGS [real-time gross settlement]; (iii) the end of fiscal 2001 (the partial removal of blanket deposit insurance).
Source: Bank of Japan.

the dispersion of NCD issue rates that was very high during the period of financial instability declined after the adoption of the ZIRP in February 1999, and has fallen further since the adoption of the QMEP in March 2001. Also, the levels of dispersion during the ZIRP and QMEP periods have been lower than that in years preceding the financial instability.

Table 1. Test of the Difference in Dispersion of NCD Interest Rates

A. Between the Period of Financial Instability and the ZIRP Period

Maturity	Mean of the Standard Deviation		Difference	t-value
	Financial Instability (A)	ZIRP (B)	C = B − A	Null Hypothesis: C = 0
Less than 30 days	0.089	0.030	−0.060	−8.795***
Less than 60 days	0.152	0.045	−0.107	−8.312***
Less than 90 days	0.124	0.049	−0.075	−5.663***

B. Between the ZIRP and the QMEP Periods

Maturity	Mean of the Standard Deviation		Difference	t-value
	ZIRP (B)	QMEP (D)	E = D − B	Null Hypothesis: E = 0
Less than 30 days	0.030	0.013	−0.016	−5.046***
Less than 60 days	0.045	0.020	−0.025	−6.049***
Less than 90 days	0.049	0.026	−0.023	−4.601***

C. Between the Period Before Financial Instability and the ZIRP Period

Maturity	Mean of the Standard Deviation		Difference	t-value
	Before Financial Instability (F)	ZIRP (B)	G = B − F	Null Hypothesis: G = 0
Less than 30 days	0.097	0.030	−0.067	−11.444***
Less than 60 days	0.079	0.045	−0.034	−4.927***
Less than 90 days	0.038	0.049	−0.011	1.966**

(*continued*)

Table 1 (continued). Test of the Difference in Dispersion of NCD Interest Rates

D. Between the Period Before Financial Instability and the QMEP Period

Maturity	Mean of the Standard Deviation		Difference	t-value
	Before Financial Instability (F)	QMEP (D)	H = D − F	Null Hypothesis: H = 0
Less than 30 days	0.097	0.013	−0.083	−20.1165***
Less than 60 days	0.079	0.020	−0.060	−16.542***
Less than 90 days	0.038	0.026	−0.012	−3.950***

Notes: Each period is defined as follows. The period before financial instability is April 5, 1993–October 27, 1997. The period of financial instability is November 3, 1997–December 28, 1998. The ZIRP period is February 15, 1999–August 14, 2000. The QMEP period is March 26, 2001–May 9, 2005.
The following dates are excluded as event dates: (i) the end of 1999 (Y2K problem); (ii) the end of 2000 (preparation for the introduction of RTGS [real-time gross settlement]; (iii) the end of fiscal 2001 (the partial removal of blanket deposit insurance)).
*** and ** denote the 1 percent and 5 percent significance level, respectively.

3. Estimating Credit Curves from Interest Rates on Newly Issued NCDs

The interpretation of the preceding section's finding, declines in the standard deviation of NCD issuance rates among banks, is not straightforward. One possible interpretation is that financial strains have gradually eased since 1999 and the resultant improvements of the credit ratings for many banks have lowered the standard deviation as well as the levels of credit risk premiums. In order to statistically address this issue, we estimate credit curves at various points in time.

3.1 Estimation Method

First, we define the credit spread for a bank as the interest rate on NCDs issued by the bank with each maturity (less than 30 days, 60 days, and 90 days) minus the weighted average of uncollateralized overnight call rate over all banks.[12] Then, we run cross-sectional time-series regressions of the credit spreads on dummy variables corresponding to sample banks' credit ratings for each of the following four representative years under study: (i) 1997, the year of the financial instability, (ii) 1999, a year when the ZIRP was in full swing, (iii) 2002, one year after the adoption of the QMEP, and (iv) 2004, the last year of our sample period. We also estimate the credit curve in 2005 with a view to following the most recent development. Note, however, that it only covers the period up to May 9, 2005. Our sample consists of city and trust banks. The data on NCD rates are available weekly, resulting in more than 500 observations for almost all cases.[13] We also include end of March, September, and December dummies to control for seasonal market tightness in year-end and annual/semi-annual book-closing months. The credit spreads for each credit rating category, derived from the coefficients on credit

[12]Precisely, the maturity of less than 60 days denotes the maturity of 60 days to 89 days, and the maturity of less than 90 days denotes the maturity of 60 days to 179 days, respectively.

[13]The number of observations for later years is smaller for the following two reasons: (i) there have been mergers among banks; (ii) some banks were not able to issue NCDs in later years, since their credit ratings fell below the investment grades.

rating dummies along with the constant term, map out the "credit curve" for each year.

3.2 Estimated Credit Curves

The estimation result is shown in table 2.[14] Dummy variables for credit ratings are statistically significant in many cases, particularly for the maturity of less than 30 days. Figure 2 draws the credit curves derived by the estimation result. As a general tendency, the credit curves are sloped upward for ratings of A2 or lower for each maturity. The credit curve is sloped downward between A1 and A2 for 1999 in the case of less than 30-day maturity. The number of banks with a rating of A1 or higher, however, is very small for 1999. In fact, the coefficient on A1 dummy is insignificant for 1999. Thus, it seems that we do not have to take this part of the result too seriously.

Figure 2 also demonstrates how the slope of the credit curve became flatter over time. A notable exception is the movement of the spread of Baa2 rating between 1997 and 1999 for each maturity. This coefficient, however, is insignificant even at the 5 percent level in 1999. Except this, it seems fair to say that the credit curves flattened after the introduction of the ZIRP in 1999, flattened further following the introduction of the QMEP in 2002, and almost flattened out in 2004 for all maturities.[15]

The estimation result indicates that the credit risk premiums among major banks are currently close to zero, and that the differences in credit ratings among them are now hardly reflected in their fund-raising costs in the money market. Therefore, the narrowed dispersion of fund-raising costs among banks, shown in figure 1, is

[14]For estimation results for maturities other than less than 30 days, see appendix table 1 at www.ijcb.org.

[15]For the 2004 credit curve, we statistically tested for differences in credit spreads between credit ratings. Although the null hypothesis that the credit risk premiums are the same was rejected between the A2 and Baa2 ratings at the 5 percent significance level for all maturities, the null hypothesis between the A2 and Baa1 ratings was not rejected at the 5 percent significance level for the maturity of less than 30 days. This result shows that the credit curve for this maturity became completely flat between the A2 and Baa1 ratings. We also tested for differences in credit spreads between 2002 and 2004 for the same credit ratings, and found that the null hypothesis that the credit risk premiums are the same was rejected for Baa1, Baa2, and Baa3 for all maturities. This result statistically supports the flattening of the credit curves under the QMEP.

Table 2. Estimation Results of Credit Curves of NCD Spreads

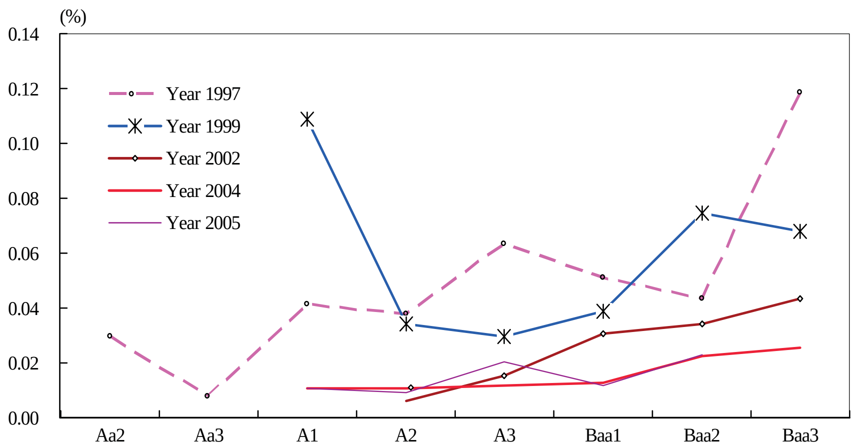
Dependent Variable: NCD Issuance Interest Rate (Less than 30 Days) —Uncollateralized Overnight Call Rate

	1997	**1999**	**2002**	**2004**	**2005**
Constant	−0.075***	0.156***		0.034***	
Aa2	0.105***				
Aa3	0.082***				
A1	0.116***	−0.047		−0.023***	0.011***
A2	0.113***	−0.122***	0.006	−0.023***	0.009***
A3	0.139***	−0.126***	0.015***	−0.022***	0.020***
Baa1	0.126***	−0.117***	0.030***	−0.021***	0.012***
Baa2	0.119***	−0.081*	0.034***	−0.011***	0.023***
Baa3	0.194***	−0.088**	0.043***	−0.008***	
Year-end dummy	0.493***	0.191***	−0.002	−0.000	
Fiscal year-half dummy	0.025**	−0.031**	−0.006	0.004***	
Fiscal year-end dummy	0.030**	0.038***	0.087***	0.003***	0.004***
Adjusted R-squared	0.658	0.253	0.296	0.449	0.197
Number of observations	865	826	667	668	243
Aa2	52	0	0	0	0
Aa3	52	0	0	0	0
A1	80	8	0	47	132
A2	128	78	52	71	19
A3	52	72	202	198	29
Baa1	227	252	109	204	55
Baa2	104	262	155	58	8
Baa3	155	149	149	50	0
No rating	15	5	0	40	0

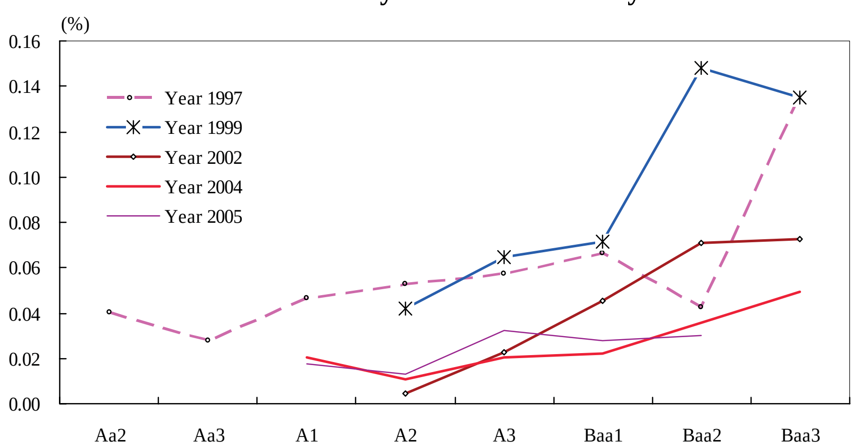
Notes: Estimation is by OLS. ***, **, and * denote the 1, 5, and 10 percent significance level, respectively.
The year 2005 covers the period up to May 9, 2005.
Credit ratings are the long-term ratings of Moody's.

Figure 2. Credit Curves of NCD Spreads

A. Maturity of Less than 30 Days



B. Maturity of Less than 60 Days

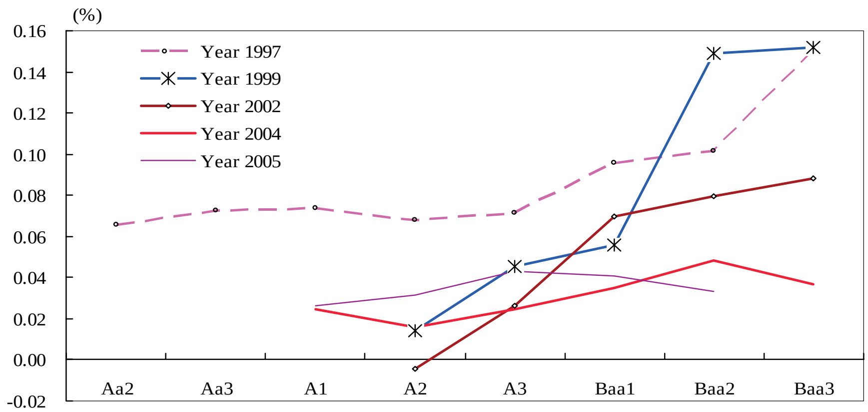


(*continued*)

more likely to be a result of declines in risk premiums across the board in the money market, rather than a result of a lowered dispersion of credit ratings among major banks.

Figure 2 (continued). Credit Curves of NCD Spreads

C. Maturity of Less than 90 Days



Notes: Each curve is drawn from parameter estimates shown in table 2. Credit ratings are the long-term ratings of Moody's.

4. The Flattening of Credit Curves and the BOJ's Monetary Policy

4.1 Three Hypotheses

In this section, we attempt to investigate further the reasons behind the recent flattening of the credit curves for NCD issuance rates, as well as the declines in the dispersion of those rates. We came up with at least the following three possible explanations for this phenomenon.

First, although the analysis in the last section showed that improvements in banks' credit ratings are not the whole explanation, credit ratings may not be the best indicator of credit risks. Sometimes, they are known to lag behind evolving market perceptions of credit risks. With a more proper measure of credit, credit curves may not have flattened.

Second, the BOJ's monetary policy may have exerted non-negligible effects on risk premiums. There are several possibilities here. Easy monetary policy in general, through lower interest rates, raises asset prices and lowers risk premiums. In addition, the BOJ's increasingly strong commitment to maintain a zero interest rate until deflation ends may have played a role. Under the ZIRP (February 1999–August 2000), the BOJ committed itself to maintain a zero short-term interest rate until "deflationary concerns were dispelled." Under the QMEP (March 2001–present), the BOJ promised to maintain the level of the CABs well in excess of required reserves, thus a zero short-term interest rate until the actual and expected core CPI inflation turns positive. In October 2003, the exit condition was further clarified to read "until at least actual core CPI inflation and its forecast by monetary policy board members exceed zero." These commitments may have lowered risk premiums in the money market by substantially reducing the risk that banks fail to meet payment obligations, which makes the near-term chance of a default smaller. Furthermore, the BOJ's attempt to supply huge amounts of excess reserves well above the levels necessary to keep short-term interest rates around zero percent may have played a role. In order to abide by the target on the CABs of 30–35 trillion yen, the BOJ has had to increasingly offer long-dated fund-supplying operations in the money market.[16] The average maturity of bill-purchasing operations was two to three months at the start of the QMEP, but it rose to close to nine months recently. Any banks eligible for the BOJ's money market operations can take such long-term funds from the BOJ with almost zero interest rates. Arbitrage activities across the money markets may have lowered the level and the dispersion of rates even on instruments that are not directly used in the BOJ's money market operations—NCD rates, for example.

The third possible explanation for the decline in the dispersion of NCD issuance rates is that it is partially irrational. In the environment of easy monetary policy and low returns, investors may have carried out "reach for yield" activities aggressively, buying assets with returns too low to be justified by rational economic calculation. While rigorously distinguishing between these hypotheses is beyond the scope of the present paper, in what follows we offer several pieces of evidence that we think are helpful in speculating on the importance of each.

---

[16] At the time when the QMEP was adopted in March 2001, the target on the CABs was 5 trillion yen. It was raised several times and reached 30–35 trillion yen in January 2004.

4.2 Some Evidence

We first take a look at the evolution of credit curves and the dispersion of rates for a wide range of financial instruments. This exercise will reveal that, while factors common to many instruments have been at work, there is something distinct about the decline in risk premiums in the money market. We then proceed to carry out a more formal analysis concerning whether or not this distinct movement of money market credit spreads is related to monetary policy, and, if so, in what ways.

4.2.1 Credit Curves for Other Financial Instruments

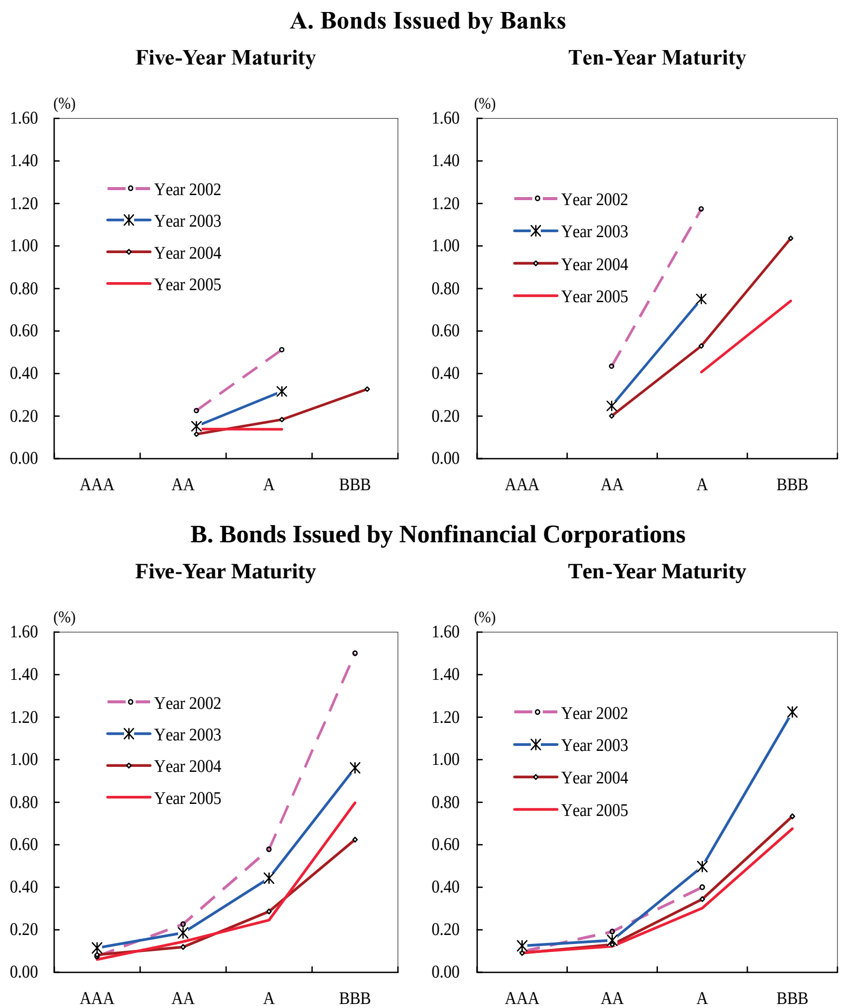
Figure 3 shows credit curves of bond spreads with maturities of five and ten years for banks and nonfinancial corporations, respectively, which we estimated using the same methodology as in the case of NCD credit curves.[17] We can see that credit curves have become flatter over time as in the case of NCD issuance rates, but that risk premiums seem to remain other than a few cases for 2004 and 2005: (i) the five-year spread for banks between the ratings of A and AA and (ii) five- and ten-year spreads for nonfinancial corporations between AA and AAA.[18] In general, the curves are flatter for five-year maturity than ten-year maturity and for banks than for nonfinancial corporations. Thus, the tendency for spreads to decline does exist even in the long-term bond market for both banks and nonfinancial corporations, but it is stronger for relatively shorter-dated bonds and for banks.

Regarding Japanese corporate bond pricing, Nishioka and Baba (2004) showed that narrowed credit spreads on Japanese corporate bonds under the ZIRP and QMEP cannot be explained unless they assume a risk-loving representative investor. This leads to an underpricing of negative-skewness risk that is inherent in defaultable bonds. Thus, together with our observation that the curves are flatter

[17]The bond spread is defined as the bond issuance rate minus the Japanese government bond (JGB) yield with the same maturity. Credit curves are derived by estimating credit rating dummies after controlling for year-end, semi-annual, and fiscal year-end dummies. The bond yield data is available only from 2002.

[18]The bond issuance rates for the year 2005 cover the period up to the end of May.

Figure 3. Credit Curves of Bond Spreads



Notes: The bond spread is defined as the spread of bond issuance rate over the JGB yields with the same maturity.
Credit curves are derived by estimating credit rating dummies after controlling for year-end, semi-annual, and fiscal year-end dummies.
Credit ratings are the long-term ratings of Moody's.
Number of observations is as follows: (i) bank bonds: 12/4 (five-year/ten-year) for 2002, 13/7 for 2003, 12/12 for 2004, and 6/6 for 2005; (ii) nonfinancial corporate bonds: 66/51 for 2002, 85/67 for 2003, 82/52 for 2004, and 21/26 for 2005, respectively.
Sources: Bloomberg, IN database.

for banks than for nonfinancial corporations, the third hypothesis above, excessive risk taking in the money and bond markets, particularly for banks, is likely to hold at least to some extent.

Next, figure 4 presents credit curves of commercial paper (CP) issuance spreads with one- and three-month maturities.[19] As in the case with bond spreads, the curves have become flatter over time. There are, however, significant spreads remaining at ratings of below a-1. In particular, note that the difference in CP spreads between a-2 and a-1 in 2004 amounts to ten times as large as the largest one-notch difference in spreads for NCD issuance rates. This difference in the slope of credit curves is interesting. It should partially be explained by differences in credit risks perceived by investors: banks versus nonfinancial corporations. It is also interesting to look at the tight credit spread between a-1 and a-1+. The BOJ has carried out fund-supplying operations using CP, albeit in a repurchasing form, and market participants seem to recognize that most of the CP eligible for the money market operations has the rating of a-1 or higher.[20] Consequently, the very flat CP credit curve over the zone of a-1 or higher is suggestive of the direct effect of the BOJ's operations on CP rates.

4.2.2 Dispersion of Interest Rates for Other Financial Instruments

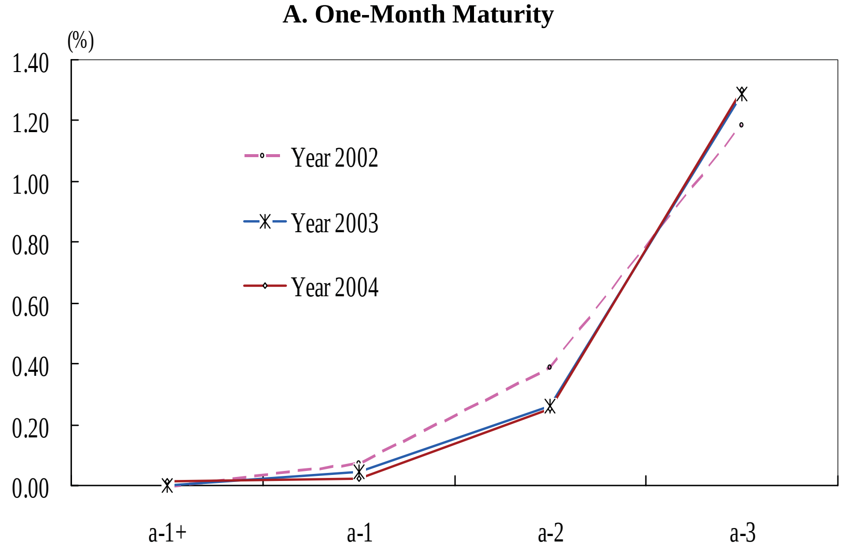
It is also of interest to compare the dispersion of NCD issuance rates with that of other financial instruments. Figure 5 shows the spreads on bank-issued bond yields over the JGB yields and the standard deviation of those spreads across the banks. It is similar to the case of NCD issuance rates in that the standard deviation has declined sharply since the beginning of 2003. The dispersion of bank bond spreads, however, rose significantly in late 2001 and stayed high

[19] The CP spread is defined as the CP issuance rate minus the uncollateralized overnight call rate. The CP issuance rates are available only from 2002. Credit curves are derived by estimating credit rating dummies after controlling for year-end, semi-annual, and fiscal year-end dummies.

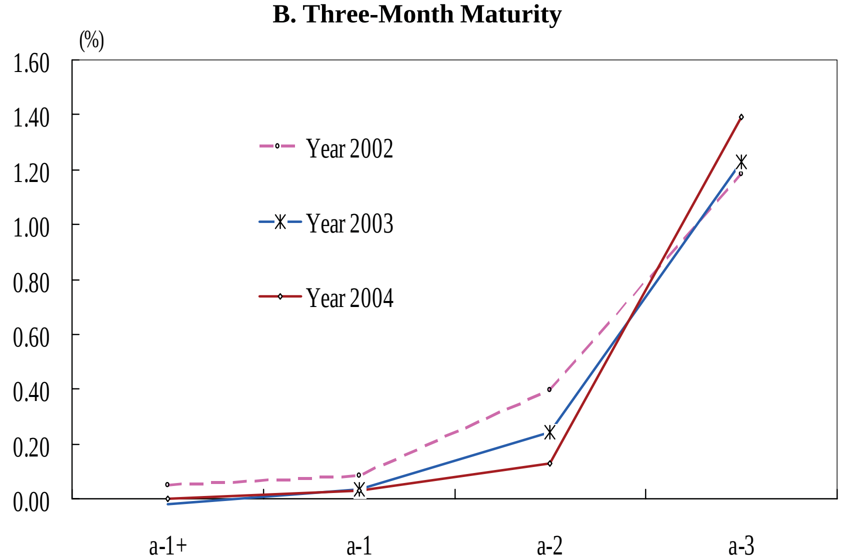
[20] The BOJ officially states the following as eligibility standards for CP: (i) those deemed appropriate in light of relevant conditions including the creditworthiness of an obligor, and (ii) those with an original maturity of up to one year.

Figure 4. Credit Curves of CP Spreads

A. One-Month Maturity

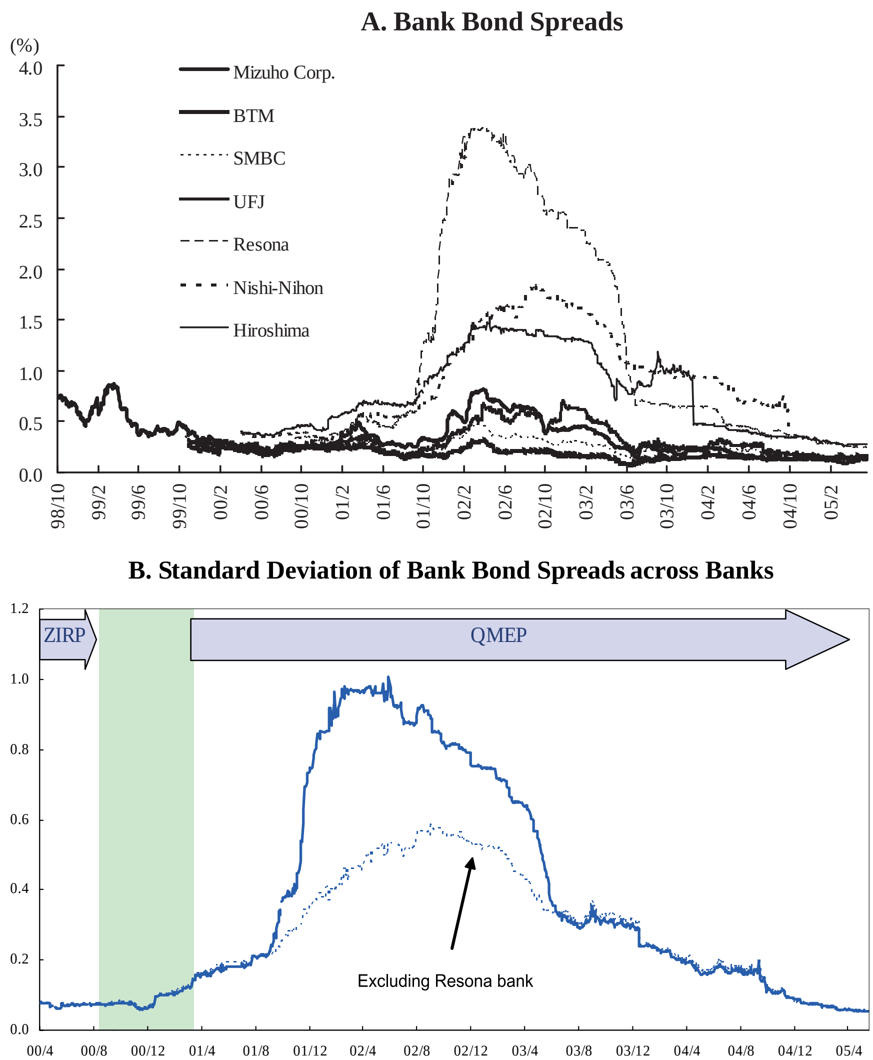


B. Three-Month Maturity



Notes: The CP spread is defined as the spread of CP issuance rate over the uncollateralized overnight call rate.
Credit curves are derived by estimating credit rating dummies after controlling for year-end, semi-annual, and fiscal year-end dummies.
Credit ratings are the short-term ratings of Moody's.
Number of observations is 2,327 for 2002; 1,975 for 2003; and 2,006 for 2004, respectively.
Sources: Finance Facsimile News, Bank of Japan.

Figure 5. Dispersion of Bank Bond Spreads



Notes: The bank bond spread is defined as the spread of bank bond yield over the JGB yield with the same maturity. The maturity of most bonds is five years.
We computed the standard deviation of bank bond yields from the date from which more than four banks' yield data are available.
Source: Japan Securities Dealers Association.

until early 2003. There are some corresponding movements in the dispersion of NCD issuance rates, but these are limited to spikes of short duration in March 2002 and March 2003.[21] The period between late 2001 and early 2003 corresponds to a recession following the collapse of the so-called IT bubble and saw many bankruptcies of both financial and nonfinancial corporations. The bankruptcy of Enron Corporation and the resultant worldwide concerns over mutual funds added to the stresses. The minutes of the BOJ's monetary policy meetings in late 2001 and early 2002 reveal that the policy board was very concerned about the rise in risk premiums in the money and bond markets. In response, the policy board decided to raise the target on CABs in December 2001 and also allowed the CABs to go above the target range temporarily in the spring of 2002. The minutes after the adoption of these measures indicate that the board thought that the measures were successful in containing the risk premiums in the money market, but not those on bonds for banks and nonfinancial corporations. This episode is again indicative of different impacts of monetary policy on money market instruments from those on others.

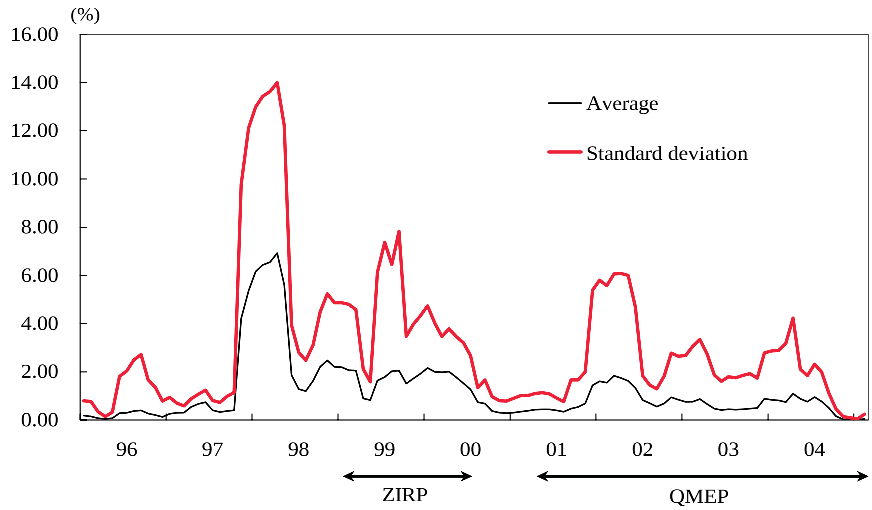
Figure 6 presents the movement of the standard deviation of another measure of bank risk, that is, the bank default probabilities implied by bank stock prices.[22] Again, we can see the general downward trend in the dispersion. The dispersion in default probabilities, however, rose between late 2001 and the middle of 2002, declined somewhat after that, but still remained at fairly high levels until the middle of 2004. This pattern is roughly the same as for bank bond yields and is not quite similar to that for NCD issuance rates.

The foregoing discussion suggests that both the levels and dispersion of interest rates on credit instruments, particularly those issued

[21] The other spikes in the dispersion of NCD issuance rates in late 1999 and late 2000 reflect the stress associated with the Y2K problem and the introduction of the real-time gross settlement (RTGS) scheme into the settlement of funds and government bonds in early 2001.

[22] We used the model of Merton (1974), under which default occurs at the maturity date of debt in the event that the issuer's assets are less than the face value of the debt. We calculated each bank's default probability on a monthly basis using daily stock returns and standard deviations derived from the past half year's returns, together with the balance sheet data. We obtained the data from Bloomberg. The time horizon for default is assumed to be one year.

Figure 6. Default Probability Implied by Stock Price



Notes: We used the model of Merton (1974) to compute the default probability for each bank. Time horizon is assumed to be one year. We calculated each bank's default probability on a monthly basis from daily stock returns, together with the balance sheet data.
The number of sample banks is 102, which include city, trust, regional, and regional II banks.
Sources: Bloomberg, Bank of Japan.

by banks, have declined, as is the case with the NCD issuance rates. The general environment of easy monetary policy probably played some role. There may have been an element of irrational pursuit of yields. The discussion, however, also suggests that the decline in the spreads or the dispersion of rates for NCDs has been more significant than those for other instruments. The rise in spreads seen in bond rates for the period of late 2001–early 2003 is limited to very temporary spikes in the case of NCDs. Bank bond yields and/or bank default probabilities estimated from stock prices may be a better indicator of bank credit risks than credit ratings. The different behavior of the dispersion of interest rates or default probabilities between NCDs and others, however, seems to suggest that problems of credit ratings as a measure of credit risk, the first hypothesis in section 4.1 above, are not the whole explanation of our finding of flat credit curves for NCD rates.[23]

It would be best if we can determine the exact contribution of monetary policy developments, irrational investor behavior, and other factors behind the declines in spreads and the dispersion of rates for a wide range of instruments such as bonds for banks and nonfinancial corporations. This, however, would go well beyond the scope of the present paper. Instead, we focus on the analysis of money market rates, to which we now turn.

4.2.3 Regression Analysis on the Role of the BOJ's Monetary Policy

In what follows, we attempt to investigate the role the BOJ's monetary policy has played in the behavior of credit spreads for NCD issuance rates. To that end, we extend the year-by-year regression analysis on the credit curves of individual NCD issuance spreads by pooling the entire cross-sectional time-series data and allowing the slope of the credit curves to depend on the variables, including those related to the BOJ's monetary policy.

Specifically, we estimate the following model:

$$\begin{aligned} NCD_{it} &= (a_0 + a_1 ZIRP + a_2 QMEP + a_3 TRANS \\ &+ a_4 CAB_t + a_5 BOND_{it})^*(\mathrm{A1}) \\ &+ (b_0 + b_1 ZIRP + b_2 QMEP + b_3 TRANS \\ &+ b_4 CAB_t + b_5 BOND_{it})^*(\mathrm{A2}), \\ &+ \ldots \end{aligned}$$

where NCD_{it} denotes the spread for NCD issuance rate for bank i at time t over the weighted average of uncollateralized overnight call rate, and $BOND_{it}$ denotes the spread of the bond yield for bank i at time t over the JGB yield with the same maturity. A1 (A2 ..) denotes the dummy variable that takes 1 if the credit rating for bank i at time t is A1 (A2 ..) and takes 0 otherwise.[24] More importantly,

[23] We also analyzed the relationship between the NCD spreads and other measures of credit risk such as interest coverage ratio, defined as the ratio of interest payment to earnings, and ROA (return on assets), but could not get any robust results.

[24] We also included seasonable dummies as in the estimation of credit curves in section 3.

we include monetary policy-related variables to investigate the relationship between the flattening of NCD credit curves and the BOJ's monetary policy. They are dummy variables corresponding to the BOJ's commitments during the ZIRP and QMEP periods and the level of aggregate CABs. That is,

ZIRP: takes on 1 when the ZIRP is in force (February 12, 1999–August 11, 2000) and 0 otherwise.

QMEP: takes on 1 when the QMEP is in force (March 19, 2001–present) and 0 otherwise.

TRANS: takes on 1 after October 10, 2003, when the BOJ enhanced the transparency of monetary policy and 0 otherwise.

CAB_t: aggregate current account balances at time t.

Those interactive terms with credit rating dummies are an attempt to estimate whether or not, and to what extent, monetary policy variables have contributed to the flattening of the credit curves that we saw in figure 2.[25]

The inclusion of bank bond spreads in the credit rating dummy coefficients is an attempt to allow for the possibility that credit ratings are inadequate measures of bank credit. Thus, it is an attempt to address the first hypothesis for the decline in rate dispersion as put forward at the beginning of this section. To the extent that bank bond spreads reflect underlying bank risks more appropriately, declines in spreads at each rating due to such mismeasurement should be captured by the bond spread terms.[26] The equation is estimated for seven banks for which the bond yield data are available.[27] The data frequency is weekly and the sample period is from October 5, 1998, to May 9, 2005.

[25]To the extent that bank bond yields have responded to monetary policy, we are underestimating the effects of monetary policy on NCD issuance spreads.

[26]We also estimated the equation including the bank bond spreads as one independent variable, not as an interactive term with credit rating dummies. The results were essentially the same.

[27]Those banks are the Mizuho Corporation Bank, Shinsei Bank, Aozora Bank, the Bank of Tokyo-Mitsubishi, Sumitomo Mitsui Banking Corporation, UFJ Bank, and Risona Bank.

Table 3 reports estimation results.[28] The results show that, even after controlling for bank bond spreads, monetary policy has significantly contributed to the declines in risk premiums in the NCD market. Specifically, the ZIRP and QMEP dummies are significant with the right (negative) sign in most cases. That is, the commitments to maintain a zero interest rate have contributed to the decline in the NCD credit spreads. Also, the coefficients on bond spreads are significantly positive at credit ratings lower than Baa1 for each maturity. This result suggests that the slope of credit curves is likely to be significantly flatter at those ratings when bond spreads continue to decline like the period under the QMEP.

Figure 7 graphically shows the effects of each commitment on the credit curve. The effects of the first two commitments are larger at lower ratings. For higher ratings, the effects of the QMEP commitment are slightly larger than those of the ZIRP commitment. Thus, the ZIRP and QMEP commitments have flattened the credit curves for, and lowered the dispersion of, NCD issuance rates by mainly reducing risk premiums for banks with relatively low ratings.[29]

In contrast, the variable *CAB* is either insignificant or significant with a wrong (positive) sign. We tried several variations of the equation reported above, finding essentially the same result. That is, there is no evidence that higher levels of CABs have reduced risk premiums in the money market over and above the effect of the QMEP dummy.[30]

Regarding the above result, one may think that what is important is not quite the level of the CABs per se, but the level relative to ex ante demand for liquidity. In fact, during the period under study, we experienced significant fluctuations in the banks' demand for liquidity mainly due to a changing perception about the health of the banking system. In order to take account of this possibility,

[28] For estimation results for maturities other than less than 30 days, see appendix table 2 at www.ijcb.org.

[29] The near absence of the significant effects of the third commitment is not easy to interpret. A casual observation suggests that it stabilized JGB yields after a spike in the summer of 2003. One possibility is that it may have influenced longer-term yields more than money market rates by its clarification of the exit conditions.

[30] Just as a robustness check on the significance of the commitment dummies, we estimated the above equation without including the CABs and did not find any significant differences in the results.

Table 3. Regression Results on the Role of the BOJ's Monetary Policy: Part I

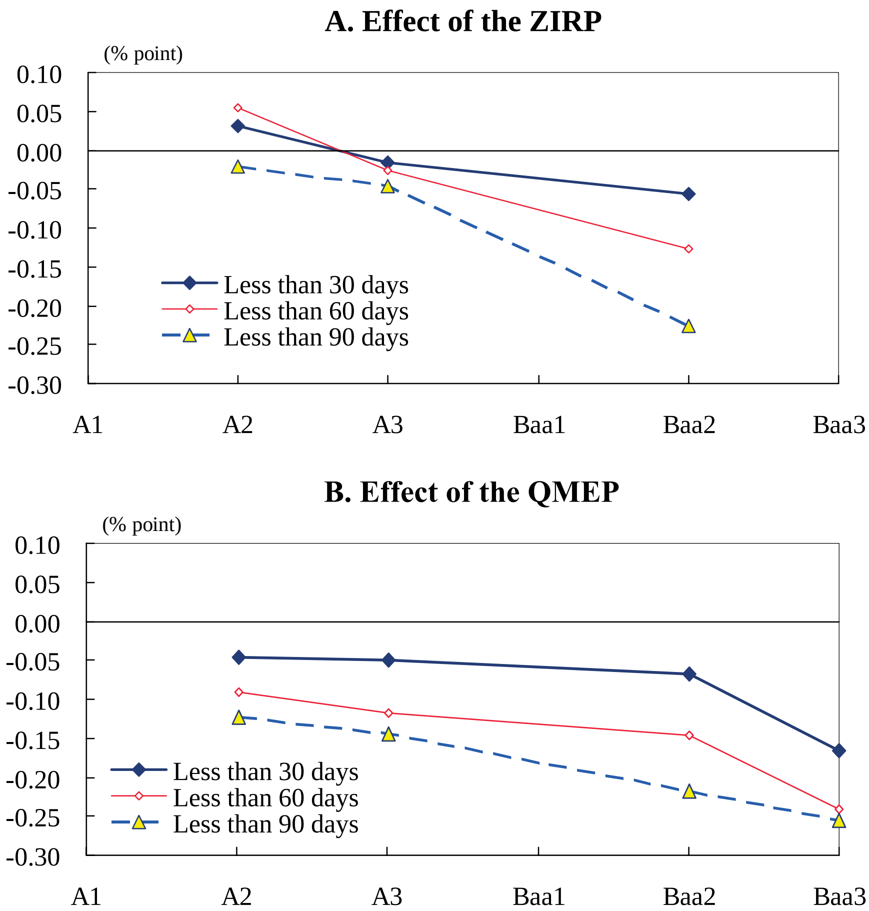
Dependent Variable: NCD Issuance Interest Rate (Less than 30 Days) —Uncollateralized Overnight Call Rate

Number of Observations: 1,929 (October 5, 1998–May 9, 2005)

Variable	Coefficient	Standard Error
A1	0.056	0.204
A1*CAB	−0.001	0.006
A1*BOND	−0.051	0.261
A2	0.030	0.028
A2*ZIRP	0.032**	0.014
A2*QMEP	−0.047***	0.013
A2*TRANS	−0.014	0.012
A2*CAB	0.001*	0.000
A2*BOND	0.021	0.100
A3	0.047***	0.009
A3*ZIRP	−0.016*	0.009
A3*QMEP	−0.051***	0.009
A3*TRANS	−0.021**	0.008
A3*CAB	0.001***	0.000
A3*BOND	0.002	0.018
Baa1	−0.127	0.159
Baa1*CAB	0.004	0.005
Baa1*BOND	0.047	0.106
Baa2	0.060***	0.008
Baa2*ZIRP	−0.057***	0.009
Baa2*QMEP	−0.068***	0.009
Baa2*TRANS	−0.025**	0.010
Baa2*CAB	0.001***	0.000
Baa2*BOND	0.014***	0.002
Baa3	0.107***	0.003
Baa3*QMEP	−0.166***	0.012
Baa3*TRANS	0.003	0.017
Baa3*CAB	0.002**	0.001
Baa3*BOND	0.039***	0.003
Year-end dummy	0.045***	0.004
Fiscal year-half dummy	−0.004	0.005
Fiscal year-end dummy	0.020***	0.004
Adjusted R-squared	0.388	

Notes: Estimation is by OLS. ***, **, and * denote the 1, 5, and 10 percent significance level, respectively.
Credit ratings are the long-term ratings of Moody's.

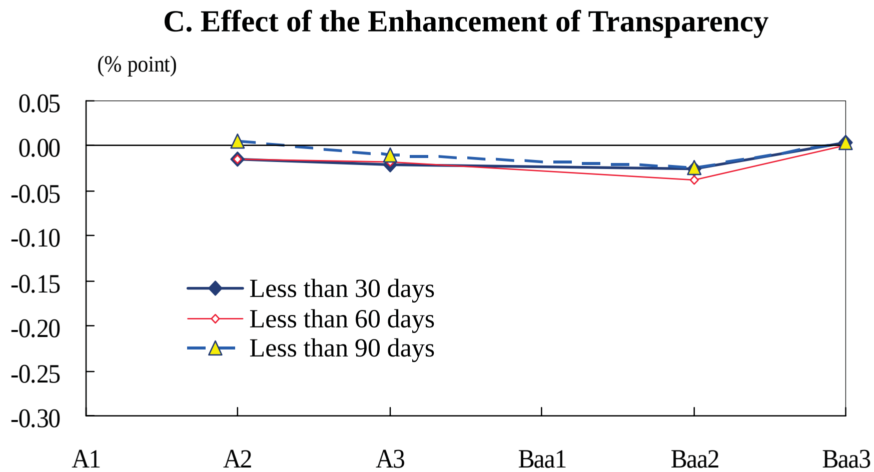
Figure 7. Effects of the BOJ's Monetary Policy on NCD Credit Curves



(*continued*)

we reran the regression using the variable *TERM* in place of *CAB*. *TERM* represents the monthly average of the maturity of the BOJ's bill-purchasing operations. At times of low demand for liquidity, the BOJ had to offer longer-dated operations to meet the target on the CABs. In this sense, the variable may be regarded as a proxy for an ex ante "excess supply" of liquidity in the money market. As

Figure 7 (continued). Effects of the BOJ's Monetary Policy on NCD Credit Curves



Notes: Each curve is drawn using parameter estimates reported in table 3.
Credit ratings are the long-term ratings of Moody's.

shown in table 4, *TERM* in fact comes in negatively and is significant at many credit ratings below A1 for each maturity.[31] The rest of the estimation results, *BOND* coefficients, are similar to the results reported in table 3.[32] Thus, we cannot deny the possibility that increasingly longer-dated operations in the money market have lowered risk premiums.

To summarize, monetary policy, particularly the commitments to maintain a zero interest rate until deflation ends under the ZIRP and QMEP, has contributed to the decline in the dispersion of NCD issuance rates. The effect of the quantitative easing aspect of the QMEP on credit spreads, CABs well in excess of the levels

[31] For estimation results of maturities other than less than 30 days, see appendix table 3 at www.ijcb.org.

[32] Since *TERM* is only available after January 2001, we estimated the equation using data since then. We excluded the *QMEP* and *TRANS* dummies from the equation since the estimation period almost coincides with the QMEP period. Also, note that the *ZIRP* dummy is irrelevant in this period. *TERM* is available on the BOJ web site.

Table 4. Regression Results on the Role of the BOJ's Monetary Policy: Part II

Dependent Variable: NCD Issuance Interest Rate (Less than 30 Days) —Uncollateralized Overnight Call Rate

Number of Observations: 1,515 (January 4, 2001–May 9, 2005)

Variable	Coefficient	Standard Error
A1	−0.059	0.042
A1*TERM	0.010	0.007
A1*BOND	0.062	0.143
A2	0.018	0.014
A2*TERM	−0.003	0.002
A2*BOND	0.019	0.047
A3	0.028***	0.006
A3*TERM	−0.005***	0.001
A3*BOND	0.015*	0.009
Baa1	0.038	0.024
Baa1*TERM	−0.006*	0.003
Baa1*BOND	−0.020	0.061
Baa2	0.040***	0.007
Baa2*TERM	−0.007***	0.002
Baa2*BOND	0.013***	0.003
Baa3	0.048***	0.009
Baa3*TERM	−0.060**	0.002
Baa3*BOND	0.008***	0.002
Year-end dummy	0.002	0.002
Fiscal year-half dummy	−0.000	0.002
Fiscal year-end dummy	0.015***	0.002
Adjusted R-squared	0.167	

Notes: Estimation is by OLS. ***, **, and * denote the 1, 5, and 10 percent significance level, respectively.
Credit ratings are the long-term ratings of Moody's.

necessary to keep a short-term interest rate zero, was less clear. We do find, however, some evidence that the particular types of operations that the BOJ carried out—that is, longer-dated operations in the money market—have exerted the effect of lowering risk premiums in the money market. The informal discussion of the spreads on CP offered in this section also accords well with such a finding.

It is important to note that we have not attempted to estimate the effects of the BOJ's monetary policy on a wider range of instruments. The commitments of the maintenance of a zero interest rate may have had significant effects on longer-term interest rates.[33] Quantitative easing and/or targeted asset purchases may have also affected asset prices other than money market rates.[34]

5. Concluding Remarks

We have shown that not just the levels of money market rates but also their dispersion have declined since 1999. We have documented this in detail for NCD issuance rates. In particular, the decline in rate dispersion cannot be fully accounted for by improvements in bank credit. That is, risk premiums have declined sharply across the board in the money market. We have found a similar tendency for a decline in spreads for longer-dated bank liabilities and for bonds issued by nonfinancial corporations. Many factors, including monetary policy, probably played a role behind the declines in risk premiums for such a wide range of instruments.

We have provided evidence, however, for a stronger tendency for risk premiums on NCD issuance rates to decline than for other longer-maturity instruments, as well as for instruments issued by nonfinancial corporations. We have found that the BOJ's monetary policy has played a role here. In particular, the commitments to maintain a zero interest rate until deflationary pressure ends both under the ZIRP and QMEP have significantly contributed to the declines in the spreads. We have not found a similar effect from increases in the CABs, but have identified the possibility that some particular operations that the BOJ carried out to increase the supply of liquidity—for example, longer-dated money market operations—have lowered the spreads.

[33]Bernanke, Reinhart, and Sack (2004) and Oda and Ueda (2005) present evidence consistent with such a view.

[34]Bernanke, Reinhart, and Sack (2004) find significant links between the BOJ's JGB purchases and JGB yields and between quantitative easing and stock prices.

References

Baba, Naohiko, Shinichi Nishioka, Nobuyuki Oda, Masaaki Shirakawa, Kazuo Ueda, and Hiroshi Ugai. 2005. "Japan's Deflation, Problems in the Financial System and Monetary Policy." *Monetary and Economic Studies* 23:47–111.

Bernanke, Ben S., Vincent R. Reinhart, and Brian P. Sack. 2004. "Monetary Policy Alternatives at the Zero Bound: An Empirical Assessment." *Brookings Papers on Economic Activity* 2:1–100.

Merton, Robert C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29 (2): 449–70.

Nishioka, Shinichi, and Naohiko Baba. 2004. "Credit Risk Taking by Japanese Investors: Is Skewness Risk Priced in Japanese Corporate Bond Market?" Working Paper No. 04-E-7, Bank of Japan.

Oda, Nobuyuki, and Kazuo Ueda. 2005. "The Effects of the Bank of Japan's Zero Interest Rate Commitment and Quantitative Monetary Easing on the Yield Curve: A Macro-Finance Approach." Working Paper No. 05-E-6, Bank of Japan.

Totan Research. 2002. *Shin Tokyo Money Market* (in Japanese).

Appendix Table 1. Estimation Results of Credit Curves for NCD Issuance Rates

A. Maturity of Less than 60 Days

Dependent Variable: NCD Issuance Interest Rate

—Uncollateralized Overnight Call Rate

	1997	**1999**	**2002**	**2004**	**2005**
Constant	−0.043	0.237***		0.044***	
Aa2	0.083				
Aa3	0.071				
A1	0.089			−0.023***	0.018***
A2	0.095	−0.195***	0.005	−0.033***	0.013***
A3	0.100	−0.173***	0.023***	−0.023***	0.033***
Baa1	0.109	−0.166***	0.045***	−0.022***	0.028***
Baa2	0.085	−0.089*	0.071***	−0.008**	0.030***
Baa3	0.179	−0.102**	0.072***	0.006	
Year-end dummy	0.298***	0.278***	−0.002	−0.000	
Fiscal year-half dummy	0.026*	−0.064***	−0.005	0.004***	
Fiscal year-end dummy	0.012	0.147***	0.088***	0.010***	0.006***
Adjusted R-squared	0.371	0.453	0.401	0.554	0.213
Number of observations	723	710	665	640	236
Aa2	52	0	0	0	0
Aa3	48	0	0	0	0
A1	79	8	0	47	133
A2	128	73	52	69	19
A3	52	71	207	193	28
Baa1	205	243	106	202	48
Baa2	56	216	155	56	8
Baa3	102	99	145	44	0
No rating	1	0	0	29	0

Notes: Estimation is by OLS. ***, **, and * denote the 1, 5, and 10 percent significance level, respectively.

The year 2005 covers the period up to May 9, 2005.

Credit ratings are the long-term ratings of Moody's.

Appendix Table 1 (continued). Estimation Results of Credit Curves for NCD Issuance Rates

B. Maturity of Less than 90 Days

Dependent Variable: NCD Issuance Interest Rate
—Uncollateralized Overnight Call Rate

	1997	**1999**	**2002**	**2004**	**2005**
Constant	-0.060	0.093^{**}		0.045^{***}	
Aa2	0.125				
Aa3	0.132				
A1	0.134			-0.021^{***}	0.026^{***}
A2	0.128	-0.079^{*}	-0.004	-0.030^{***}	0.031^{***}
A3	0.131	-0.048	0.027^{***}	-0.021^{***}	0.043^{***}
Baa1	0.155	-0.038	0.069^{***}	-0.011^{**}	0.041^{***}
Baa2	0.161	0.056	0.079^{***}	0.003	0.034^{***}
Baa3	0.210^{*}	0.059	0.088^{***}	-0.009^{*}	
Year-end dummy	0.129^{***}	0.202^{***}	-0.002	0.004^{*}	
Fiscal year-half dummy	0.021	-0.028^{*}	-0.001	0.006^{***}	
Fiscal year-end dummy	-0.030^{**}	0.246^{***}	0.070^{***}	0.009^{***}	0.002
Adjusted R-squared	0.194	0.554	0.332	0.321	0.164
Number of observations	588	511	523	497	193
Aa2	52	0	0	0	0
Aa3	50	0	0	0	0
A1	74	8	0	44	126
A2	119	64	40	54	18
A3	50	70	202	177	19
Baa1	160	198	81	151	24
Baa2	31	125	122	40	6
Baa3	51	46	78	20	0
No rating	1	0	0	11	0

Notes: Estimation is by OLS. ***, **, and * denote the 1, 5, and 10 percent significance level, respectively.
The year 2005 covers the period up to May 9, 2005.
Credit ratings are the long-term ratings of Moody's.

Appendix Table 2. Regression Results on the Role of the BOJ's Monetary Policy: Part I

A. Maturity of Less than 60 Days

Dependent Variable: NCD Issuance Interest Rate

—Uncollateralized Overnight Call Rate

Number of Observations: 1,890 (October 5, 1998–May 9, 2005)

Variable	**Coefficient**	**Standard Error**
A1	−0.022	0.295
A1*CAB	0.001	0.009
A1*BOND	−0.134	0.378
A2	0.077*	0.042
A2*ZIRP	0.055**	0.021
A2*QMEP	−0.092***	0.020
A2*TRANS	−0.014	0.018
A2*CAB	0.001	0.001
A2*BOND	0.020	0.149
A3	0.116***	0.013
A3*ZIRP	−0.026*	0.013
A3*QMEP	−0.117***	0.013
A3*TRANS	−0.018	0.011
A3*CAB	0.001	0.001
A3*BOND	0.015	0.027
Baa1	−0.226	0.230
Baa1*CAB	0.006	0.007
Baa1*BOND	0.142	0.156
Baa2	0.125***	0.011
Baa2*ZIRP	−0.126***	0.014
Baa2*QMEP	−0.146***	0.013
Baa2*TRANS	−0.039**	0.014
Baa2*CAB	0.002***	0.001
Baa2*BOND	0.042***	0.003
Baa3	0.173***	0.015
Baa3*QMEP	−0.242***	0.017
Baa3*TRANS	0.001	0.024
Baa3*CAB	0.002*	0.001
Baa3*BOND	0.051***	0.004
Year-end dummy	0.064***	0.005
Fiscal year-half dummy	−0.004	0.005
Fiscal year-end dummy	0.028***	0.005
Adjusted R-squared	0.506	

Notes: Estimation is by OLS. ***, **, and * denote the 1, 5, and 10 percent significance level, respectively.
Credit ratings are the long-term ratings of Moody's.

Appendix Table 2 (continued). Regression Results on the Role of the BOJ's Monetary Policy: Part I

B. Maturity of Less than 90 Days

Dependent Variable: NCD Issuance Interest Rate
—Uncollateralized Overnight Call Rate
Number of Observations: 1,540 (October 5, 1998–May 9, 2005)

Variable	**Coefficient**	**Standard Error**
A1	0.034	0.247
A1*CAB	−0.001	0.007
A1*BOND	−0.081	0.314
A2	0.146***	0.041
A2*ZIRP	−0.021	0.017
A2*QMEP	−0.124***	0.018
A2*TRANS	0.005	0.018
A2*CAB	−0.001	0.001
A2*BOND	−0.071	0.146
A3	0.142***	0.011
A3*ZIRP	−0.047***	0.011
A3*QMEP	−0.146***	0.011
A3*TRANS	−0.010	0.009
A3*CAB	0.001	0.000
A3*BOND	0.041*	0.023
Baa1	−0.271	0.236
Baa1*CAB	0.007	0.007
Baa1*BOND	0.269	0.165
Baa2	0.199***	0.013
Baa2*ZIRP	−0.226***	0.014
Baa2*QMEP	−0.218***	0.013
Baa2*TRANS	−0.023*	0.014
Baa2*CAB	0.001**	0.001
Baa2*BOND	0.059***	0.004
Baa3	0.216***	0.015
Baa3*QMEP	−0.255***	0.017
Baa3*TRANS	0.004	0.023
Baa3*CAB	0.001	0.001
Baa3*BOND	0.050***	0.004
Year-end dummy	0.040***	0.005
Fiscal year-half dummy	0.003	0.005
Fiscal year-end dummy	0.019***	0.005
Adjusted R-squared	0.615	

Notes: Estimation is by OLS. ***, **, and * denote the 1, 5, and 10 percent significance level, respectively.
Credit ratings are the long-term ratings of Moody's.

Appendix Table 3. Regression Results on the Role of the BOJ's Monetary Policy: Part II

A. Less than 60 Days

Dependent Variable: NCD Issuance Interest Rate

—Uncollateralized Overnight Call Rate

Number of Observations: 1,490 (January 4, 2001–May 9, 2005)

Variable	**Coefficient**	**Standard Error**
A1	-0.154^{***}	0.049
A1*TERM	0.025^{***}	0.008
A1*BOND	0.126	0.165
A2	0.022	0.016
A2*TERM	-0.007^{***}	0.003
A2*BOND	0.054	0.055
A3	0.048^{***}	0.007
A3*TERM	-0.010^{***}	0.002
A3*BOND	0.026^{**}	0.011
Baa1	0.048^{*}	0.028
Baa1*TERM	-0.008^{**}	0.004
Baa1*BOND	-0.015	0.071
Baa2	0.041^{***}	0.008
Baa2*TERM	-0.008^{***}	0.002
Baa2*BOND	0.044^{***}	0.004
Baa3	0.082^{***}	0.010
Baa3*TERM	-0.010^{***}	0.003
Baa3*BOND	0.007^{***}	0.003
Year-end dummy	0.006^{**}	0.003
Fiscal year-half dummy	0.004	0.003
Fiscal year-end dummy	0.035^{***}	0.003
Adjusted R-squared	0.390	

Notes: Estimation is by OLS. ***, **, and * denote the 1, 5, and 10 percent significance level, respectively.
Credit ratings are the long-term ratings of Moody's.

Appendix Table 3 (continued). Regression Results on the Role of the BOJ's Monetary Policy: Part II

B. Less than 90 Days

Dependent Variable: NCD Issuance Interest Rate

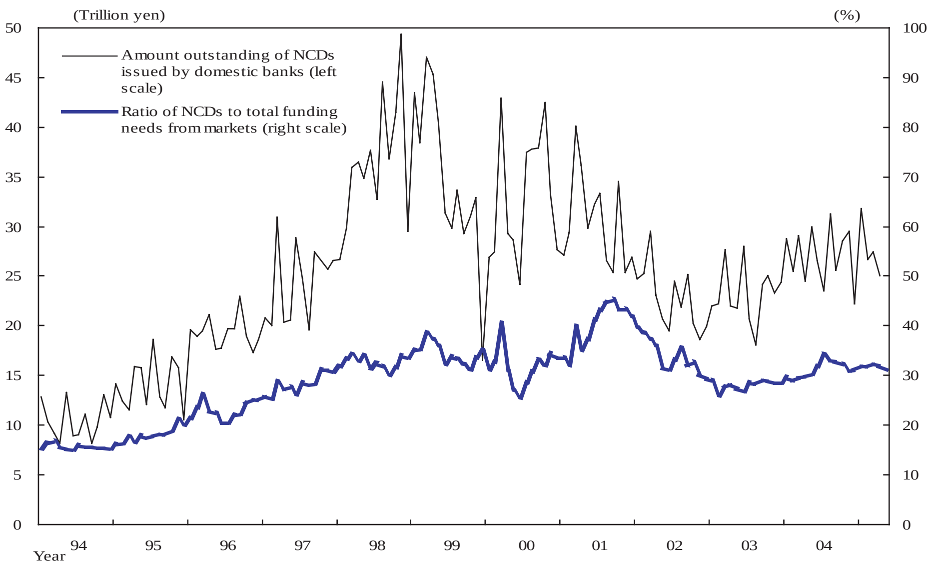
—Uncollateralized Overnight Call Rate

Number of Observations: 1,248 (January 4, 2001–May 9, 2005)

Variable	Coefficient	Standard Error
A1	-0.193^{***}	0.053
A1*TERM	0.033^{***}	0.009
A1*BOND	0.095	0.179
A2	0.055^{***}	0.021
A2*TERM	-0.013^{***}	0.003
A2*BOND	-0.028	0.066
A3	0.056^{***}	0.007
A3*TERM	-0.011^{***}	0.002
A3*BOND	0.025^{**}	0.011
Baa1	0.055	0.038
Baa1*TERM	-0.011^{*}	0.006
Baa1*BOND	0.029	0.095
Baa2	0.032^{***}	0.010
Baa2*TERM	-0.006^{***}	0.002
Baa2*BOND	0.059^{***}	0.005
Baa3	0.082^{***}	0.011
Baa3*TERM	-0.011^{***}	0.003
Baa3*BOND	0.014^{***}	0.003
Year-end dummy	0.012^{***}	0.003
Fiscal year-half dummy	0.008^{**}	0.003
Fiscal year-end dummy	0.048^{***}	0.003
Adjusted R-squared	0.448	

Notes: Estimation is by OLS. ***, **, and * denote the 1, 5, and 10 percent significance level, respectively.
Credit ratings are the long-term ratings of Moody's.

Appendix Figure 1. Size of the NCD Market

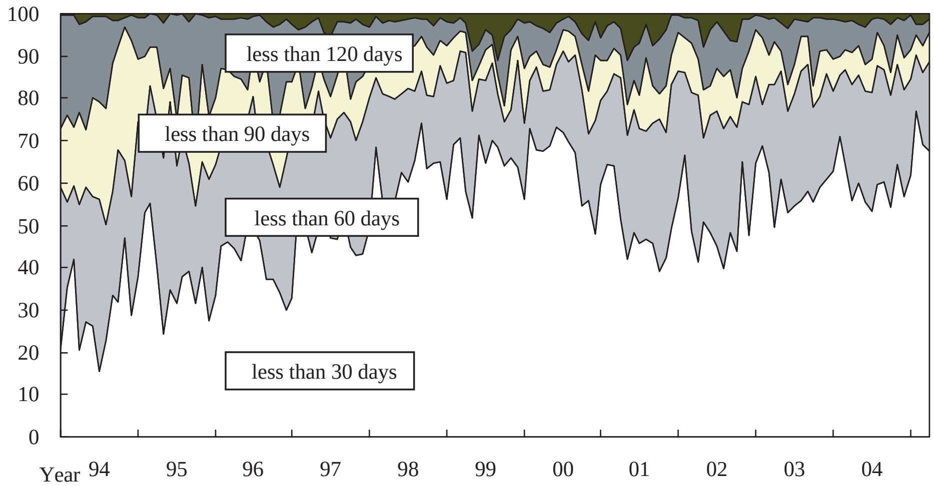


Notes: The amount outstanding is based on the banking account of domestically licensed banks and the ratio of NCDs to total funding needs from markets is based on city and trust banks.

Total funding needs from markets is defined as the sum of NCDs, CP, call money, bank bonds (including bank debentures), and repurchase agreements.

Source: Bank of Japan.

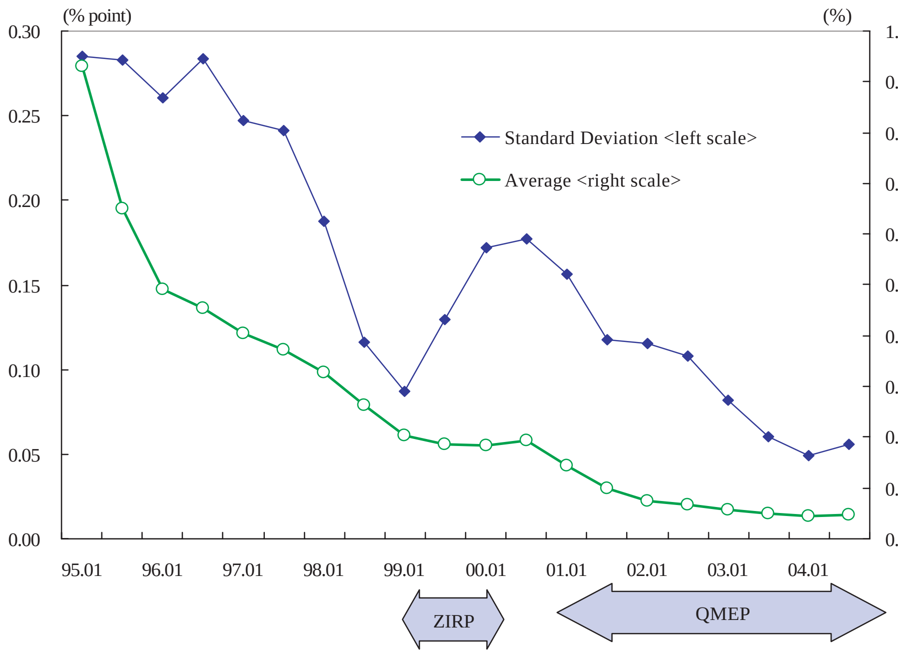
Appendix Figure 2. NCD Issuance by Maturity



Notes: Calculation is based on city and trust banks.
The maturity of less than 60 days denotes the maturity of 60 days to 89 days, and the maturity of less than 90 days denotes the maturity of 60 days to 179 days, respectively.

Source: Bank of Japan.

Appendix Figure 3. Standard Deviation of Fund-Raising Costs via Deposits Among Banks



Note: Sample banks include city banks, regional banks, and regional banks II.

Source: Financial statements released by each bank.

Using Market Information for Banking System Risk Assessment*

Helmut Elsinger,[a] Alfred Lehar,[b] and Martin Summer[c]
[a]Department of Finance, University of Vienna
[b]Haskayne School of Business, University of Calgary
[c]Economic Studies Division, Oesterreichische Nationalbank

We propose a new method for the analysis of systemic stability of a banking system relying mostly on market data. We model both asset correlations and interlinkages from interbank borrowing so that our analysis gauges two major sources of systemic risk: correlated exposures and mutual credit relations that may cause domino effects of insolvencies. We apply our method to a data set of the ten major UK banks and analyze insolvency risk over a one-year horizon. We also suggest a stress-testing procedure by analyzing the conditional asset return distribution that results from the hypothetical failure of individual institutions in this system. Rather than looking at individual bank defaults ceteris paribus, we take the change in the asset return distribution and the resulting change in the risk of all other banks into account. This takes previous stress tests of interlinkages a substantial step further.

JEL Codes: G21, C15, C81, E44.

*Martin Summer thanks the Bank of England for its hospitality and support during the work on this project. Helmut Elsinger and Alfred Lehar are grateful for financial support from the Jubiläumsfonds der Oesterreichischen Nationalbank under grant number 10972. We thank Nyeong Lee for valuable research assistance. We thank Charles Goodhart, Mathias Drehmann, Miguel Segoviano, Glenn Hoggarth, Alistair Cunningham, Garry Young, and Simon Wells as well as seminar participants at the Bank of England, the London School of Economics, Imperial College London, the European Central Bank, the University of Frankfurt, and the University of Munich for helpful discussions and comments. The views expressed in this paper are entirely the views of the authors and do not necessarily reflect the views of OeNB. Corresponding author: Summer: Otto-Wagner-Platz 3, A-1011 Wien, Austria; e-mail: martin.summer@oenb.co.at, Tel: +43-1-40420 7212, Fax: +43-1-40420 7299. Other author contact: Elsinger: Brünner Strasse 72, A-1210 Wien, Austria; e-mail: helmut.elsinger@univie.ac.at, Tel: +43-1-4277 38057, Fax: +43-1-4277 38054. Lehar: 2500 University Drive NW, Calgary, AB, Canada T2N 1N4; e-mail: alehar@ucalgary.ca, Tel: +1-(403) 220 4567.

1. Introduction

We suggest a new method for analyzing systemic financial stability of banking systems relying on market data and nonproprietary accounting data. The central idea is to combine concepts from finance and modern risk management with a network model of interbank loans to analyze the probability of *simultaneous* failures of banks—often referred to as *systemic risk*—and to develop a simple stress-testing procedure. We apply our ideas to a data set describing the system of the ten major UK banks and find that this system appears to be very stable. In particular, the likelihood of domino effects of bank insolvencies is very low. We also gain three more general insights. First, we see that for the analysis of systemic risk, defined as the probability assessment of joint default events, the analysis of both correlations and interlinkages is important. An analysis based on single institutions underestimates these events. Second, we see that stress testing of interbank linkages based on idiosyncratic default events only underestimates the impact of bank defaults on the rest of the system by a considerable margin. Third, we see that a simultaneous risk analysis of all major banks in a system can be done even when access to large proprietary microdata sets about individual banks is not available.

1.1 Related Research

In a series of recent papers analyzing interbank exposures such as Humphrey (1986), Angelini, Maresca, and Russo (1996), Furfine (2003), Wells (2004), Degryse and Nguyen (2004), VanLelyveld and Liedorp (2004), Upper and Worms (2004), and Mistrulli (2005), it has become common practice to investigate contagious defaults that result from the hypothetical failure of some single institution. This sort of analysis is able to capture the effect of idiosyncratic bank failures (e.g., because of fraud). It emphasizes one source of systemic risk, namely interbank linkages, and ignores the other, i.e., it is silent on correlation between banks' exposures. We believe that a meaningful risk assessment is only possible by studying both aspects in conjunction. Our paper builds on the model developed in Elsinger, Lehar, and Summer (2004), which incorporates both sources of systemic risk simultaneously. While in their model the distribution of bank asset returns is inferred from bank-specific data on market and credit risk exposures derived from a combination of various proprietary data sets of the Austrian Central Bank (OeNB), in contrast, in this paper the distribution of bank asset returns is inferred indirectly from stock market return data. The method of indirectly inferring bank asset return correlations from market data builds on the work of Lehar (2005).

1.2 An Overview of the Model and Main Results

We reconstruct a time series for the market values of assets for ten large publicly traded UK banks by viewing equity as a call option on total assets. We analyze the *covariance structure* of asset returns and simulate potential risk situations for the banking system as a whole based on this analysis. The advantage of this approach to model the uncertainty of bank asset returns lies in the fact that it does not depend on proprietary data sources. Of course, this advantage does not come without a price. While in highly developed financial systems stock market data are likely to incorporate all relevant public information on a bank's risk exposure, the data do not necessarily incorporate private information that is often contained in supervisory bank microdata and loan registers. Private information is, however, likely to be important for assessing the risks of a bank due to the opaque nature of bank asset values. One way to see the approach to bank asset risk modeling suggested in this paper is that it offers an alternative approach when private information—as is very often the case in practice—is not available.

Using a network model of the interbank market (following the model of Elsinger, Lehar, and Summer 2004) we investigate default probabilities and so-called domino effects. More significantly, we analyze the differences that arise in risk assessment when we take a naive approach, neglecting correlations; when we analyze correlations but ignore interlinkages; and finally, when we additionally take interlinkages into account. We then model the impact of various stress scenarios for the banking system by using a method that preserves the idea of previous papers examining scenarios where each bank in the system fails one at a time. But in contrast to this literature, we do so in a way that is consistent with the correlation structure of asset returns. Put another way, rather than simply removing a bank from the system one at a time (leaving everything else equal) we look at the *conditional distribution* of asset returns resulting from the event that one bank fails.

The empirical analysis gives the following main insights. First, the UK banking system appears to be very stable. In particular, the likelihood of domino effects is very low. Second, the simultaneous consideration of correlation and interlinkages does indeed make a difference for the assessment of systemic financial stability. In particular, the probability of systemic events such as the *joint breakdown of major institutions* is underestimated when correlations between banks are ignored. We can also show that ignoring interlinkages leads to an underestimation of joint default events. Third, the analysis uncovers substantial differences between banks concerning their impact on others in stress scenarios and clearly identifies institutions with a high systemic impact.

Finally, we demonstrate the importance of the assumption about the source of the shock when studying the consequences of a bank default. While the previous literature has studied idiosyncratic shocks, only our model captures systematic shocks too. We suggest a hypothetical decomposition into idiosyncratic and systematic sources of a shock that may hit a bank. In this way we can investigate not only the extreme cases studied in the existing literature but also intermediate cases. By measuring the expected shortfall for all other banks in the system conditional on the default of one bank, we demonstrate that a systematic shock has a much higher impact on financial stability than an idiosyncratic one. Basing a stress test entirely on idiosyncratic shock scenarios may therefore considerably underestimate the impact of the shock on the banking system as a whole. The impact of a bank's default on the banking system is much smaller if we assume an idiosyncratic shock than if we assume that the bank defaults following a macroeconomic shock.

2. A System Perspective on Risk Exposure for Banks

Our network model of interbank credits is a version of the model of Eisenberg and Noe (2001). We refer the reader to this paper for technical details. For our purpose of risk analysis, we extend their model to include uncertainty. Consider a set $\mathcal{N} = \{1, ..., N\}$ of banks. Each bank $i \in \mathcal{N}$ is characterized by a given value e_i net of interbank positions and its nominal liabilities l_{ij} against other banks $j \in \mathcal{N}$ in the system. The entire banking system is thus described by an $N \times N$ matrix L and a vector $e \in \mathbb{R}^N$. We denote this system by the pair (L, e).

The total value of a bank is the value of e_i plus the value of all payments *received* from counterparties in the interbank market minus the interbank liabilities. If for a given pair (L, e) the total value of a bank becomes negative, the bank is insolvent. In this case we assume that creditor banks are rationed proportionally. Denote by $d \in \mathbb{R}_+^N$ the vector of total obligations of banks toward the rest of the system, i.e., $d_i = \sum_{j \in \mathcal{N}} l_{ij}$. Define a new matrix $\Pi \in [0,1]^{N \times N}$ which is derived from L by normalizing the entries by total obligations.

$$\pi_{ij} = \begin{cases} \frac{l_{ij}}{d_i} & \text{if } d_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

We describe a banking system as a tuple (Π, e, d) for which we define a *clearing payment vector* p^*. The clearing payment vector has to respect limited liability of banks and proportional sharing in case of default. It denotes the total payments made by the banks under the clearing mechanism. It is defined by

$$p_i^* = \begin{cases} d_i & \text{if } \sum_{j=1}^{N} \pi_{ji} p_j^* + e_i \geq d_i \\ \sum_{j=1}^{N} \pi_{ji} p_j^* + e_i & \text{if } d_i > \sum_{j=1}^{N} \pi_{ji} p_j^* + e_i \geq 0 \\ 0 & \text{if } \sum_{j=1}^{N} \pi_{ji} p_j^* + e_i < 0 \end{cases} \tag{2}$$

This can be written more compactly as

$$p^* = \min\left[d, \max\left(\Pi' p^* + e, 0\right)\right], \tag{3}$$

where the min and max operators denote the componentwise maximum and minimum. The clearing payment vector directly gives us two important insights: for a given structure of liabilities and bank values (Π, e, d) we can identify insolvent banks ($p_i^* < d_i$) and derive the recovery rate for each defaulting bank ($\frac{p_i^*}{d_i}$).

To find a clearing payment vector, we employ a variant of the fictitious default algorithm developed by Eisenberg and Noe (2001). They prove that under mild regularity conditions, a unique clearing payment vector for (Π, e, d) always exists. These results extend to our framework as well.

From the solution of the clearing problem, we can gain additional economically important information with respect to systemic stability. Default of bank i is called fundamental if bank i is not able to honor its promises under the assumptions that all other banks honor their promises

$$\sum_{j=1}^{N} \pi_{ji} d_j + e_i - d_i < 0.$$

A contagious default occurs when bank i defaults only because other banks are not able to keep their promises, i.e.,

$$\begin{aligned} \sum_{j=1}^{N} \pi_{ji} d_j + e_i - d_i &\geq 0 \\ &\text{but} \\ \sum_{j=1}^{N} \pi_{ji} p_j^* + e_i - d_i &< 0. \end{aligned}$$

To use the model for risk analysis, we extend it to an uncertainty framework by assuming that e is a random variable. As there is no closed-form solution for the distribution of p^*, given the distribution of e, we have to resort to a simulation approach where each draw is called a scenario. By the theorem of Eisenberg and Noe (2001) we know that there exists a (unique) clearing payment vector p^* for each scenario. Thus from an ex ante perspective we can assess expected default frequencies from interbank credits across scenarios as well as the expected severity of losses from these defaults given that we have an idea about the distribution of e. Furthermore, we are able to decompose insolvencies across scenarios into fundamental and contagious defaults.

To pin down the distribution of e we choose the following approach: assume that there are two dates: $t = 0$, which is the *observation date,* and $t = T$, which is a *hypothetical clearing date* where all interbank claims are settled according to the clearing mechanism. At $t = 0$ the interbank exposures are observed. Assuming that these positions remain constant for the time horizon under consideration, they constitute the matrix L at T. This implies that the liability structure of the banks remains constant. Yet, there is empirical evidence (see Shibut 2002) that the creditors of distressed banks withdraw unsecured funds before the bank fails.[1] If creditors learn between $t = 0$ and $t = T$ that a bank is distressed, they will try to withdraw their unsecured funds.[2] If this were interbank funds, this would change the assumed seniority structure. Though this could reduce the risk of contagion and increase the loss to a deposit insurer, it will not change the risk of fundamental default.

Given the assumption of constant interbank claims, the value of the banks at $t = T$ depends solely on the realization of the random value of e at T, which is defined as the net assets before interbank positions are taken into account, i.e.,

$$e_i = V_i(T) - D_i(T) - \left(\sum_{j=1}^{N} \pi_{ji} d_j - d_i \right),$$

where $V_i(T)$ is the value of total assets of bank i and $D_i(T)$ is the value of total liabilities of bank i at time T. As in Duan (1994) we assume that the liabilities are insured and hence accrue at the risk-free interest rate. Therefore, $D_i(T) = D_i(0)e^{rT}$ and the distribution of e_i is determined by the distribution of $V_i(T)$ only.

Given the lack of available data on UK banks' net asset positions, we model $V_i(t)$ as a geometric Brownian motion under the objective probability measure P, i.e.,

$$dV_i = \mu_i V_i dt + V_i \sigma_i dB_i,$$

where B_i is a *one-dimensional* Brownian motion.[3] An important innovation in our research is that we explicitly allow the asset values

[1]On the other hand, Cocco, Gomes, and Martins (2004) show that for overnight loans, lending relationships do play a role in the interbank market and that during the Russian financial crisis, banks relied on relationship lending even more than usual.

[2]In case of netting agreements, another way to reduce exposures could be to take up funds from the troubled institution.

[3]This approach follows Merton (1974) and has been applied to banking systems as a whole by Lehar (2005).

of different banks to be correlated, i.e., the instantaneous correlation of $B_i(t)$ and $B_j(t)$ denoted by ρ_{ij} might be different from zero for all i and j. Given the drift parameters μ_i and the variance-covariance matrix Σ where $\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$, we are able to simulate the future asset values $V_i^s(T)$ of all banks simultaneously taking the correlation structure between their asset values into account. For details we refer the reader to Elsinger, Lehar, and Summer (2005).

By correcting $V_i^s(T)$ for interbank positions and deducting total liabilities $D_i(T)$ in each scenario, we construct the net income position for each bank as follows:

$$e_i^s = V_i^s(T) - D_i(T) - \left(\sum_{j=1}^{N} \pi_{ji} d_j - d_i \right).$$

This together with the interbank matrix L determines a clearing payment vector for each realization. Based on this information, we conduct our risk analysis.

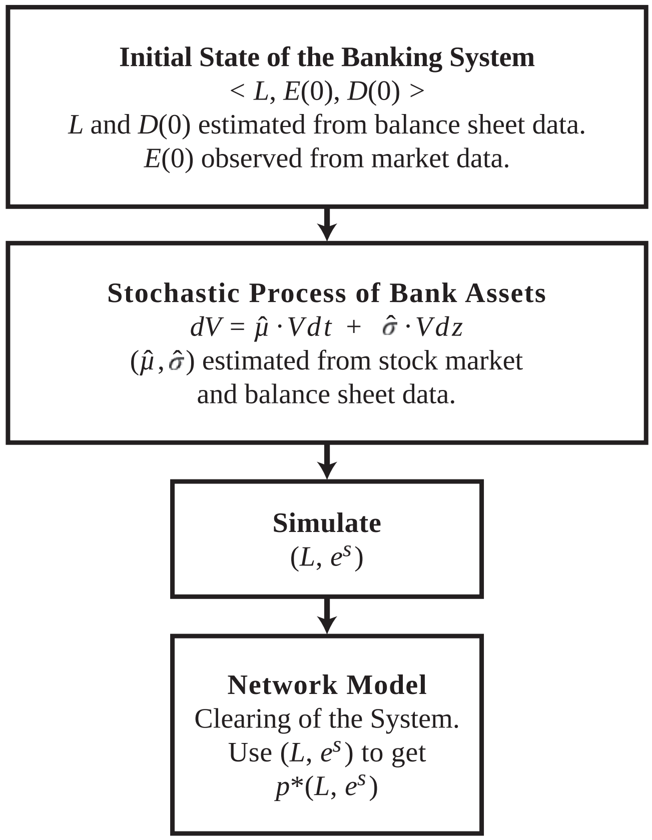
Neither the initial bank asset value $V(0)$ nor the drift μ nor the variance covariance matrix Σ are observable. Our approach therefore requires not only an estimate of interbank liabilities, but also estimates of the parameters of the stochastic processes governing bank assets, and of the market values of total assets. The simulation is then performed using the estimated values. An overview of the model is given in figure 1. Like all market or credit risk models, we have to assume a time horizon, which we set to one year.

3. Estimating Bank Asset Risk from Market Data

A bank's asset portfolio consisting of loans to nonbanks, interbank loans, traded securities, and many other items is funded by debt and equity. So in order to estimate the value of total assets, we need information on the future development of asset values and the face value of debt. The problem is that the actual market value of assets is not directly observable.[4] What is, however, observable is the market value of equity and the face value of debt for each publicly traded

[4]The dynamics of the market value of a bank's liabilities are not important, as the bank is assumed to default whenever the market value of the assets is below the promised payments, which is the book value of liabilities.

Figure 1. The Structure of the Model



bank. By viewing equity as a European call option on the bank's assets with a strike price equal to the value of debt at maturity, we can make use of this information to get an estimate of the market value of assets for each publicly traded bank.[5]

Denote the equity of bank i at t by $E_i(t)$ and the total face value of its interest-bearing debt by $D_i(t)$, which is assumed to have a time to maturity of T_1. We assume that all bank debt is insured and will

---

[5] This idea goes back to Black and Scholes (1973) and Merton (1973).

therefore grow at the risk-free rate.[6] The value of bank equity is then given by the call option price formula:

$$E_i(t) = V_i(t)\ \Phi(k_i(t)) - D_i(t)\ \Phi(k_i(t) - \sigma_i \sqrt{T_1}), \tag{4}$$

where

$$k_i(t) = \frac{\ln(V_i(t)/D_i(t)) + (\sigma_i^2/2)T_1}{\sigma_i \sqrt{T_1}} \tag{5}$$

and $\Phi(\cdot)$ is the cumulative standard normal distribution.[7] This formula is invertible in the sense that given $(E_i(t), D_i(t), \sigma_i, T_1)$ are all larger than 0, the value of total assets $V_i(t)$ is uniquely determined. Hence, given an estimate of σ_i, we can infer the market value of total assets from observable data.

The parameters of the stochastic processes are estimated using a maximum likelihood approach as developed in Duan (1994) and Duan (2000). As we are interested in the joint behavior of the total assets, we extend this technique by estimating the parameters of all banks simultaneously.

Given sequences $\boldsymbol{E}_i = (E_i(t))$ and $\boldsymbol{D}_i = (D_i(t)), t \in \{1 \dots m\}$ and $i \in \{1 \dots N\}$ of observed historical equity and debt values, respectively, the parameters (μ, Σ) of the asset value processes can be estimated by maximizing the following log-likelihood function:[8]

$$\begin{aligned} L(E) &= -\tfrac{(m-1)N}{2} ln(2\pi) - \tfrac{m-1}{2} ln|\Sigma| \\ &- \sum_{t=2}^{m} \left\{ \tfrac{N}{2} ln(h_t) + \tfrac{1}{2h_t} \left(\hat{x}_t - h_t\alpha\right)' \Sigma^{-1} \left(\hat{x}_t - h_t\alpha\right) \right\} \\ &- \sum_{t=2}^{m} \sum_{i=1}^{N} \left[ln\hat{V}_{i,t}(\sigma_i) + ln\Phi(\hat{k}_{i,t}) \right], \end{aligned}$$

where $\alpha_i = \mu_i - \frac{1}{2}\sigma_i^2$; h_t denotes the time increment from $t-1$ to t; $\hat{V}_{i,t}(\sigma_i)$ is the solution of equation (4) given σ_i; $\hat{k}_{i,t}$ corresponds to $k_i(t)$ in equation (5) with $V_i(t)$ replaced by $\hat{V}_{i,t}(\sigma_i)$; and $\hat{x}_{it} = \ln(\hat{V}_{i,t}(\sigma_i)/\hat{V}_{i,t-1}(\sigma_i))$.

[6] Relaxing this assumption will not dramatically change the results, since the paper's focus is not on deposit insurance pricing. From the available data, we cannot determine the amount of uninsured debt for every bank.

[7] Note, as the strike price equals $D_i(t)e^{rT_1}$, r cancels out in the Black-Scholes formula.

[8] For the derivation of the likelihood function see Elsinger, Lehar, and Summer (2005).

For the estimation of the parameters μ and Σ, we assume that the time to maturity of debt, T_1, equals one year. We use one year of weekly market values of total equity $E_i(t)$. From the estimation we get a set of parameters for every bank in the sample, which can then be used to back out the estimated asset values $\hat{V}_i(t)$ for every given equity price for each week during the past year. Put another way, we are able to estimate the value of total assets at each observation date for each bank.

In line with the standard risk management literature, we assume throughout the paper that the returns on the banks' asset portfolios are normally distributed. One could consider alternative distributions to include frequently observed characteristics of equity return series like fat tails.[9] However, this would be inconsistent with the assumptions of the estimation procedure in equation (4).

4. The Data

To apply the framework described in section 2 to the data, we need to determine the interbank exposures (the matrix L) as well as non-interbank exposures (the net worth positions e_i) for each bank. Since we describe the risks to e_i by the stochastic process approach, we can only consider banks that are publicly traded. All banks that are not in this category are summarized in a residual position. To estimate the parameters of the stochastic process governing the value of banks' assets, we use weekly stock market data for 2003 from Bloomberg. Total liabilities are taken from the Bank of England's bank balance sheet data.

Central banks usually have quite detailed information about their domestic banks' on-balance-sheet interbank positions. This information is available in form of balance sheet reports and supervisory data. The information is partial in several dimensions. First, the balance sheet does not contain exposures at a bilateral level. Some bilateral exposures can, however, be recovered by combining balance sheet information with other data sources.[10] Second, the balance

[9]Note that normality is assumed for the asset returns. The equity returns, where most studies document skewness and kurtosis, are not normally distributed in this setting.

[10]For instance, in their study for Austria, Elsinger, Lehar, and Summer (2004) can reconstruct 72 percent of on-balance-sheet interbank exposures exactly. Wells

sheet data allow a reconstruction of the interbank network only for the domestic banks, as data on overseas banks are usually only available as an aggregate position. The procedure thus can usually cover only banks that are owned domestically or branches and subsidiaries of foreign banks located within the country. Finally, off-balance-sheet information and exposures arising from intraday payment and settlement are not included.

For the estimation of an interbank exposure matrix, we look at the ten largest UK resident banks, an aggregate position for all other UK resident banks, and an aggregate position for foreign banks (i.e., branches and subsidiaries of overseas banks located within the United Kingdom).[11] This gives us a 10 by 10 matrix of interbank exposures of money market loans and deposits. As mentioned in Wells (2004), these data are unconsolidated. This is a measurement problem because the UK banking system is highly concentrated and the largest banking groups often have significant overseas subsidiaries and/or other subsidiaries located within the United Kingdom. But although potentially important exposures are excluded, we believe our data set provides an adequate estimate of the interbank liabilities. Wells (2004) finds that the data cover around 75 percent of total (on-balance-sheet) unsecured interbank assets.[12] He furthermore finds that OTC derivative exposures are small relative to on-balance-sheet interbank exposures. In table 1, we give an account of the size of on-balance-sheet interbank business for the last quarter of 2003.

Partial information about the interbank liability matrix L is available from balance sheet data. The bank-by-bank record of total interbank assets and liabilities provides the column and row sums of the matrix L. Further, some structural information is available. For example, the diagonal of L must contain only zeros since banks do

(2004) combines balance sheet data with the large exposure statistics to get an improved estimate on bilateral positions compared to an estimate that relies on balance sheet information only.

[11]As we have no information on default probabilities of foreign banks and the other UK banks, we assume in the following analysis that the exposure to these banks is well diversified and thus has zero default probability. To analyze the impact of interbank exposure to these banks, one could come up with ad hoc scenarios, like assuming that a certain fraction of foreign interbank debt is lost. Our framework allows us to analyze the impact of such scenarios on contagion.

[12]The other 25 percent is accounted for by commercial paper and certificates of deposit.

Table 1. UK-System Interbank Loans and Deposits in 2003:Q4

	Interbank Assets		Interbank Liabilities	
Bank Group	**Billion GBP**	**% of Total**	**Billion GBP**	**% of Total**
Major UK Banks	269.97	67.78%	270.07	67.81%
Other UK Banks	3.81	0.96%	2.89	0.72%
Foreign Banks	124.51	31.26%	125.39	31.47%
Total	398.29	100%	398.29	100%

Source: Bank of England.

not have claims and liabilities against themselves. For the UK banking system, limited information about certain large bilateral exposures is also available (see Wells 2004). But these data are based on a different definition of interbank exposure; for example, they include some off-balance-sheet exposures and so are not directly comparable with the loan and deposit data that we use to estimate the matrix L. Given our aim of using only market data, we do not incorporate these data into our analysis.

As fundamental defaults are determined by the sum of all claims and liabilities in the interbank market, the sum of individual rows and columns is sufficient for this purpose. But to calculate a clearing payment vector and to identify contagious defaults, the bilateral exposures have to be *estimated* based on this partial information. The fact that we cannot observe individual bilateral exposures should be reflected in the fact that these entries in the matrix are treated homogeneously in the estimation process. We formulate the estimation of the unobservable parts of the L matrix as an *entropy optimization problem.*

Intuitively, this procedure finds a matrix that treats all entries as balanced as possible and satisfies all known constraints. This can be formulated as minimizing a suitable measure of distance between the estimated matrix and a matrix that reflects our a priori knowledge. The so-called *cross-entropy* measure is a suitable concept for this task (see Fang, Rajasekera, and Tsao [1997] or Blien and Graef [1997]). A detailed description of the estimation procedure and the estimated matrix can be found in Elsinger, Lehar, and Summer (2004).

Our assumption on the structure of L will not affect fundamental defaults but will certainly have an impact on the number of contagious defaults. On the one hand, spreading out interbank loans among many banks might make the banking system more resilient toward shocks (Allen and Gale 2000); on the other hand, it might allow contagion to spread out more (consistent with the empirical findings of Elsinger, Lehar, and Summer 2004). To check for robustness we also estimated L matrices that are as sparse as possible.[13] Table 7 (shown at the end of section 7) contains some results of this robustness check.

5. Risk Analysis: Status Quo

For the estimate of the interbank matrix and the observed values of total equity and liabilities at the end of December 2003, our framework provides statistics of default scenarios in one year's time, i.e., at the end of 2004. Note that our model allows for a decomposition of default events into "fundamental" and "contagious" defaults. The results of the simulation are reported in table 2.

We see that the UK banking system—at least as far as the ten largest institutions are concerned—appears to be extremely stable. There are scenarios with nine defaults in total; however, their probability is practically zero, since it occurs in only one scenario out of 100,000. The probability that one or more defaults occur in the entire system over a one-year horizon given the December 2003 starting position is 4.7 percent. The probability of observing a domino effect is practically zero.

Various parameters in the clearing process can be changed to check the sensitivity of the results on the banking system's aggregate default statistics. When we change the procedure by netting

[13] We had to rely on heuristics for this estimation, since we are not aware of a well-suited algorithm for our problem.

Table 2. Frequency of Fundamentally and Contagiously Defaulting Banks Grouped by the Number of Fundamental Defaults (First Column)

Fundamental Defaults	Contagious Defaults								
	No Netting			Netting					
	0	**1**	**2**	**0**	**1**	**2**	**3**	**4**	**5**
0	95335	0	0	95335	0	0	0	0	0
1	3985	34	2	3971	49	1	0	0	0
2	409	37	8	402	45	5	1	0	1
3	98	23	2	93	25	5	0	0	0
4	31	11	4	27	14	3	2	0	0
5	11	4	2	6	6	5	0	0	0
6	2	1	0	0	2	1	0	0	0
7	0	0	0	0	0	0	0	0	0
8	0	1	0	0	1	0	0	0	0
Total	99871	111	18	99834	142	20	3	0	1

Note: The total number of scenarios is 100,000.

all bilateral exposures before the clearing mechanism is applied, the mean default probability as well as its standard deviation increase slightly compared to the case without netting. This is due to increasing second round effects or contagious defaults (see table 2).[14] If in addition we assume that insolvent institutions do not repay their interbank creditors after netting of bilateral exposures—which might be interpreted according to Elsinger, Lehar, and Summer (2004) as a "short-term" scenario—the probability of contagious defaults hardly rises at all and the default statistics remain virtually unchanged.

[14]Netting bilateral exposures might increase or decrease contagion (see Elsinger, Lehar, and Summer [2005] for examples). In our data set most of the banks are harmed by bilateral netting.

Table 3. Distribution of Individual Default Probabilities and Distance to Default

Bank	1	2	3	4	5	6	7	8	9	10
Default Prob	0%	0.01%	0.02%	0.02%	0.04%	0.10%	0.23%	0.59%	0.68%	4.06%
DD	7.11	4.78	3.48	3.46	3.31	3.10	2.90	2.42	2.45	1.73

Looking at the distribution of the individual Merton-default probabilities of the ten banks in our system, we see that the system is very stable. We have one outlier with a one-year default probability of 4 percent; all other individual default probabilities are in the range between 0 percent and 0.68 percent. The distribution of individual default probabilities is shown in table 3. The table also shows the distance to default under the objective probability, which is measured as

$$dd_i(T) = \frac{\left(\hat{\mu}_i - \frac{1}{2}\hat{\sigma}_i^2\right) T + ln \frac{V_i(0)}{D_i(T)}}{\hat{\sigma}_i \sqrt{T}}.$$

The results should be interpreted with caution. The focus of our model is not to derive individual default probabilities but rather to investigate the impact of correlation between bank portfolios versus contagion as well as to derive a stress-testing framework to identify system-relevant banks. The default probabilities of the Merton model should mainly be seen as providing a ranking of default risk among banks.[15]

6. The Role of Correlation and Interlinkages

Banking regulation has traditionally been more focused on individual banks than on the system as a whole. Hence, regulators are typically interested in the marginal distribution of $V_i(t)$ and less attention is given to the joint distribution of $V(t)$. Whereas this *marginal approach* gives the correct default probabilities of individual banks, the estimates for joint defaults based on the marginal distributions are, in general, not correct. The question is whether the improvement in estimating the probability of joint defaults by taking the correlation

[15] To get a precise default probability estimate, one could follow KMV and use a mapping of Merton default probabilities into empirical PDs.

Table 4. Number of Simultaneously Defaulting Banks across Simulations

Simultaneous Defaults	Marginal Distribution	Joint Distribution	Interbank Market	
			No Netting	Full Netting
0	94523	95335	95335	95335
1	5421	4021	3985	3971
2	56	454	443	451
3	0	123	137	139
4	0	46	62	57
5	0	17	24	26
6	0	3	10	9
7	0	0	3	10
8	0	1	0	1
9	0	0	1	1
10	0	0	0	0

Note: Simulations are based on the marginal distribution only (second column), on the joint distribution (third column), and on the joint distribution together with contagion (fourth column).

structure into account makes this more elaborate technique really necessary. To examine this, we compare the (simulated) number of joint defaults for three different procedures

1. based on the marginal distributions only, i.e., assuming that the covariances are zero,[16]

2. based on the joint distribution, and

3. based on the joint distribution taking the financial linkages between banks into account.

The results, shown in table 4, demonstrate that taking the correlation structure into account can have a considerable impact on estimates of default. The number of scenarios with a single defaulting bank decreases. In contrast, both the number of scenarios with no default at all and the number of scenarios where two or more

[16]For a description of the simulation procedure, see appendix 1.

banks default simultaneously increase. This result is further amplified when bank interlinkages (i.e., the potential for contagion) are taken into account.

This analysis shows that, from the viewpoint of systemic stability, both correlated exposures and interlinkages do matter. Ignoring the systemwide perspective—i.e., ignoring correlations and interlinkages—leads to a considerable underestimation of the probability of a systemic crisis. If we do not take into account interlinkages, the amount of underestimation of joint default probabilities is, from a practical point of view, perhaps not too big. Ignoring correlations, however, leads to an underestimation of joint default events by a significant margin.[17]

7. Risk Analysis: Stress Testing

Stress testing provides another measure of systemic stability, but importantly it also allows financial regulators to identify individual banks that may pose systemic risks. With the exception of Elsinger, Lehar, and Summer (2004), the literature on interbank linkages and domino effects has focused on stress tests that assume the default of single institutions, leaving the financial condition of the other banks unaffected. The implicit assumption of this previous research is that the cause of bank failure is an idiosyncratic shock that hits just one bank at a time. This approach is useful to study the consequences of fraud or to study the contagion impact within a banking system where banks' asset portfolios are rather uncorrelated, e.g., geographical diversification. But to look at stress testing from a more general perspective, we have to be more specific on the source of the assumed default.

From the perspective of systemic stability, the assumption of idiosyncratic shocks might lead to an underestimation of systemic risk, as there is evidence that the correlation between bank portfolios is generally positive.[18] When conducting stress testing on a system

[17] The results are quite robust with respect to the estimation procedure of the interbank matrix L. Using different estimates for L, we got similar results in terms of contagion (see table 7).

[18] See, for example, Nicolo and Kwast (2002) or Lehar (2005). While the new internal ratings-based approach of Basel II considers correlations of bank loans within a bank portfolio, our focus is on the correlation between bank portfolios.

level, the impact of a macroeconomic shock that hits the whole banking system should be a major concern for institutions charged with maintaining financial stability. Such a shock affects all banks to a certain degree, depending on their asset composition. Thus, we extend the current stress-testing framework by modeling a second reason for a bank's default—a systematic shock. If there is a positive correlation in banks' asset values, it is likely that if one bank defaults because of a declining asset value, other banks may also be expecting difficulties.

We model systematic shocks by deriving the multivariate conditional distribution for the banks' asset values. The idea is as follows: suppose that the regulator knows the joint unconditional distribution of the banks' asset values and observes that one bank has defaulted, partly due to a systematic shock. It is now rational for the regulator to update his or her beliefs on the joint distribution and compute the conditional distribution of all the other banks' asset values, given that one bank's asset value is below the bankruptcy threshold. Under this conditional distribution, default probabilities, the probability of contagion, and the losses to the deposit insurer would be expected to increase if bank asset values are positively correlated. Conducting such an analysis ex ante will allow the regulator to rank banks according to the impact of their default on the banking system and thus identify system-relevant banks. Appendix 2 outlines the simulation technique in detail.

Table 5 shows each bank's probability of default conditional on the default of bank i. We find a large variation across banks. For instance, the first bank has only a very small impact on the fundamental default probability of all the other banks, but is itself affected most by the hypothetical defaults of all the others. Banks 1, 4, and 7 have, on average, a much weaker impact on the others than banks 2, 3, 5, 6, 8, 9, and 10. On the other hand, banks 1, 3, 4, 5, and 7 are, on average, most affected by the change in asset correlations brought about by the default of other banks in the system. The pattern that seems to appear in this table is that the larger the distance to default, the higher the impact of a default on the other banks. The reason is that a bank with a large distance to default needs a large negative shock to make the bank default. In conjunction with the positive asset correlations, all other banks are seriously hit by this (systematic) shock, too. Yet, a closer look at table 5 reveals that,

Table 5. Probabilities of Default Conditional on the Failure of One Bank

Banks										DD
No 1	**No 2**	**No 3**	**No 4**	**No 5**	**No 6**	**No 7**	**No 8**	**No 9**	**No 10**	
–	92.7%	62.6%	49.0%	62.9%	98.8%	55.5%	64.3%	77.0%	97.7%	1.73
0.5%	–	5.4%	2.0%	6.5%	56.8%	2.1%	14.2%	12.0%	60.1%	3.46
1.2%	16.1%	–	5.7%	10.6%	79.4%	6.3%	19.3%	31.4%	88.4%	3.10
7.9%	55.6%	46.9%	–	42.4%	74.4%	26.9%	51.4%	44.5%	97.1%	2.45
3.8%	54.8%	34.0%	14.3%	–	55.0%	14.7%	38.6%	68.7%	77.6%	2.90
0.0%	0.0%	0.0%	0.0%	0.0%	–	0.0%	0.0%	0.0%	0.0%	7.11
8.3%	52.7%	47.6%	24.3%	32.7%	68.5%	–	48.5%	70.2%	81.6%	2.42
0.6%	23.7%	9.7%	3.0%	7.2%	65.1%	3.2%	–	14.9%	44.9%	3.31
0.3%	6.3%	6.0%	1.0%	5.4%	32.3%	1.8%	5.2%	–	18.2%	3.48
0.1%	6.6%	4.5%	0.6%	1.6%	73.8%	0.7%	4.1%	7.3%	–	4.78

Note: Each column i shows the default probabilities of the other banks, conditional on the default of bank i. The last column shows the distance to default for all banks.

for instance, the default of bank 6 hits bank 10 harder than bank 5, although bank 10 has a larger distance to default than bank 5. So, the distance to default, which is based on the marginal distribution of the asset value only, is a reasonable but not perfect indicator of whether the bankruptcy of a bank will have a small or large impact on the system.

To demonstrate the difference between systematic and idiosyncratic shocks, we assume that a fraction $(1 - a)$ of the distance to default dd_i hits bank i as an idiosyncratic shock $z_i^{idio} = -(1-a)dd_i$. We then draw a systematic shock z_i^s such that this bank is in default, i.e., $z_i^s + z_i^{idio} \leq -dd_i$. Given this systematic shock, we simulate the conditional distribution of all banks' asset values using the technique described in appendix 2. Hence, the simulation for the other banks is conditioned on the systematic shock only. This simulation is done for various levels of a ranging from 0 to 1. We run 100,000 simulations where $z_i^s + z_i^{idio} \leq -dd_i$ for each defaulting bank and each level a.

Note that we compute the conditional distribution of V using the estimated covariance matrix $\hat{\Sigma}$. As an alternative, one could assume a factor model, which would also allow a decomposition into

Table 6. For Each Bank i, Expected Shortfall for All Other Banks Conditional on the Default of Bank i

	Bank									
a	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	77	168	192	139	267	170	235	144	165	180
0.1	85	223	241	169	305	267	256	188	218	262
0.25	103	385	370	241	404	643	315	306	366	558
0.5	157	1170	897	492	769	3059	528	835	1133	2541
0.75	263	3835	2424	1118	1669	12568	1011	2611	4227	11615
0.9	370	7566	4411	1882	2721	25791	1547	5201	9111	24845
1	469	11550	6498	2676	3782	38814	2069	8071	14460	37513

Note: The shortfall (in £m) is computed for different ratios of idiosyncratic to systematic shocks (first column). The shock that causes bank i's default is assumed to consist of a systematic part (a) and an idiosyncratic part (1–a).

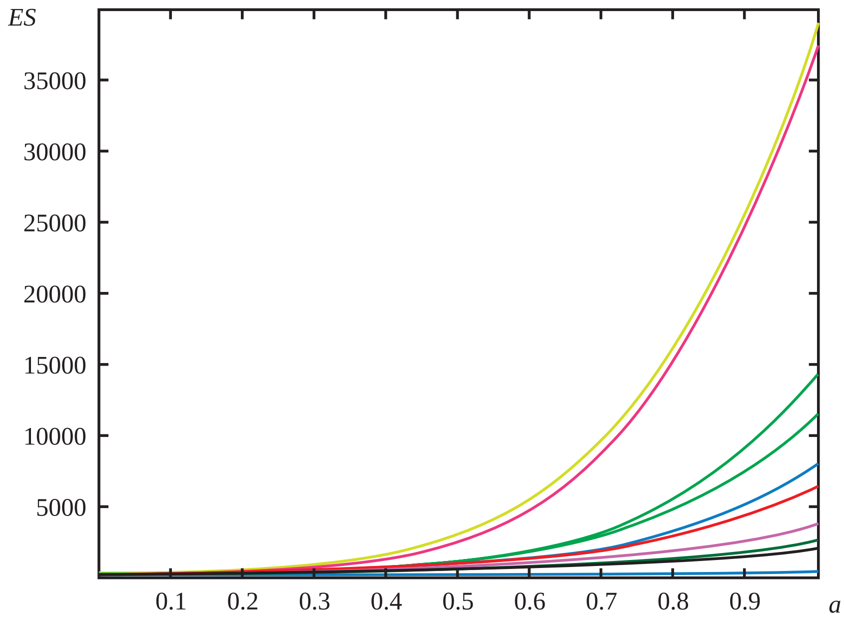
systematic and idiosyncratic shocks. Such a model, however, would just be equivalent to imposing a special structure on $\hat{\Sigma}$. If the aim is to get a quick impression of the difference in magnitude of expected shortfall that comes with the stress assumption, our suggested decomposition is perhaps the simplest and most direct way. As a measure of systemic importance of bank i, we compute the expected shortfall for all other banks conditional on the default of bank i. That is

$$ES_i = \frac{1}{S}\sum_{s=1}^{S}\sum_{k=1,k\neq i}^{N}\max(D_k(T) - V_k^s(T), 0), \tag{6}$$

where N is the number of banks and S is the number of simulation runs. If all deposits are insured, the expected shortfall is equal to the liability of the deposit insurer. Therefore, we can interpret ES_i as the increase in the liability of the deposit insurer that results from the failure of bank i.

In line with our intuition, we find that systematic shocks constitute a much bigger threat for financial stability than idiosyncratic shocks. Table 6 shows expected shortfall (in £m) conditional on each

Figure 2. Systematic versus Idiosyncratic Shocks



Note: For each bank i of the ten banks, the expected shortfall ES for all other banks conditional on the default of bank i is plotted for different weights of the systematic component a of the shocks. The shock that causes bank i's default is assumed to consist of a systematic part (a) and an idiosyncratic part ($1 - a$).

bank's default for different levels of a. A completely idiosyncratic shock is simulated whenever $a = 0$ and the shock is assumed to be completely systematic in the case of $a = 1$. Figure 2 illustrates the results.

From the results, we can see that when defining a stress-testing framework for a financial stability assessment, we have to be precise about which situation we want to analyze. Idiosyncratic shocks because of fraud will have a much smaller impact on the banking system than a systemwide shock of similar magnitude. Our approach allows us to come up with measures of systemic importance that combine both aspects of systemic risk, the correlation between banks' assets

Table 7. Number of Simultaneously Defaulting Banks across Simulations Based on Different Estimates of the Matrix *L*

Defaults	Entrop	A	B
0	95335	95335	95335
1	3985	4005	3941
2	443	448	473
3	137	123	144
4	62	51	55
5	24	26	19
6	10	8	14
7	3	3	9
8	0	0	7
9	1	1	3
10	0	0	0

Note: The results in the column labeled “Entrop” are based on the solution of the relative entropy minimization. The matrix A (B) is an estimate of L with the highest (lowest) probability of a single default we were able to find.

as well as contagion. Regulators can therefore identify banks that are crucial for the stability of the banking sector.

8. Conclusions

This paper has outlined a new framework for systemic financial stability analysis for banking systems, which relies mainly on easily observable market data. We apply this framework to the ten major UK banks and suggest a stress-testing procedure. Our motivation stems from the fact that for the analysis of systemic risk—the large-scale breakdown of financial intermediation—the main events of interests are the *joint failures of major financial institutions*. Therefore it is essential to capture two major sources of risk that can lead to simultaneous insolvencies. This requires the consideration of both *correlated exposures* and *credit interlinkages*. In most existing studies, attention is focused exclusively on domino effects that result from interlinkages, when single institutions fail ceteris paribus. One of our main results is that the existing approach potentially underestimates joint default events by a significant margin and that considering the two sources of systemic risk indeed matters.

For stress testing we demonstrate how the assumption of a default of a major institution can be simulated consistently with the risks inherent in the bank's assets. We do so by considering the conditional covariance structure of bank asset returns that result from the failure of one institution and study how this changed covariance structure influences domino effects of defaults. Thus we carry previous stress tests for interlinkages a significant step further by embedding these stress tests in a coherent risk analysis. Furthermore, we analyze the role of the assumption of idiosyncratic defaults in the stress testing of interlinkages that was frequently used in the previous literature. We demonstrate that this assumption leads to a much lower impact on the rest of the banking system than assuming that the source of the shock is systematic. Stress tests of interlinkages therefore underestimate the impact of bank breakdowns on the stability of the financial system. The empirical analysis uncovers substantial differences between individual banks concerning their impact on others in stress scenarios and clearly identifies institutions with a high systemic impact.

We hope that our results will be useful in the search for a canonical model to perform risk assessment for banking systems for institutions in charge of systemic financial stability. Since our method relies mainly on market data, it can be more easily applied than methods relying strongly on proprietary information such as loan registers and supervisory data. While such data sources are very rich and allow a more detailed analysis of risk factors, their drawback is that they are not widely available and usually under the close control of national supervisory bodies. Provided the system under consideration is financially highly developed—such as, for instance, in the United Kingdom—our method shows a workable alternative to naive single-institution analysis for systemic risk monitoring. We therefore believe that the approach outlined here is interesting for supranational institutions like the International Monetary Fund or the European Central Bank who do not have access to proprietary supervisory data sources but who are interested in financial stability assessment. The parsimony in data has the advantage that our approach is more easily replicable than proprietary data models and might thus be a useful building block to enhance our understanding of systemic risk monitoring for financial stability analysis through studies of other banking systems.

Appendix 1. The Marginal Approach

To simulate joint defaults neglecting the correlation structure, we use the following procedure. The marginal distribution of $V_i(T)$ is given by

$$V_i(T) = V_i(0) * exp\left(\left[\mu_i - \frac{1}{2}\sigma_i^2\right]T + \sigma_i B_i(T)\right),$$

where $B_i(T) \sim N(0, T)$. To generate a scenario s we randomly draw an $N \times 1$ vector $\tilde{B}^s$ of independent standard normal random variables and calculate

$$V_i^s(T) = V_i(0) * exp\left(\left[\hat{\mu}_i - \frac{1}{2}\hat{\sigma}_i^2\right]T + \hat{\sigma}_i\sqrt{T}\tilde{B}_i^s\right),$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i$ are the estimates of μ_i and σ_i. Then we count the number of banks for which their asset values $V_i^s(T)$ is less than their total liabilities $D_i(T)$.

Appendix 2. Conditional Default

In section 7 we assume that the regulator learns that bank i is in default. We ask the question, what can be deduced about the stability of the system given this information, i.e., what is the conditional distribution of the asset values of all other banks given the default of bank i? To do the simulations, we first reorder the banks such that the defaulting bank is the first one. Then we simulate asset returns according to the procedure below and count the number of *conditionally* defaulting banks.

The (asset) return of bank i is defined as $R_i(T) = ln(V_i(T)/V_i(0))$. We denote the vector of joint returns by $R(T) = (R_1(T), \ldots, R_N(T))'$. $R(T)$ is a multivariate normal random variable with $E[R_i(T)] = T(\mu_i - \frac{1}{2}\sigma_i^2) = T\alpha_i$ and $Var[R(T)] = T\Sigma$, i.e., $R(T) \sim MVN(\alpha, T\Sigma)$, where $\alpha = (\alpha_1, \ldots, \alpha_N)'$. Consider the following partition

$$R(T) = \begin{bmatrix} R^1(T) \\ R^2(T) \end{bmatrix} \quad \alpha = \begin{bmatrix} \alpha^1 \\ \alpha^2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix},$$

where the N random variables are partitioned into n_1 and n_2 variates $(n_1 + n_2) = N$. $R^2(T)$ given $R^1(T)$ is multivariate normally distributed with $E[R^2(T) \mid R^1(T)] = T\alpha^2 + \Sigma^{21}(\Sigma^{11})^{-1}(R^1 - T\alpha^1)$ and $Var[R^2(T) \mid R^1(T)] = T(\Sigma^{22} - \Sigma^{21}(\Sigma^{11})^{-1}\Sigma^{12})$.[19]

For our simulation, we factor Σ using the Cholesky decomposition such that $\Sigma = U'U$. Now define the random variable $S = T\alpha + \sqrt{T}U'Z$ where $Z \sim MVN(0_{N,1}, I_{N,N})$. Evidently, S has the same distribution as R, i.e., $S \sim MVN(T\alpha, T\Sigma)$. Partitioning S, U, Z conformably to R gives

$$S = \begin{bmatrix} S^1 \\ S^2 \end{bmatrix} \quad Z = \begin{bmatrix} Z^1 \\ Z^2 \end{bmatrix} \quad U = \begin{bmatrix} U^{11} & U^{12} \\ 0 & U^{22} \end{bmatrix}.$$

This means that

$$S^1 = T\alpha^1 + \sqrt{T}(U^{11})'Z^1.$$

and

$$S^2 = T\alpha^2 + \sqrt{T}(U^{12})'Z^1 + \sqrt{T}(U^{22})'Z^2.$$

To simulate the conditional distribution of S^2 given $S^1 = R^1(T)$, we first calculate Z^1 as

$$Z^1 = \frac{1}{\sqrt{T}} \left((U^{11})'\right)^{-1} \left(R^1(T) - T\alpha^1\right).$$

Plugging this into the definition of S^2 yields

$$S^2 = T\alpha^2 + (U^{12})' \left((U^{11})'\right)^{-1} \left(R^1(T) - T\alpha^1\right) + \sqrt{T}(U^{22})'Z^2.$$

[19]See Ramanathan (1993, 109).

We know that S^2 given S^1 is multivariate normally distributed. It remains to be shown that $E[S^2 \mid R^1(T))] = E[R^2(T) \mid R^1(T)]$ and $Var[S^2 \mid R^1(T)] = Var[R^2(T) \mid R^1(T)]$. Note that $E[S^2 \mid R^1(T)] = T\alpha^2 + (U^{12})' \left((U^{11})'\right)^{-1} \left(R^1(T) - T\alpha^1\right)$ and

$$(U^{12})' \left((U^{11})'\right)^{-1} = (U^{12})' U^{11} (U^{11})^{-1} \left((U^{11})'\right)^{-1}.$$

Now $(U^{12})'U^{11} = \Sigma^{21}$ and $(U^{11})^{-1} \left((U^{11})'\right)^{-1} = (\Sigma^{11})^{-1}$. Hence

$$E[S^2 \mid R^1(T)] = T\alpha^2 + \Sigma^{21}(\Sigma^{11})^{-1}(R^1(T) - T\alpha^1).$$

The variance of S^2 given $S^1 = R^1(T)$ is $T(U^{22})'U^{22}$. By the definition of U it holds that

$$\begin{aligned} (U^{22})'U^{22} &= \Sigma^{22} - (U^{12})'U^{12} \\ &= \Sigma^{22} - (U^{12})'U^{11}(U^{11})^{-1} \left((U^{11})'\right)^{-1} (U^{11})'U^{12} \\ &= \Sigma^{22} - \Sigma^{21}(\Sigma^{11})^{-1}\Sigma^{12}, \end{aligned}$$

which is the same as the variance of $R^2(T)$ given $R^1(T)$. Hence, the conditional distribution of S^2 given $S^1 = R^1(T)$ is just the same as that of $R^2(T)$ given $R^1(T)$.

To generate a scenario s we assume that bank 1 defaults ($n_1 = 1$). Let $R_1^*(T)$ be such that $V_1(T) = V_1(0)exp(R_1^*(T)) = D_1(T)$. Now we randomly draw $R_1^s \leq R_1^*(T)$. Given this realization of $R_1(T)$, we simulate S^2 and calculate the asset values of the banks, $V_2^s(T), \ldots, V_n^s(T)$. Finally, we count the number of (conditionally) defaulting banks in scenario s. The results are based on 100,000 simulations. Note that the procedure can easily be extended to the case where several banks are assumed to be in default.

References

Allen, Franklin, and Douglas Gale. 2000. "Financial Contagion." *Journal of Political Economy* 108 (1): 1–34.

Angelini, Paolo, G. Maresca, and Daniela Russo. 1996. "Systemic Risk in the Netting System." *Journal of Banking and Finance* 20 (5): 853–86.

Black, Fischer, and Myron Scholes. 1973. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 8: 1637–59.

Blien, Uwe, and Friedrich Graef. 1997. "Entropy Optimizing Methods for the Estimation of Tables." In *Classification, Data Analysis, and Data Highways*, ed. Ingo Balderjahn, Rudolf Mather, and Martin Schader. Berlin: Springer Verlag.

Cocco, Joao F., Francisco J. Gomes, and Nuno C. Martins. 2004. "Lending Relationships in the Interbank Market." IFA Working Paper No. 384.

Degryse, Hans, and Gregory Nguyen. 2004. "Interbank Exposures: An Empirical Analysis of Systemic Risk in the Belgian Banking System." Working Paper, National Bank of Belgium.

Duan, Jin-Chuan. 1994. "Maximum Likelihood Estimation Using Price Data of the Derivative Contract." *Mathematical Finance* 4 (2): 155–67.

———. 2000. "Correction: Maximum Likelihood Estimation Using Price Data of the Derivative Contract." *Mathematical Finance* 10 (4): 461–62.

Eisenberg, Larry, and Thomas Noe. 2001. "Systemic Risk in Financial Systems." *Management Science* 47:236–49.

Elsinger, Helmut, Alfred Lehar, and Martin Summer. 2004. "Risk Assessment for Banking Systems." Working Paper, University of Vienna.

———. 2005. "Using Market Information for Banking System Risk Assessment." SSRN Working Paper, http://ssrn.com/abstract = 787929.

Fang, Shu-Cherng, Jay R. Rajasekera, and H.-S. Jacob Tsao. 1997. *Entropy Optimization and Mathematical Programming*. Boston, London, Dordrecht: Kluwer Academic Publishers.

Furfine, Craig. 2003. "Interbank Exposures: Quantifying the Risk of Contagion." *Journal of Money, Credit, and Banking* 35 (1): 111–28.

Humphrey, David B. 1986. "Payments Finality and Risk of Settlement Failure." In *Technology and the Regulation of Financial Markets: Securities, Futures and Banking*, ed. Anthony Saunders and Lawrence J. White. Lexington, MA: Lexington Books.

Lehar, Alfred. 2005. "Measuring Systemic Risk: A Risk Management Approach." *Journal of Banking and Finance* 29 (10): 2577–2603.

Merton, Robert C. 1973. "A Rational Theory of Option Pricing." *Bell Journal of Economics and Management Science* 4: 141–83.

———. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29 (2): 449–70.

Mistrulli, P. E. 2005. "Interbank Lending Patterns and Financial Contagion." Mimeo, Banca d'Italia.

Nicolo, Gianni De, and Myron L. Kwast. 2002. "Systemic Risk and Financial Consolidation: Are They Related?" *Journal of Banking and Finance* 26 (5): 861–80.

Ramanathan, Ramu. 1993. *Statistical Methods in Econometrics*. San Diego: Academic Press.

Shibut, Lynn. 2002. "Should Bank Liability Structure Influence Deposit Insurance Pricing?" FDIC Working Paper No. 2002-01.

Upper, Christian, and Andreas Worms. 2004. "Estimating Bilateral Exposures in the German Interbank Market: Is There a Danger of Contagion?" *European Economic Review* 48:827–49.

VanLelyveld, Iman, and Frank Liedorp. 2004. "Interbank Contagion in the Dutch Banking Sector." DNB Working Paper Series.

Wells, Simon. 2004. "Financial Interlinkages in the United Kingdom's Interbank Market and the Risk of Contagion." Working Paper, Bank of England.

Measuring Investors' Risk Appetite*

Prasanna Gai and Nicholas Vause
Bank of England

This paper proposes a method for measuring investor *risk appetite* based on the variation in the ratio of risk-neutral to subjective probabilities used by investors in evaluating possible future returns to an asset. Unlike other indicators advanced in the literature, our measure of market sentiment distinguishes risk appetite from *risk aversion*, and is reported in levels rather than changes. Implementation of the approach yields results that respond to crises and other major economic events in a plausible manner.

JEL Codes: G10, G12, G13.

Financial market practitioners often cite market sentiment as a key factor driving broad trends in asset prices. The prices of financial assets frequently move together, even though many of the factors affecting valuations in different asset markets can be quite different. The Asian financial crisis illustrates how shifting attitudes toward risk can generate correlation among the prices of seemingly unrelated assets. Following the devaluation of the Thai baht in July 1997, investors reduced their risk exposures across a range of emerging markets, causing a rise in the cost of borrowing beyond Asia, and into Latin America and Emerging Europe. The spillover of financial stress across borders could not be explained by domestic fundamentals alone and coincided with claims that a decline in "risk appetite" was an underlying reason for contagion and financial instability.

*We thank Alex Bowen, Damien Lynch, Paul Robinson, Hyun Shin, and Peter Westaway for helpful comments and encouragement. The usual caveat applies. The views expressed are those of the authors and do not reflect those of the Bank of England. Corresponding author: Vause: Bank of England, Threadneedle Street, London EC2R 8AH, United Kingdom; e-mail: nicholas.vause@bankofengland.co.uk. Other author contact: Gai: Bank of England, Threadneedle Street, London EC2R 8AH, United Kingdom; e-mail: prasanna.gai@bankofengland.co.uk.

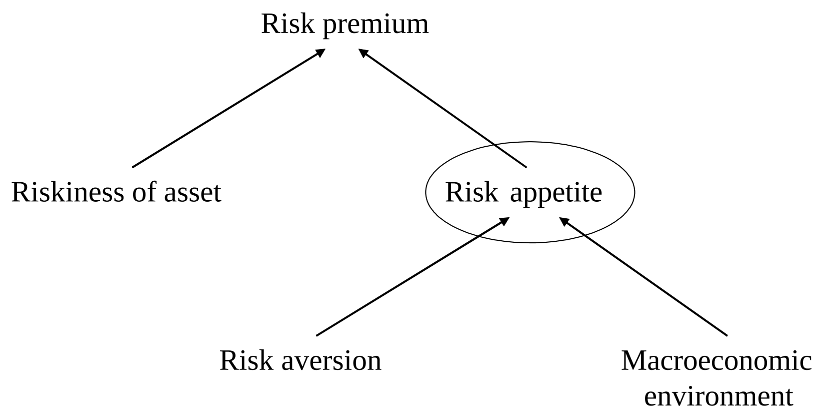
The terms "risk appetite," "risk aversion," and "risk premium" are frequently used interchangeably to refer to sentiment in asset markets. But the concepts are very distinct, and inappropriate use makes it difficult to assess and convey the true extent of the willingness to hold risky assets. Investors dislike uncertainty surrounding the future consumption implied by their asset holdings. *Risk appetite*—the willingness of investors to bear risk—depends on both the degree to which investors dislike such uncertainty *and* the level of that uncertainty. The level of uncertainty about consumption prospects depends on the macroeconomic environment. And the degree to which investors dislike uncertainty reflects underlying preferences over lotteries. This *risk aversion* is part of the intrinsic makeup of investors. It is a parameter that our theoretical priors suggest is unlikely to change markedly, or frequently, over time.[1]

Risk appetite, by contrast, is likely to shift periodically as investors respond to episodes of financial distress and macroeconomic uncertainty. In adverse circumstances, investors will require higher excess expected returns to hold each unit of risk and risk appetite will be low—it is the inverse of the price of risk. And when the price of risk is taken together with the quantity of risk inherent in a particular asset, the expected return required to compensate investors for holding that asset is the *risk premium*. Figure 1 illustrates these concepts. It is clearly difficult to disentangle risk appetite from risk aversion and, as Pericoli and Sbracia (2004) note, an increase in either one of them causes asset prices to decline and risk premia to increase.

In what follows, we formally distinguish risk appetite from risk premia and aversion. Specifically, we propose a measure based on the variation in the ratio of risk-neutral to subjective probabilities used by investors in evaluating the expected payoff of an asset. By exploiting the linkages between the risk-neutral and subjective probabilities that can be extracted from financial market prices, we follow Hayes, Panigirtzoglou, and Shin (2003), Tarashev, Tsatsaronis, and Karampatos (2003), and Bollerslev, Gibson, and Zhou (2004). Unlike these papers, however, we are able to extract an indicator of market sentiment that is quite distinct from risk aversion. Moreover, the

[1]For recent market-based estimates of risk aversion, see Bliss and Panigirtzoglou (2004).

Figure 1. Relationship between Risk Concepts



index of risk appetite based on our approach appears to respond to crises and other economic events in a plausible fashion and, as such, compares favorably with other measures advanced in the literature.

The paper is organized as follows. Section 1 sets out the theoretical basis of our risk appetite measure. Section 2 presents the data and discusses the empirical strategy used to obtain estimates of risk-neutral and subjective probabilities from options prices and presents our measure of risk appetite. We contrast our approach with the recent literature in section 3, and a final section concludes.

1. The Concept of Risk Appetite

The standard treatment of asset pricing theory (e.g., Cochrane 2001) states that in an efficient market, with fully rational and informed investors, the current price of an asset, p_t, should equal the expected discounted value of its possible future payoffs, x_{t+1}. These payoffs comprise income (such as dividend payments) received over the horizon, plus the ongoing value of the asset as implied by its future price. More formally,

$$p_t = \mathbb{E}_t(m_{t+1} \cdot x_{t+1}), \tag{1}$$

where x_{t+1} denotes the payoff in period $t + 1$, and m_{t+1} denotes the discount factor—the marginal rate at which the investor is willing to substitute consumption at time $t + 1$ for consumption at time t.

Both x_{t+1} and m_{t+1} vary across states of the world. Indeed, m_{t+1} is usually referred to as the *stochastic* discount factor.

We ensure that m_{t+1} is unique by assuming that the asset market is complete. This means that it is possible to form portfolios as linear combinations of the assets traded in the market that have positive payoffs in a single state of the world, and otherwise pay zero. Furthermore, it is possible to create many of these portfolios, so that there is a positive payoff for every state. So, if m_{t+1} were not unique, multiple prices would be a possibility for at least one of the portfolios. But this is inconsistent with the absence of arbitrage opportunities that is associated with rational investors. Hence, m_{t+1} is a unique stochastic discount factor that prices all assets.[2]

The basic asset pricing equation can also be expressed in terms of gross returns, R_{t+1}, by dividing equation (1) by current prices. Thus,

$$1 = \mathbb{E}_t(m_{t+1} \cdot R_{t+1}). \tag{2}$$

Although, in general, different assets have different expected returns, all assets have the same expected *discounted* return in equilibrium (of unity). Since both the gross return and the stochastic discount factor are random variables, equation (2) can be written as

$$1 = \underbrace{\mathbb{E}_t(m_{t+1}) \cdot \mathbb{E}_t(R_{t+1})}_{\text{risk-neutral component}} + \underbrace{cov_t(m_{t+1}, R_{t+1})}_{\text{risk adjustment}}. \tag{3}$$

The first term on the right-hand side of equation (3) reflects the mean return required by investors to hold the asset *if* they were indifferent to risk. The second term is a risk correction required by risk-averse investors. Noting that the gross risk-free rate is given by $R^f_{t+1} = 1/\mathbb{E}_t(m_{t+1})$, we can rearrange to obtain the familiar expression

$$\underbrace{\mathbb{E}_t(R_{t+1}) - R^f_{t+1}}_{\text{risk premium}} = -R^f_{t+1} cov_t(m_{t+1}, R_{t+1}). \tag{4}$$

Equation (4) states that the expected return of a risky asset in excess of that available on a risk-free asset is proportional to *minus* the covariance of its state-contingent rate of return and the stochastic

[2]See Danthine and Donaldson (2005, chap. 11) or Milne (1995, chap. 5) for further detail.

discount factor. Intuitively, an asset that pays a high return in good times when investors have a high level of consumption, but fails to pay out in bad times when investors' consumption is lower, has a disadvantageous pattern of returns. So to encourage investors to hold this asset, the expected return must exceed the risk-free rate, i.e., the asset must offer a *risk premium*.

The risk premium can, in turn, be decomposed into the quantity of risk, β_i, inherent in each asset and the unit price of risk that is common across assets, λ_t. In particular,

$$\mathbb{E}_t(R_{t+1}) - R^f_{t+1} = \underbrace{\frac{-cov_t(m_{t+1}, R_{t+1})}{var(m_{t+1})}}_{\beta_i} \cdot \underbrace{var(m_{t+1}) \cdot R^f_{t+1}}_{\lambda_t}. \quad (5)$$

The price of risk, λ_t, is the expected excess return that investors require to hold each unit of risk in equilibrium. *Risk appetite*—the willingness of investors to bear risk—can therefore be defined as the inverse of the price of risk. So when risk appetite falls, larger expected excess returns are required to hold risky assets.

It is apparent from equation (5) that risk appetite reflects variation in the stochastic discount factor, $var(m_{t+1})$. Since the stochastic discount factor specifies the marginal rate at which the investor is willing to substitute uncertain future consumption for present consumption, risk appetite depends on the *degree* to which investors dislike uncertainty about their future consumption and on factors that determine the overall *level* of uncertainty surrounding consumption prospects. The degree of such uncertainty corresponds to *risk aversion*, since the more risk averse the investor, the more valuable is additional income in bad states of the world. Accordingly, risk aversion reflects innate preferences over uncertain future consumption prospects—the curvature of individuals' utility functions—that are unlikely to vary significantly over time.

The factors underpinning risk appetite can be seen more clearly by imposing some structure on the stochastic discount factor. In particular, if consumption growth is log-normally distributed with variance, $\sigma^2_t(c_{t+1})$, and investors have power utility functions, then the price of risk is

$$\lambda_t = \gamma\sigma^2_t(c_{t+1}), \quad (6)$$

where γ is the coefficient of absolute risk aversion.[3] So a rise in γ would mean a fall in risk appetite. But risk appetite will also fall if uncertainty about future consumption growth increases. The expected volatility of future consumption is likely to depend on factors such as unemployment prospects, the stance of macroeconomic policy, and so on. In general, one would expect that the periodic shifts in market sentiment witnessed over time are more likely to be driven by the macroeconomic environment rather than by changes in the risk aversion of investors.

The analysis of asset pricing above is couched in terms of investors' subjective probabilities about various states of the world. But the risk aversion of investors—their tendency to value more highly assets that produce high payoffs in bad states—means that the expected payoff of an asset can also be evaluated using a set of adjusted probabilities. These adjusted probabilities are risk neutral, as by assigning greater weight to undesirable states they generate the same utility for a risk-neutral investor as for a risk-averse investor with the original probabilities. As discussed in section 2 below, these adjusted probabilities can be inferred from the prices of options contracts on the underlying asset.

Assets can, therefore, be priced by (i) evaluating the expectation of discounted payoffs using investors' best guesses of the probabilities of different states of the world occurring or, equivalently, by (ii) discounting payoffs by the risk-free rate and evaluating expectations using a set of adjusted probabilities. If there are S possible future states of the world, indexed by $s = 1, 2, 3, ...S$, then the expected discounted return of an asset can be expressed either as the sum of the discounted returns in each state, weighted by investors' subjective probability of the state occurring,

$$1 = \mathbb{E}_t(m_{t+1} \cdot R_{t+1}) = \sum_{s=1}^{S} m_{t+1}(s) \cdot R_{t+1}(s) \cdot \pi_{t+1}(s), \qquad (7)$$

[3] This is a standard result in asset pricing. See Cochrane (2001) for a detailed exposition. Asset pricing models that employ these restrictions do, however, significantly underestimate the risk premia observed in practice due to the low volatility of consumption. Models with less restrictive utility functions and, hence, stochastic discount factors that depend on a broader set of variables may help to reconcile such anomalies (see, for example, Barberis, Huang, and Santos 2001).

or in terms of risk-neutral probabilities ($\pi^*_{t+1}(s)$), discounted with the risk-free interest rate,

$$1 = \mathbb{E}_t(m_{t+1}) \cdot \mathbb{E}^*_t(R_{t+1}) = \sum_{s=1}^{S} \frac{1}{R^f_{t+1}} \cdot R_{t+1}(s) \cdot \pi^*_{t+1}(s). \tag{8}$$

Taken together, equations (7) and (8) imply that the ratio of the risk-neutral to subjective probabilities is proportional to the stochastic discount factor, where the constant of proportionality is given by the gross risk-free rate of return, i.e.,

$$\frac{\pi^*_{t+1}(s)}{\pi_{t+1}(s)} = m_{t+1}(s) \cdot R^f_{t+1}. \tag{9}$$

Note that the risk-neutral probability distribution is *pessimistic* in the sense that it assigns excessive probability to low-income states and too little probability to high-income states. The mean of the risk-neutral density is given by $R^f_{t+1} = 1/\mathbb{E}_t(m_{t+1})$, whereas the mean of the subjective density is given by equation (2). The difference between the two means is therefore the risk premium.

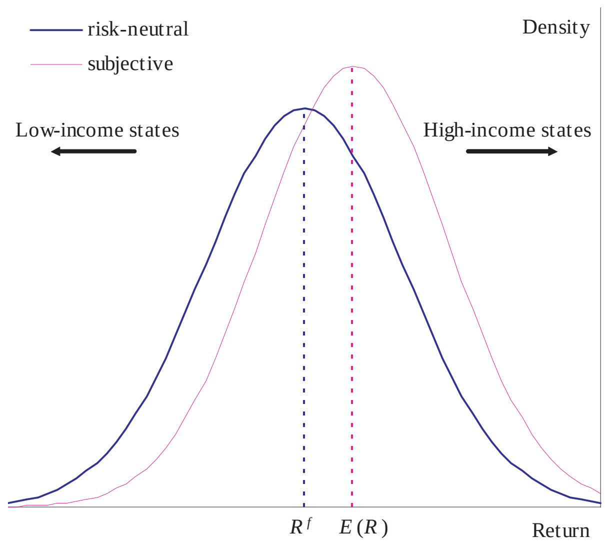
Investors' risk aversion also enters the risk-neutral probabilities. Since risk-averse investors value additional income more highly in poor states of the world, low-income states receive an increased weight when computing the expected return of an asset using the risk-neutral asset pricing relationship. When the marginal utility of consumption is high in state s, the risk-neutral probability is greater than the true probability and vice versa. Figure 2 provides a stylized illustration of the two probability distributions.

An increase in the ratio between the risk-neutral and subjective probabilities may therefore reflect either an increase in risk aversion or changes in other state variables that increase the marginal utility of consumption. As we have seen, the willingness of the investor to pay for insurance across such states—the investor's risk appetite—depends on the variance of the stochastic discount factor across states of the world. It follows from equations (5) and (9) that

$$\lambda_t = \frac{1}{R^f_{t+1}} \cdot var\left(\frac{\pi^*_{t+1}(s)}{\pi_{t+1}(s)}\right) \tag{10}$$

is a measure of risk appetite, once the two probability densities over future returns are derived.

Figure 2. Risk-Neutral and Subjective PDFs



2. Estimating Risk Appetite

Our analysis suggests that a measure of risk appetite may be derived by computing the variation in the ratio of risk-neutral to subjective probabilities used by investors in evaluating the expected payoff of an asset. This requires estimating two probability density functions over future returns—one risk-neutral distribution and one subjective distribution—on an index such as the S&P 500. To generate a time series for risk appetite, these distributions are estimated every three months, at the end of each quarter. As the return forecasts for the end of a particular quarter are made at the end of the previous quarter, the corresponding estimate of risk appetite would also be for the previous quarter. In what follows, we outline the approach used in estimating these distributions.

2.1 Risk-Neutral Densities

Option prices offer a forward-looking guide to the likelihood investors attach to future values of asset prices. But it is only a guide, because the price that investors will pay for an option depends both on their subjective beliefs about the relative likelihoods of returns having particular future values and on their attitude to risk. If investors were neutral toward risk, however, option prices would only reflect expectations about returns. So, by comparing options with different strike prices, we can infer the risk-neutral probabilities attached by market participants to an asset being within a range of possible prices at some future date. Indeed, the whole risk-neutral density function can be inferred from the prices of marketed options using the no-arbitrage argument of Breeden and Litzenberger (1978), who demonstrate that the density function is the second derivative of the option price with respect to the option strike.

We use risk-neutral density functions for the S&P 500 index constructed by the Bank of England. These are estimated by a two-step procedure. The first step is to estimate a call price function, which shows how the prices of call options with identical maturities vary as strike prices change. This is achieved by applying a cubic-spline interpolation to the available data on pairs of call and strike prices. For more robust results, the interpolation is actually applied to transformations of these prices (see Clews, Panigirtzoglou, and Proudman [2000] for details) and the resulting function is converted back into a smooth and continuous relationship between call and strike prices. The second step is then to twice differentiate the resulting call price function. As demonstrated by Breeden and Litzenberger, this delivers the density function of the underlying asset based on the assumption that investors are risk neutral.

2.2 Subjective Densities

Estimation of the subjective probability distributions of returns follows the approach of Hayes, Panigirtzoglou, and Shin (2003).[4] It is

[4] An alternative approach is suggested by Bliss and Panigirtzoglou (2004), who estimate the subjective probability by hypothesizing a specific utility function for a representative agent and then using it to convert the estimated risk-neutral density function into a subjective density using the method suggested by Ait-Sahalia and Lo (2000). As Bliss and Panigirtzoglou observe, knowledge of any two of three functions—the risk-neutral density, the subjective density, and the utility function—allows the third to be inferred. So it is not immediately obvious whether this alternative is superior to the approach suggested by Hayes, Panigirtzoglou, and Shin (2003).

based on the following threshold-GARCH model of returns on the S&P 500 index, r_t:

$$\begin{aligned} r_t &= c + x'_{t-1}\theta + \delta\sigma_t + \varepsilon_t, \\ \sigma_t^2 &= \omega + y'_{t-1}\lambda + \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 d_{t-1} + \beta\sigma_{t-1}^2 \\ d_t &= 1 \quad \text{if} \quad \varepsilon_t > 0, \text{ and } 0 \text{ otherwise,} \end{aligned}$$

where x_t and y_t are vectors of explanatory variables and σ_t^2 is the variance of the residuals, ε_t.[5]

The shape of the subjective density of returns is equated to the shape of the density of the standardized residuals, ε_t/σ_t. But to construct the precise subjective density of one-quarter-ahead returns, the variance of the density of the standardized residuals is multiplied by the forecast conditional variance, σ_{t+1}, and the mean of the resulting density is set equal to a particular value. In principle, this value could be the forecast conditional mean, r_{t+1}, but in practice this occasionally implies that the mean of the subjective density is smaller than the mean of the risk-neutral density, i.e., that the risk premium is negative, which seems implausible for an equity index. Instead, we locate each subjective density such that the difference between its mean and the mean of the corresponding risk-neutral density is equal to the value of the equity risk premium implied by the Bank of England's three-stage dividend discount model.[6]

The threshold-GARCH model is initially estimated using quarterly data from 1920:Q1 to 1983:Q1. The fitted model is then used to forecast the conditional variance in 1983:Q2, which, as noted above, is used to construct the subjective density of returns in 1983:Q2. The model is then reestimated using data from 1920:Q1 to 1983:Q2 and the new model is used to forecast the conditional variance and hence construct the subjective return density in 1983:Q3, and so on.

[5]Hence, the modeled variance of returns depends on previous errors in modeling the level of returns. So, extreme returns that are not fully captured by the GARCH model would generate large residuals, and these would affect the subsequent modeled variance of return via the parameter β_4. Furthermore, extreme returns can have differential effects on the modeled variance of returns depending on whether they are extremely high or extremely low and hence whether residuals are positive or negative. The scale of the difference is governed by the parameter β_5. Finally, the variance of returns is postulated to exhibit some persistence according to β_6.

[6]See Panigirtzoglou and Scammell (2002).

Variables included in x_t and y_t were selected by adopting a general-to-specific modeling approach. From an initial list comprised of the natural logarithm of the dividend yield on the S&P 500, the spread between the yield on BBB- and AAA-rated U.S. corporate bonds, the yield on three-month U.S. Treasury bills, the term spread between the yields on ten-year U.S. government bonds and three-month U.S. Treasury bills, the rate of commodity price growth according to the Commodity Research Bureau, U.S. consumer price inflation, and the rate of unemployment in the United States,[7] the natural logarithm of the dividend yield was selected as the only variable to include in x_t, while y_t was selected to be empty. Variables were selected by deleting any found to be insignificant in the most general specification and reestimating the model until only statistically significant variables remained. This choice of variables also optimized the Hannan-Quinn information criterion.

The parameter estimates for the preferred model, estimated over 1920:Q1 to 1983:Q1, are reported in the first column of table 1. These estimates are quite stable as the sample period is lengthened. The positive coefficient associated with the logarithm of the dividend yield implies that returns tend to be low when prices are high relative to dividends. This generates a degree of mean reversion in the dividend yield that is consistent with the findings of empirical finance. Returns are also found to vary positively with their standard deviation. This is consistent with theoretical models in which risk and expected returns are positively associated.

The conditional variance equation generates three further features of equity returns that are commonly found in empirical work: fat tails, negative skewness, and volatility clustering. The ARCH term (ε_{t-1}^2), which has a positive coefficient, means that a significant shock to returns will boost the conditional variance, so that extreme returns are more likely to follow an initial extreme return than an initial moderate return. This increases the thickness of tails in the distribution of returns. The threshold-ARCH term ($\varepsilon_{t-1}^2 d_{t-1}$), which also has a positive coefficient, implies that negative shocks are more

[7]The choice of variables was motivated by the literature on equity return predictability, e.g., Lamont (1998) and Kothari and Shanken (1992). Also, variables such as inflation, the unemployment rate, and commodity price growth were included to capture potential business cycle effects on returns and their variability (see, for example, Chen, Roll, and Ross 1986).

Table 1. Coefficient Estimates of GARCH Model

	1920:1 to 1983:1	**1920:1 to 1983:1**	**1920:1 to 1983:1**	**1945:4 to 1983:1**
c	−0.0702 (0.0302)	−0.0241 (0.0172)	−0.0782 (0.0282)	−0.0475 (0.0356)
θ	0.0436 (0.0209)		0.0419 (0.0186)	0.0314 (0.0235)
δ	0.3989 (0.2041)	0.4820 (0.2077)	0.4261 (0.1949)	0.2531 (0.4113)
ω	0.0017 (0.0007)	0.0020 (0.0007)	0.0014 (0.0006)	0.0016 (0.0006)
α	0.0997 (0.0813)	0.0803 (0.0888)	0.2911 (0.0913)	−0.0481 (0.0955)
γ	0.2836 (0.1313)	0.3224 (0.1403)		0.3156 (0.1303)
β	0.5611 (0.1329)	0.5355 (0.1465)	0.5755 (0.1186)	0.6120 (0.1555)

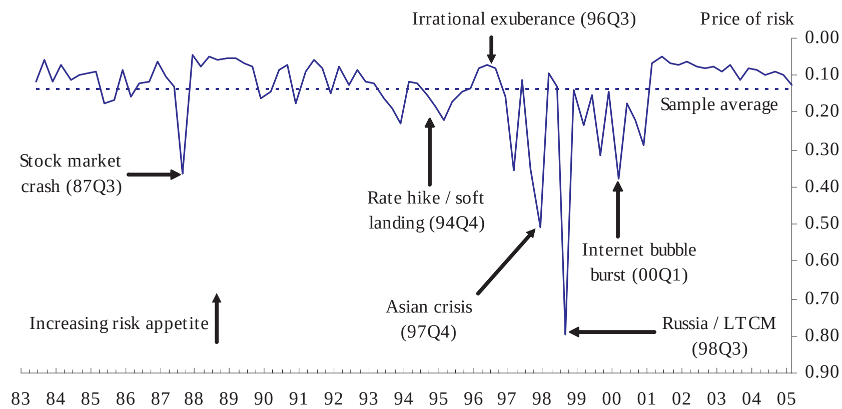
Note: Standard errors are shown in parentheses.

likely to be followed by high volatility than positive shocks. This generates negative skewness. Finally, the GARCH term ($\beta_6\sigma_{t-1}^2$), with its positive coefficient, generates persistence in volatility, resulting in clusters of high and low volatility.

2.3 Comparing the Two Densities

We plot histograms of the estimated risk-neutral and subjective densities. For each bin of the histograms, we compute the ratio of π^*/π, as required by equation (10). Due to inaccuracy of estimation, however, this ratio is sometimes spuriously high in the tails of the histograms. Therefore, any bins for which $\pi^*/\pi > 10$ are dropped from the histograms, which are subsequently rescaled so that the probabilities of the various feasible returns continue to sum to unity. Finally, risk appetite is computed in accordance with equation (10), with the yield on three-month U.S. Treasury bills serving as a proxy for the risk-free rate.

Figure 3. Estimated Risk Appetite



2.4 The "Variance" Measure

Figure 3 shows the quarterly time series of risk appetite from our estimation procedure. The illustrated series fluctuates close to its average for most of the time, but has occasional sharp downward movements. The sharp downward movements coincide with the 1987 stock market crash, the Asian financial crisis, the Russian/LTCM crisis, and the Internet stock crash. The series suggests that investors' risk appetite is likely to be fairly stable during "tranquil" periods, but move sharply in response to exogenous shocks. More recently, investors' appetite for risk has been strong, above the sample average and at levels comparable to those of 1996 when Alan Greenspan spoke of irrational exuberance. Of course, the true path of investors' risk appetite remains unobserved, but the behavior of the measure during the period in question (1983–2005) seems plausible.[8]

3. Comparison with Existing Approaches

A number of recent papers have also attempted to measure market sentiment. A first approach is based on changes in excess returns. Equation (5) showed how the excess return required by investors to

[8]The appendix investigates the robustness of the risk appetite measure to changes in the assumptions made in its derivation.

hold an asset depends on the level of risk inherent in the asset and the risk appetite of the investor. Kumar and Persaud (2002) propose a measure of risk aversion based on the distribution of excess returns across assets. Their hypothesis is that when risk appetite increases, excess returns of very risky assets increase by more than for less-risky assets. In contrast, changes in the overall level of risk across assets should not have a differential impact on expected returns. Thus, the degree of correlation between changes in excess returns and the level of risk across a number of assets should indicate any change in risk appetite.[9]

There are a number of difficulties with this measure, however. First, the measure only indicates changes in risk aversion and does not suggest what its level might be. Second, the measure does not give an indication of the magnitude of the change in risk aversion. The rank correlation is theoretically unity when risk aversion is driving returns and zero when changing risk is driving returns. And finally, a rank correlation may be detected even when risk aversion is constant, if the level of risk associated with different assets changes to differing degrees. For example, if the volatility of the market return increased, this would increase the risk of some assets more than others and lead to a rank correlation.

A second approach, emphasized by Tarashev, Tsatsaronis, and Karampatos (2003) and Hayes, Panigirtzoglou, and Shin (2003), focuses on a comparison of the risk-neutral and subjective probability densities.[10] They interpret the ratio on the left-hand side of equation (9), evaluated at a particular percentile, as an indicator of risk aversion. As we have argued, however, the stochastic discount factor generally reflects rather more than just investor preferences. So movements in the probability ratio over time are more likely to reflect factors other than risk aversion. Recognizing this shortcoming,

[9]See Misina (2003) and Pericoli and Sbracia (2004) for a reconciliation of the Kumar and Persaud measure with the general asset pricing framework outlined above.

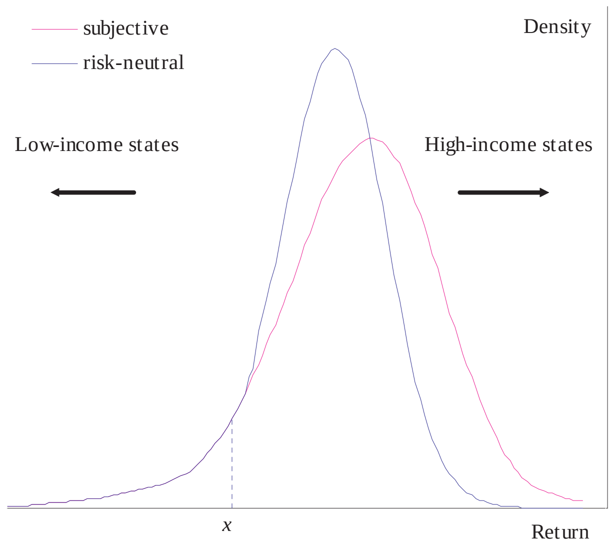
[10]See also Scheicher (2003). Jackwerth (2000) also uses the probability ratio to obtain a function for risk aversion that can be computed from option contracts on the market portfolio. But his approach has two drawbacks. First, the risk aversion function can take on negative values in some states of the world, suggesting that risk aversion may (on occasion) increase with increasing wealth. And second, the risk aversion schedule does not allow a measure of market sentiment to be readily tracked over time.

Hayes, Panigirtzoglou, and Shin suggest that movements in the ratio might reflect investors' concerns about liquidity. Their hypothesis is that investors discount asset returns less heavily when their wealth is illiquid because it is more difficult to support consumption from retained wealth in such circumstances. They suggest that the importance of illiquidity in the stochastic discount factor is greatest in bad states of the world that are characterized by low asset returns. This is supported by the fact that, in such states, there is a positive relationship between implied volatilities (which tend to increase when market liquidity falls) and the estimated probability ratio. But in other states of the world, a better indication of risk aversion may be obtained since the liquidity factor is less likely to be important.

A further drawback of such an approach is that, by estimating the stochastic discount factor at a particular percentile, a "ratio" measure can misrepresent investors' overall attitude to risk. By contrast, our "variance" measure uses estimates of the stochastic discount factor across many states of the world, in which asset returns differ. If the subjective and risk-neutral distributions differ in shape markedly, then using all the information contained in the distributions is likely to offer a more reliable indicator of sentiment. For example, a ratio measure evaluated at a point like x in figure 4 would suggest that investors were risk neutral, as the tails of the risk-neutral and subjective densities coincide. As the densities diverge away from the left tail, however, the variance measure would suggest that investors disliked risk.

An approach that is very close in spirit to our own is that of Bollerslev, Gibson, and Zhou (2004), who essentially compare estimates of the standard deviations (or volatilities) of the risk-neutral distribution and the subjective distribution, rather than the whole distributions. The difference between the two standard deviations reflects a "volatility risk premium." The higher the risk appetite, the smaller the degree to which implied (risk-neutral) volatilities derived from option prices will exceed realized (subjective) volatilities. An advantage of their approach is in its use of model-free volatilities. The computation of implied volatility does not rely on the accuracy of the Black-Scholes model, for example, while the subjective volatility is computed using high-frequency historical data without imposing a GARCH model. A further advantage is that the authors are able to relate their measure of risk appetite to macroeconomic factors. One

Figure 4. Importance of Using Whole Densities



of their key findings is that risk appetite appears to increase with industrial production, which supports the notion advanced earlier that risk appetite is positive related to the macroeconomic environment that investors inhabit. A potential disadvantage of this study, however, is that by focusing only on the standard deviations of the risk-neutral and subjective distributions, and ignoring their higher moments, an incomplete picture of risk appetite may be obtained.

A final approach to measuring risk appetite relies on cross-border portfolio flows (Froot and O'Connell 2003). By assuming that investors have constant absolute risk aversion (CARA) utility functions, the authors show that each investor's demand for a risky asset will depend on the investor's wealth, the variance of the risky asset's excess returns, the covariance of these excess returns with the excess returns to other risky assets, as well as on the risk aversion parameter. Investors are then divided into two categories: international investors, who can purchase all assets, and domestic investors, who can only purchase the asset of the market that they inhabit. Froot and O'Connell show that cross-border portfolio flows will reflect only the risk aversion of international investors *relative to* the risk aversion of domestic investors. Using data on cross-border portfolio flows, they infer the relative measure of risk aversion in the form of an "investor confidence index." By contrast, our approach provides an absolute measure of risk appetite, rather than a measure of relative risk aversion that relies on particular utility functions.

4. Conclusion

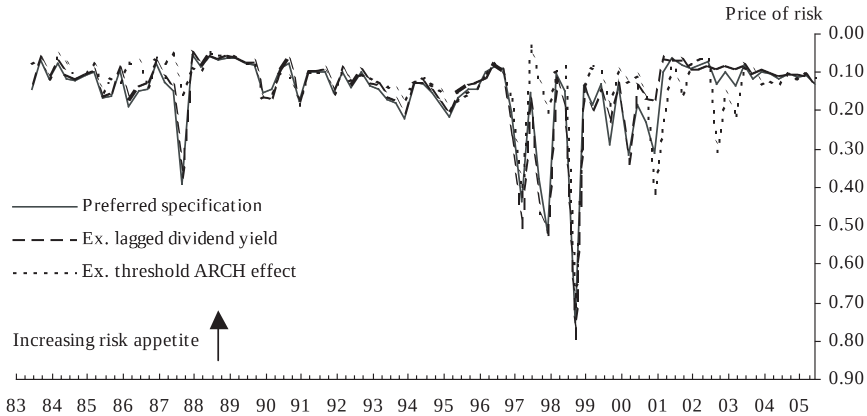
This paper has proposed a measure of market sentiment that is distinct from risk aversion and can be used to gauge how investors' appetite for risk evolves over time. The empirical analysis suggests that a measure based on the variation in the ratio of risk-neutral to subjective probabilities, derived from equity index option prices, appears to generate results that conform to intuition—the measure responds to major financial events in a plausible manner. Our approach has a number of advantages over existing measures of market sentiment. In particular, it does not rely on restrictive assumptions on investor preferences and it uses all the available information in the risk-neutral and subjective probability distributions.

Appendix. Robustness of Risk Appetite Measure

This appendix investigates the robustness of the risk appetite measure to changes in the assumptions made in its derivation. In particular, we investigate the effects of changing (i) the specification of the GARCH model used in constructing the subjective density of returns, (ii) the sample period over which the coefficients of the GARCH model are estimated, (iii) the estimates of the risk premium used to separate the means of the risk-neutral and subjective densities, and (iv) the cutoff point at which we reject our estimates of the ratio of the risk-neutral probability to the subjective probability.

First, we adapt our specification of the GARCH model, dropping the lagged dividend yield (i.e., setting $\theta = 0$) and the threshold ARCH effect (i.e., setting $\gamma = 0$) from our preferred specification. Regression results for these two modifications are respectively displayed in columns 2 and 3 of table 1, shown previously in section 2.2. Coefficient estimates are quite similar to those of our preferred specification. As a result, the risk appetite measure derived using these alternative GARCH specifications is also quite similar to our preferred measure of risk appetite, as can be seen in figure 5.

Figure 5. Robustness of Risk-Appetite Measure to Alternative GARCH Specifications



Second, we change the sample period over which we estimate the coefficients of our GARCH model to the postwar interval of 1945:Q4 to 1983:Q1. Regression results are displayed in the final column of table 1. Again, the coefficient estimates are similar to those of our preferred specification and the resulting profile of risk appetite is broadly similar to that of our preferred measure, although the correlation between the two does fall in the last few years of the sample (see figure 6). The gap between the measure based only on postwar data and our preferred measure of risk appetite that emerges at certain times is attributable to a change in the estimated shape of the subjective density. As the shape of the subjective density is constructed from the GARCH residuals, it is affected by the change of sample period. In particular, some probability mass is removed from the left tail of the subjective density due to the exclusion of the 1929 crash from the data sample. This results in higher ratios of risk-neutral probabilities to subjective probabilities and, hence, higher estimates of the price of risk.

Third, we change the estimates of the equity risk premium used to separate the means of the risk-neutral and subjective densities from the time-varying estimates obtained from the Bank of England's discounted dividend model to a constant estimate of 3.3 percent, which is taken from Taylor (2005). The latter is an estimate of the average equity risk premium on the S&P 500 since the beginning of the 1980s.

Figure 6. Robustness of Risk-Appetite Measure to Alternative GARCH Estimation Periods

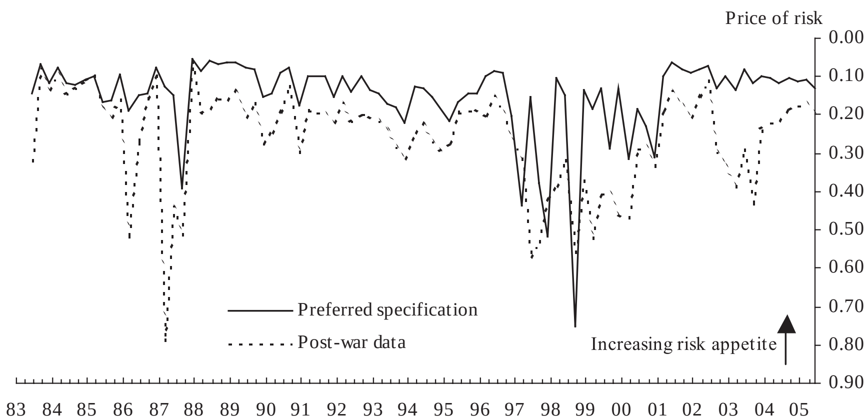
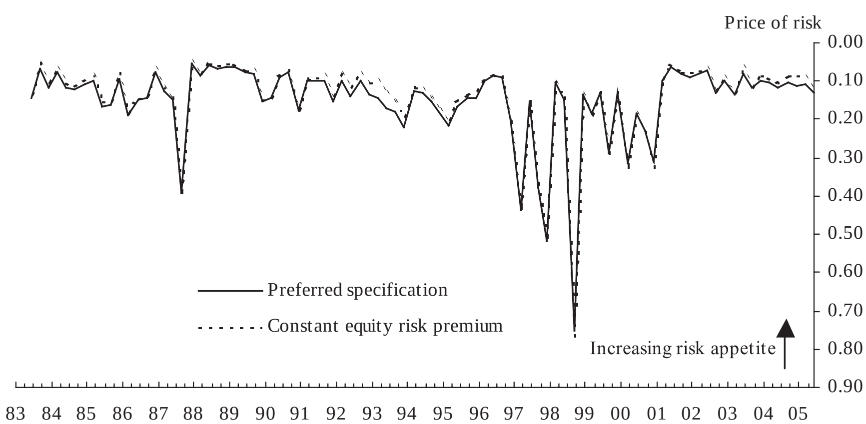


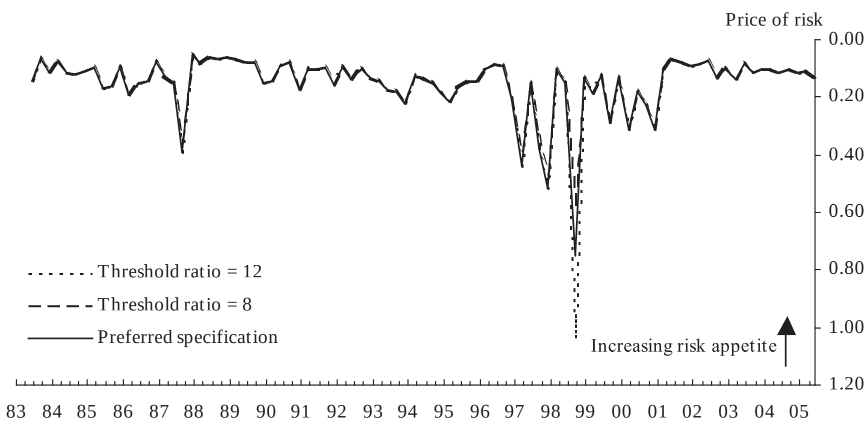
Figure 7. Robustness of Risk-Appetite Measure to Alternative Estimates of the Equity Risk Premium



As figure 7 indicates, the risk appetite measure is highly robust to alternative estimates of the equity risk premium, with constant and time-varying estimates producing very similar profiles.

Finally, we experiment by changing the threshold above which we reject our estimates of the ratio of risk-neutral probability to subjective probabilities. These ratios are occasionally found in the

Figure 8. Robustness of Risk-Appetite Measure to Alternative Threshold Ratios



tails of return distributions, where errors can result in subjective probability estimates that are very close to zero. This produces very high estimates of the ratio of risk-neutral probability to subjective probability. As the risk appetite measure is derived from variation in this ratio across the estimated probability distributions, it could potentially become driven by spuriously high ratios in the tails of the distributions. Hence, our preferred measure of risk appetite is computed by omitting any ratio estimates greater than ten from the variance calculation of equation (10). Figure 8 shows the effect of varying this threshold. The threshold appears to affect the degree to which crisis periods stand out as episodes of low risk appetite, while leaving the broad profile of the series essentially unchanged.

References

Ait-Sahalia, Yacine, and Andrew W. Lo. 2000. "Non-parametric Risk Management and Implied Risk Aversion." *Journal of Econometrics* 94 (1–2): 9–51.

Barberis, Nicholas, Ming Huang, and Tano Santos. 2001. "Prospect Theory and Asset Prices." *Quarterly Journal of Economics* 116 (1): 1–53.

Bliss, Robert R., and Nikolaos Panigirtzoglou. 2004. "Option-Implied Risk Aversion Estimates." *Journal of Finance* 59 (1): 407–46.

Bollerslev, Tim, Michael Gibson, and Hao Zhou. 2004. "Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities." Finance and Economics Discussion Series Working Paper No. 2004-56, Board of Governors of the Federal Reserve System.

Breeden, Douglas T., and Robert H. Litzenberger. 1978. "Prices of State-Contingent Claims Implicit in Options Prices." *Journal of Business* 51:621–51.

Chen, Nai-Fu, Richard Roll, and Stephen A. Ross. 1986. "Economic Forces and the Stock Market." *Journal of Business* 59:383–403.

Clews, Roger, Nikolaos Panigirtzoglou, and James Proudman. 2000. "Recent Developments in Extracting Information from Options Markets." *Bank of England Quarterly Bulletin* (February):50–60.

Cochrane, John H. 2001. *Asset Pricing.* Princeton, NJ: Princeton University Press.

Danthine, Jean-Pierre, and John B. Donaldson. 2005. *Intermediate Financial Theory.* Elsevier Academic Press.

Froot, Kenneth, and Paul O'Connell. 2003. "The Risk Tolerance of International Investors." NBER Working Paper No. 10157.

Hayes, Simon, Nikolaos Panigirtzoglou, and Hyun Song Shin. 2003. "Liquidity and Risk Appetite: Evidence from Equity Index Option Prices." Mimeo, Bank of England.

Jackwerth, Jens. 2000. "Recovering Risk Aversion from Option Prices and Realized Returns." *Review of Financial Studies* 13 (2): 433–51.

Kothari, S. P., and Jay Shanken. 1992. "Stock Return Variation and Expected Dividends: A Time-Series and Cross-Section Analysis." *Journal of Financial Economics* 31 (2): 177–210.

Kumar, Manmohan, and Avinash Persaud. 2002. "Pure Contagion and Investors' Shifting Risk Appetite: Analytical Issues and Empirical Evidence." *International Finance* 5 (3): 401–26.

Lamont, Owen. 1998. "Earnings and Expected Returns." *Journal of Finance* 53 (5): 1563–87.

Milne, Frank. 1995. *Finance Theory and Asset Pricing.* Oxford University Press.

Misina, Miroslav. 2003. "What Does the Risk Appetite Index Measure?" Working Paper No. 2003-23, Bank of Canada.

Panigirtzoglou, Nikolaos, and Robert Scammell. 2002. "Analysts Earning Forecasts and Equity Valuations." *Bank of England Quarterly Bulletin* (Spring):59–66.

Pericoli, Marcello, and Massimo Sbracia. 2004. "The CAPM and the Risk Appetite Index: Theoretical Differences and Empirical Similarities?" Mimeo, Bank of Italy.

Scheicher, Martin. 2003. "What Drives Investor Risk Aversion? Daily Evidence from the German Equity Market." *BIS Quarterly Review* (June):67–74.

Tarashev, Nikola, Kostas Tsatsaronis, and Dmitrios Karampatos. 2003. "Investors' Attitude Towards Risk: What Can We Learn from Options?" *BIS Quarterly Review* (June):57–66.

Taylor, Bryan. 2005. "The Equity Risk Premium." Mimeo, Global Financial Data.