Federal Reserve Bank of New York

Economic Policy Review

April 1999
Volume 5 Number 1

The Economic Policy Review *is published by the Research and Market Analysis Group of the Federal Reserve Bank of New York. The views expressed in the articles are those of the individual authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.*

Federal Reserve Bank of New York

Economic Policy Review Advisory Board

Federal Reserve Bank of New York
Economic Policy Review

Table of Contents

April 1999
Volume 5 Number 1

ARTICLES

41 EXCHANGE RATES AND PROFIT MARGINS: THE CASE OF JAPANESE EXPORTERS
Thomas Klitgaard

When exchange rates shift, exporters must decide whether it is more important to maintain profit margins or to maintain stable export prices. This examination of Japanese exporters finds that these firms have taken a middle course: By altering their profit margins to some degree, the exporters moderate the exchange-rate-induced changes in prices seen by their foreign customers. The analysis finds that in the three major exporting industries—industrial machinery, electrical machinery, and transportation equipment—a 10 percent rise in the yen leads firms to lower profit margins on exports by 4 percent relative to the margins on their sales in Japan. That is, the exporters pass on more than half of any change in the yen to the price seen by their foreign customers and absorb the remainder by adjusting profit margins on foreign sales.

Supervisory Information and the Frequency of Bank Examinations

Beverly J. Hirtle and Jose A. Lopez

Bank supervisors need timely and reliable information about the financial condition and risk profile of banks in order to conduct effective supervision. Although such information can be obtained in part from regulatory reports and public disclosures, a key source is the on-site bank examination. Bank examinations enable supervisors to confirm the accuracy of information in regulatory reports. More important, perhaps, the examinations allow supervisors to gather confidential information about banks' financial conditions and to assess qualitative attributes, such as internal controls and risk management procedures, that affect bank risk profiles.

Such confidential information may be valuable to supervisors, yet it is costly to obtain: bank examinations absorb considerable resources on the part of supervisors as well as banks. Thus, supervisors face a trade-off between the timeliness of the information gathered from bank examinations and the costs of obtaining it. In particular, the longer the time since a bank's most recent examination, the higher the likelihood that conditions at the bank will have changed in a way that diminishes the current value of that information. This concern must be balanced against the costs of conducting more frequent examinations.

The potential "time decay" of bank examination information has been a concern for both supervisors and policymakers. Supervisors have developed a number of approaches for allocating scarce examination resources, including off-site monitoring systems to help identify banks whose financial conditions may have deteriorated. Concern about the timeliness of examination information also motivated provisions in the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA), which mandates annual on-site examinations for most commercial banks. In this case, legislators felt that frequent examinations would be useful in limiting losses to the deposit insurance system.

Beverly J. Hirtle is a vice president at the Federal Reserve Bank of New York. Jose A. Lopez, formerly an economist at the Federal Reserve Bank of New York, is now an economist at the Federal Reserve Bank of San Francisco.

In this article, we provide insight into the policy aspects of this informational time decay by assessing how the length of time between bank examinations affects the quality of information available to supervisors. For these purposes, we define the quality of information in terms of how accurately the information from a prior examination reflects the current condition of a bank. Our analysis assumes that examination information incorporates two types of data: information available from public sources and updated regulatory reports, and private supervisory information gathered from on-site examinations. That is, examination findings contain information that is readily available from public financial and regulatory reports as well as private information that can be obtained only through on-site examinations, such as confidential information about a bank's troubled assets or the examiner's assessment of a bank's internal controls. Changes over time in the quality of the examination information will be affected by changes in both of these components.

Because the information in regulatory reports is readily available to supervisors, an on-site examination is not required to update this component.[1] However, supervisory information can be acquired only through an examination. Thus, the rate of decay in the accuracy of this private supervisory information should be the key determinant in the timing of bank examinations. The faster this information decays over time, the more frequent these examinations must be to ensure that the quality of information does not drop below some critical level.

Our analysis suggests that the private supervisory component of examination information ceases to provide useful information about the current condition of a bank after six to twelve quarters, or one and a half to three years. This rate of information decay seems to be cyclical, in that the quality of this private supervisory information appears to decay faster during years in which the U.S. banking industry experiences financial difficulties. Consistent with this finding, our analysis further suggests that the decay rate depends on the initial financial condition of the bank: the rate of decay in the quality of private supervisory information appears to be significantly greater for troubled banks than for healthy ones. This latter result is consistent with the idea that conditions change more rapidly at institutions experiencing financial difficulty and that more frequent examinations of these institutions may be warranted.

Our results provide insight into how often a bank should be examined. The range of six to twelve quarters indicated by our results is really an upper bound, since it reflects the point at which the supervisory information from the previous examination contains *no* useful information about the current condition of the bank.[2] As a matter of practice, it is probably desirable to examine banks somewhat more frequently—that is, when the supervisory information from the previous examination still contains *some* useful information about the bank's current condition. Our results also suggest that more frequent examinations may be warranted during periods of banking industry stress and for banks that are financially troubled. Taken together, these results imply that the annual examination frequency mandated in FDICIA is reasonable, particularly during times of financial difficulties for the banking industry.

The Timing and Frequency of Bank Examinations

The timing and frequency of bank examinations have received increased public scrutiny in recent years, especially in light of the significant loan losses and number of bank failures suffered by the U.S. banking industry during the 1980s and early 1990s.[3] Debate has centered around the idea that supervisors, banks, and the tax-paying public face a trade-off between the costs and benefits of more frequent examinations. On the one hand, more frequent examinations would generate more timely information about the current condition of banks and could allow supervisors to address emerging problems more quickly. This quicker response could reduce the exposure of the deposit insurance system—and ultimately the taxpayer—to loss. On the other hand, examinations are resource-intensive for both banks and supervisors, and maintaining large supervisory and examination staffs can be costly.[4]

The balance of this trade-off has shifted over the years in response to conditions in the banking industry. For instance, the Federal Deposit Insurance Corporation recently reported on the efforts of the federal bank supervisory agencies to extend the examination cycle as a means of reducing the size of their examination staffs, especially during the early-to-mid-1980s (Federal Deposit Insurance Corporation 1997). According to the report, these agencies focused their resources on the institutions most likely to present systemic risk concerns. In many cases, targeted or limited-scope examinations—that is, examinations that assess only selected bank activities or that involve less detailed evaluations of a bank's overall activities, respectively—were used in place of more resource-intensive, full-scope examinations. As a result, the frequency of full-scope examinations fell considerably during this period, especially for smaller banks and banks believed to be in sound financial condition. Taken together, such measures in particular allowed the FDIC and the Office of the Comptroller of the Currency to reduce their examination staffs.

The FDIC report concluded that such cutbacks in examination staffs "reduced the ability of supervisors to detect problems early enough to take corrective actions."[5] As a result, state and federal banking supervisors increased their examination staffs and conducted bank examinations more frequently, on average, as problems in the banking industry increased through the latter part of the 1980s. As these problems intensified, the issue of the frequency and scope of bank examinations increasingly became part of the public policy debate. This process culminated in the passage of FDICIA, which mandates full-scope, on-site examinations each year for U.S. commercial banks.[6]

In general, bank examinations are scheduled at least several months in advance, both to give banks time to prepare and to allow supervisors to develop an overall schedule and individual examination plans that make efficient use of available resources. Given this advanced scheduling, changing the timing of one examination typically also entails rescheduling others to free up the needed resources. Thus, several factors work to reinforce the timing implicit in the original examination schedule. However, even after the passage of FDICIA, supervisors continue to have some discretion over the timing of examinations for individual banks. To some extent, the size and perceived condition of a bank can influence the planned time between full-scope examinations, but there is now less scope for supervisors to lengthen this time period without violating FDICIA's annual examination requirement. Supervisors can and do accelerate full-scope examinations and undertake targeted examinations if other factors indicate that problems are developing at a particular bank. In fact, supervisors employ fairly extensive off-site monitoring procedures—including the use of statistical models—to help identify those banks where problems might be emerging.[7]

The CAMEL Rating System

A key outcome of an examination is a supervisory rating of the bank's overall financial condition. For the purposes of this article, we focus on these supervisory ratings—known as CAMEL ratings—as a proxy for the information resulting from a bank examination. CAMEL ratings, which are assigned by examiners at the conclusion of an examination, are numerical ratings of the quality of a bank's financial condition, risk profile, and overall performance. This rating system is used by the three federal banking supervisors—the OCC, the FDIC, and the Federal Reserve—as well as by the Office of Thrift Supervision, the National Credit Union Administration, and state banking supervisors to provide a convenient summary rating of each bank's condition at the time of the examination. In addition, CAMEL ratings are increasingly being used for other supervisory purposes, such as setting deposit insurance rates and expediting bank applications for various regulatory purposes.

CAMEL ratings, which are assigned by examiners at the conclusion of an examination, are numerical ratings of the quality of a bank's financial condition, risk profile, and overall performance.

The acronym CAMEL refers to the five components of a bank's condition assessed by examiners: *C*apital adequacy, *A*sset quality, *M*anagement, *E*arnings, and *L*iquidity.[8] Ratings are assigned for each component, and a composite rating is assigned for the overall condition and performance of the bank. These component and composite ratings are assigned on a scale of 1 to 5, with 1 representing the highest rating (strongest performance) and 5 representing the lowest (weakest performance). Banks with composite CAMEL ratings of 1 or 2 are generally considered to present few, if any, significant supervisory concerns, while those with ratings of 3, 4, or 5 are considered to present moderate to extreme degrees of supervisory concern.[9]

The CAMEL rating is only one aspect of the information produced during a bank examination. Examiners also put together a detailed report that describes the bank's activities and management structure; assesses the bank's performance along the dimensions reflected in the CAMEL rating as well as in other areas; and, where appropriate, contains recommendations for changes and improvements in the bank's policies and procedures. This report is backed by extensive examination work papers detailing the process leading to the examiners' conclusions. The CAMEL rating, while not a comprehensive indicator of all this information, nonetheless provides a convenient summary measure of the examination findings.

A bank's CAMEL rating is known only by the bank's senior management and appropriate supervisory staff at the relevant supervisory agencies. CAMEL ratings are never made publicly available, even on a lagged basis.

All of this examination material, including the CAMEL rating, is highly confidential. A bank's CAMEL rating is known only by the bank's senior management and appropriate supervisory staff at the relevant supervisory agencies. CAMEL ratings are never made publicly available, even on a lagged basis. Thus, to a considerable degree, the CAMEL rating reflects the private supervisory information gathered during a bank examination as well as whatever public and regulatory information is available about the bank's condition. For this reason, we use the ratings as our indicator of the private supervisory information arising from bank examinations.

Previous Research on the Information in CAMEL Ratings

Other researchers have examined the role of CAMEL ratings in providing information about the financial condition of banks. For instance, Berger and Davies (1994) examine the information content of CAMEL ratings by testing for stock price reactions when new ratings are assigned. Despite the fact that CAMEL ratings are confidential, the authors find that rating downgrades seem to lead to negative excess stock returns. They interpret this result as evidence that examinations generate valuable private information and that rating downgrades reveal unfavorable private information about bank conditions. Similarly, DeYoung, Flannery, Lang, and Sorescu (1998) find that CAMEL ratings contain information useful to the market for subordinated, bank holding company debt.

Berger, Davies, and Flannery (1998) find that BOPEC ratings—the supervisory ratings given to bank holding companies—contain information about bank conditions that goes beyond the information in market data, such as bond-rating downgrades.[10] In particular, they find that supervisory data and market information Granger-cause (or are useful in forecasting) one another, suggesting that both supervisors and the financial markets have some unique information. Finally, Barker and Holdsworth (1993) find evidence that CAMEL ratings are significant predictors of bank failure, even after controlling for a wide range of publicly available information about the condition and performance of banks. Taken together, these papers suggest that supervisory ratings contain information about the condition and performance of banks that is not available to the public.

These papers suggest that newly assigned CAMEL ratings contain relevant information. Some researchers have also studied how that relevance changes over time. For example, Gilbert (1993) addresses the issue to some extent by finding that more frequent examinations reduced losses to the Bank Insurance Fund, which covers government losses when a bank fails. Cole and Gunther (1995, 1998) find that the information contained in CAMEL ratings decays quickly with respect to predicting bank failure from 1986 to 1992. In particular, they find that a model using publicly available financial data is a better indicator of the likelihood of bank failure than the previous CAMEL rating is once the rating is more than one or two quarters old. These two studies address the issue of information decay directly; however, the primary purpose of CAMEL ratings is not to identify future bank failures, but to provide an assessment of banks' overall conditions at the time of the examinations.

Focusing on this aspect of supervisory ratings, Berger, Davies, and Flannery (1998) examine BOPEC ratings in relation to market-based data and find that only very recent examinations provide useful information. The information appears to become much less useful, or "stale," over time. In our analysis, we focus directly on the time decay of the supervisory information in CAMEL ratings and the decay's impact on assessing the current condition of a bank. Thus, we view our article as complementary to, and an extension of, this general line of research.

Structure of the Data Set

The basic data set used in our analysis consists of the composite CAMEL ratings assigned to banks from 1989 to 1995 by the Federal Reserve, the FDIC, the OCC, and state banking supervisors. Each CAMEL rating was given following a full-scope, on-site examination. We eliminated from our sample any ratings associated with targeted or limited-scope examinations. CAMEL ratings are not always assigned during such examinations and, if they are, may not reflect the most up-to-date information about the full scope of a bank's activities.

For each CAMEL rating in the sample, we know the as-of date of the examination (the date as of which the condition of the bank is assessed), the supervisory entity that conducted the examination, and the identity of the bank. We matched each observation to the corresponding bank's income and balance sheet data, as reported in the *Report of Condition and Income* (the *Call Report*) maintained by the bank supervisory agencies, for the quarter prior to the as-of date of the examination. These *Call Report* data will serve as our proxy for the information available from regulatory reports and other public sources about the bank's condition at the time of the examination.

To assess how quickly the private supervisory information from a bank examination decays, we linked each observation to the CAMEL rating from the previous full-scope examination for that bank. That is, for each CAMEL rating in the sample, we know the lagged composite CAMEL rating as well as the date of the previous full-scope examination. With this information, we can calculate the time since the last examination, a key variable in our analysis.[11]

An overview of this element of the data set appears in Table 1. The table presents the number of full-scope bank examinations in our sample from 1989 to 1995, sorted by the time since the last examination. The number of examinations per year varies significantly. From about 7,000 examinations in 1989, the sample size drops to just under 4,000 in 1991 before rising again to over 8,000 starting in 1992. This variation is due to several factors. During 1990 and 1991, the number of full-scope examinations performed by the FDIC dropped significantly, while the number of limited-scope examinations rose. Given that our sample is based on full-scope examinations, this shift resulted in a sharp decline in the size of the data set.

Table 1

Number of Full-Scope Examinations in a Year, Sorted by Quarters since Last Full-Scope Examination

Quarters since Last Examination	1989	1990	1991	1992	1993	1994	1995
1	102	57	20	102	87	79	54
2	911	427	198	625	722	494	321
3	1,347	820	324	805	1,131	952	718
4	1,581	1,057	580	1,037	2,780	3,273	2,704
5	1,191	795	557	601	1,523	1,784	1,932
6	717	463	598	564	623	561	1,034
7	380	251	389	324	342	369	417
8	298	184	324	389	270	397	317
9	188	99	209	371	126	269	244
10	97	41	154	369	114	132	97
11	76	34	129	354	82	57	51
12	58	36	121	364	72	32	23
13-14	38	28	212	750	161	53	18
15-16	8	7	83	640	208	52	8
17 or more	5	7	82	1,029	757	333	74
Total	6,997	4,306	3,980	8,324	8,998	8,837	8,012

Source: Authors' calculations, based on data from the Board of Governors of the Federal Reserve System.

However, following the passage of FDICIA in 1991 with its requirement for annual full-scope examinations, the number of examinations in the data set rose significantly.

Looking down the columns for each year, we see that about three-quarters of the examinations took place within six quarters of the prior examination. There is a clear pattern of clustering of lagged examinations at three to five quarters, particularly in the latter part of the sample period. This clustering is consistent with the supervisory goal of ensuring that each bank has an annual full-scope examination. Finally, there is significant variation across the years in the share of the sample for which the time since the last examination was more than twelve quarters. The early years of the sample contain relatively few such observations, but their numbers increase sharply during 1992 and 1993 before declining significantly in later years. This sharp increase most likely reflects the impact of FDICIA, as the various supervisory agencies made efforts to examine more banks in response to the requirement for annual full-scope examinations.[12]

Empirical Approach: The Off-Site and Examination Models of CAMEL Ratings

To explore the question of how quickly private supervisory information generated during an examination decays, we develop two empirical models to predict banks' composite CAMEL ratings. The first is based on the FIMS model used by the Federal Reserve for off-site monitoring purposes (see Cole, Cornyn, and Gunther [1995] for details). Like the FIMS model, ours uses income and balance sheet data from banks' *Call Reports* to predict composite CAMEL ratings.[13] The specific variables included in the model are listed in Box A. These variables are intended to capture the five CAMEL rating components as well as other influences—such as regional factors and the time since the last full-scope exam—that could help determine the CAMEL rating. Because the variables used in the model do not incorporate the information gathered by supervisors through on-site exams, we call this model the "off-site model."

We estimated this model for each year in the sample period.[14] The overall fit is quite good with the R^2 goodness-of-fit statistics ranging from 0.50 to 0.70.[15] Although the specific variables that enter the model with statistically significant coefficients differ from year to year, a core set of variables have consistent signs and are significant in nearly every year. These variables include the log of total assets, the equity-to-capital ratio, the current and lagged ratios of net income to total assets, the ratio of residential mortgages to total loans, and the ratio of consumer loans to total loans. The coefficients on these variables suggest that, all else equal, larger banks, banks with higher capital and net income ratios, and banks with higher proportions of comparatively less risky residential mortgages and consumer loans tend to receive

Box A: Explanatory Variables Used in the Empirical Models

Capital Adequacy
- equity-to-capital ratio
- four-quarter change in equity-to-capital ratio

Asset Quality
- log of total assets
- four-quarter change in log of total assets
- loan-to-asset ratio
- commercial and industrial loans as share of total loans
- one-to-four-family mortgages as share of total loans
- real estate loans as share of total loans
- consumer loans as share of total loans
- loans past due thirty to eighty-nine days as share of total assets
- loans past due ninety or more days as share of total assets
- nonperforming loans as share of loan loss reserves
- loan loss reserves as share of total loans
- net charge-offs in year before examination as share of total assets
- year-over-year change in net charge-offs as share of total assets
- provisions in year before examination as share of total assets
- year-over-year change in provisions as share of total assets

Management
- interest rate risk exposure (assets minus liabilities that mature or reprice in more than five years)
- insider loans as share of total assets

Earnings
- ratio of net income to total assets in year before examination
- net-income-to-assets ratio lagged one year

Liquidity
- cash as share of total assets

Other Variables
- dummy variables for quarter in which examination took place (Q1, Q2, Q3, Q4)
- dummy variables for bank's Federal Reserve District
- dummy variables for agency conducting examination (Federal Reserve Bank, Federal Deposit Insurance Corporation, Office of the Comptroller of the Currency, or state regulator)
- dummy variables for number of quarters since previous examination

better CAMEL ratings. In contrast, banks with higher loan-to-asset ratios, higher amounts of past due and non-accrual loans, higher ratios of nonperforming loans to loan loss reserves, and higher interest rate risk exposures consistently receive worse CAMEL ratings.

In addition to estimating the off-site model, we estimated a second model that includes the previous composite CAMEL rating for each bank. Because this model includes the private supervisory information contained in these lagged CAMEL ratings, we call it the "examination model." The model already contains variables that control for information from updated regulatory reports, so any additional explanatory power from the lagged CAMEL rating is assumed to reflect private supervisory information.[16] By comparing the ability of the two models to explain current CAMEL ratings as the age of the lagged CAMEL rating increases, we can assess how long this supervisory information provides additional useful information on the current condition of the bank.

To conduct this comparison, we allow the coefficients on the lagged CAMEL rating to differ according to the length of time since the previous examination. In particular, we divide the observations in each year of the sample into fifteen distinct categories according to the time since the previous examination. We then let the lagged CAMEL rating enter the model with a different coefficient for each category.[17] In this way, we can test how the explanatory power of lagged CAMEL ratings varies as the ratings age.[18] This approach provides a *within-sample diagnostic*, meaning that we can test the explanatory power of the lagged CAMEL ratings on the same sample of examinations used to estimate the model.

Before presenting our empirical results, it is worth discussing the role of the fifteen dummy variables reflecting the time since the previous examination. They are included to capture the effects of any independent factors that might cause a relationship between the value of the current CAMEL rating and the time since the last examination. In that way, we can be assured that the coefficients on the interacted, lagged CAMEL ratings are capturing just the influence of the private information from the previous examination rather than these other factors. In fact, the hypothesis that these time-related factors are not meaningful (that is, the coefficients on the dummy variables are jointly equal to zero) is strongly rejected for each year in the sample, indicating that there is some independent influence of the time since the last examination on the value of the current CAMEL rating.[19]

The results of the within-sample diagnostic are presented in Table 2, which contains the coefficient estimates on the interacted, lagged CAMEL ratings in the examination model for each year in the sample. The end

Table 2
COEFFICIENTS ON INTERACTED, LAGGED CAMEL RATINGS IN THE EXAMINATION MODEL

Quarters since Last Examination	1989	1990	1991	1992	1993	1994	1995
1	2.158*	2.663*	0.803	1.893*	2.680*	3.209*	3.237*
	(0.236)	(0.437)	(0.581)	(0.259)	(0.306)	(0.367)	(0.399)
2	2.309*	2.343*	1.631*	1.347*	2.556*	2.607*	3.058*
	(0.093)	(0.133)	(0.179)	(0.099)	(0.105)	(0.134)	(0.169)
3	2.221*	2.500*	1.811*	1.728*	2.531*	2.785*	2.736*
	(0.078)	(0.109)	(0.161)	(0.094)	(0.092)	(0.105)	(0.131)
4	2.241*	2.345*	1.931*	1.728*	2.370*	2.572*	2.842*
	(0.078)	(0.105)	(0.132)	(0.091)	(0.072)	(0.077)	(0.091)
5	2.306*	2.580*	2.203*	1.624*	2.274*	2.578*	2.569*
	(0.089)	(0.130)	(0.143)	(0.117)	(0.086)	(0.093)	(0.096)
6	2.209*	2.767*	2.164*	1.596*	2.525*	2.650*	2.747*
	(0.117)	(0.158)	(0.150)	(0.126)	(0.152)	(0.167)	(0.151)
7	2.163*	2.113*	1.772*	1.324*	2.154*	2.292*	2.207*
	(0.161)	(0.200)	(0.148)	(0.142)	(0.219)	(0.212)	(0.224)
8	1.557*	2.0448	1.872*	1.524*	2.386*	2.221*	2.513*
	(0.198)	(0.237)	(0.160)	(0.140)	(0.285)	(0.223)	(0.303)
9	1.661*	2.138*	1.429*	1.245*	1.988*	2.497*	2.914*
	(0.251)	(0.344)	(0.212)	(0.128)	(0.373)	(0.277)	(0.399)
10	1.786*	1.332*	1.892*	1.300*	2.498*	1.958*	1.615*
	(0.340)	(0.587)	(0.222)	(0.134)	(0.417)	(0.378)	(0.477)
11	1.579*	1.623*	1.202*	1.187*	1.731*	1.628*	1.992*
	(0.373)	(0.497)	(0.261)	(0.137)	(0.405)	(0.485)	(0.734)
12	1.924*	1.805	1.752*	1.205*	2.157*	2.556*	1.229
	(0.606)	(1.265)	(0.296)	(0.123)	(0.568)	(1.087)	(0.696)
13-14	0.253	1.014	1.736*	1.158*	1.990*	0.988	0.779
	(0.527)	(0.700)	(0.258)	(0.100)	(0.280)	(0.638)	(0.553)
15-16	0.135	-0.553	1.486*	1.025*	1.816*	2.035*	0.697
	(0.828)	(2.874)	(0.333)	(0.120)	(0.204)	(0.523)	(0.661)
17 or more	2.075*	-0.072	1.250*	0.742*	0.597*	0.945*	0.669
	(0.772)	(1.284)	(0.467)	(0.107)	(0.131)	(0.189)	(0.414)
Memo: R^2	0.824	0.811	0.768	0.741	0.737	0.694	0.692

Source: Authors' calculations.

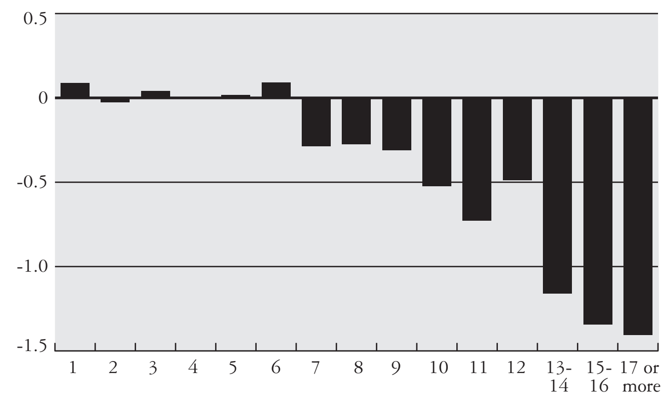
Notes: The coefficients are for the independent variables produced by interacting the lagged CAMEL ratings with dummy variables reflecting the amount of time since the last examination. Standard errors are presented in parentheses. An asterisk indicates that the coefficient is significantly different from zero at the 5 percent level.

of the shading indicates the point at which the lagged CAMEL rating generally no longer enters the model with statistical significance and thus ceases to provide useful information in modeling current CAMEL ratings. Clearly, this result varies across the sample. For 1989, 1990, and 1995, the lagged CAMEL rating is not significant beyond eleven to twelve quarters.[20] However, for the other years, CAMEL ratings older than three years provide some information regarding the current condition of the bank.

Although these results indicate that relatively old CAMEL ratings have explanatory power, further analysis shows that the value of the private supervisory information contained in the ratings decays as it ages. This evidence arises from the size of the coefficients on the lagged CAMEL ratings in addition to their statistical significance. Overall, the hypothesis that the coefficients on the lagged CAMEL ratings are stable across the age categories is strongly rejected. Furthermore, the size of the coefficients declines as the age of the lagged CAMEL rating increases, even while remaining significant. In the context of our model, smaller coefficients imply that changes in the value of the lagged CAMEL ratings have less of an impact on the value of current CAMEL ratings, even though they continue to provide some explanatory power. As shown in the chart, although the decline in the size of the coefficients is not monotonic, there is a general pattern consistent with the idea that the relationship between lagged and current CAMEL ratings decays as the age of the lagged CAMEL rating increases.[21] As the chart illustrates, there is a sharp drop-off in the size of the coefficients once the lagged CAMEL rating is more than six quarters old, suggesting that lagged CAMEL ratings have their greatest impact before they reach this age.

AVERAGE COEFFICIENTS ON LAGGED CAMEL RATINGS, SORTED BY TIME SINCE LAST EXAMINATION, 1989-95

Relative to Coefficient for Four-Quarter-Old Lagged CAMEL Ratings



Source: Authors' calculations.

OUT-OF-SAMPLE ANALYSIS

The results discussed thus far all represent a within-sample analysis of the information content of lagged CAMEL ratings, where the significance tests are carried out on the same set of data used to estimate the models. To enhance our understanding of how the value of private supervisory information changes over time, we also conduct several *out-of-sample tests*; that is, tests of the predictive power of the lagged CAMEL ratings using data other than those used to estimate the models. Out-of-sample tests are of interest for two related reasons.

First, the tests provide a more robust assessment of a model's ability to explain current CAMEL ratings. By using data outside of the estimation sample, we can assess whether the estimated model is stable over time and across different sets of observations. In our analysis, this distinction amounts to asking whether the decay rate of private supervisory information indicated by the examination model reflects the particular observations in a given year or whether the relationship is more general.

Second, out-of-sample tests more closely mirror the situation facing bank supervisors. Supervisors have information about recent bank examinations and therefore can analyze the relationship between lagged and current CAMEL ratings for those banks. Based partly on this analysis, supervisors need to infer how quickly the private supervisory information from other banks is likely to deteriorate and therefore how quickly these banks need to be examined. This situation is essentially an out-of-sample forecasting problem.

To conduct this out-of-sample analysis, we estimate our two models using data from one year and then use the estimated coefficients to forecast the CAMEL ratings to be assigned during the following year. For example, we estimate the off-site and examination models using the 1989 sample and use the coefficient estimates to forecast the CAMEL ratings for the examinations in the 1990 sample. This procedure gives us two separate forecasts of CAMEL ratings for 1990, one based on each model.

To evaluate the quality of these CAMEL rating forecasts, we need statistical tools that differ from those used in the within-sample tests of the statistical significance of the regression coefficients. The forecasts from the off-site and examination models are actually probability forecasts that a bank will receive a CAMEL rating of either 1, 2, 3, 4, or 5. For example, such a forecast might be that the bank has a 30 percent chance of being rated 2; a 50 percent chance of being rated 3; a 20 percent chance of being rated 4; and a 0 percent chance of being rated 1 or 5. We use a standard measure of forecast accuracy, known as the logarithmic scoring rule (LSR), to evaluate such multistate probability forecasts. The LSR measure examines how much weight a model's forecast places on the outcome that actually occurred. That is, if the CAMEL rating for a particular examination was 2, the LSR would assess the quality of the forecast by looking only at the probability assigned to that outcome. Under the mathematical assumptions used in computing the LSR measure (Box B), smaller LSR values imply more accurate forecasts.[22]

Box B: Model Comparisons Using the Logarithmic Scoring Rule

The off-site and examination models used in this article are ordered logit models, which provide probability forecasts for each of the five possible CAMEL ratings. In mathematical form, such an out-of-sample forecast, denoted P_n, is a (5x1) vector in which the i^{th} element represents the forecasted probability of being in state i. For example, the out-of-sample forecast might be P_n = [0; 0.30; 0.50; 0.20; 0]. Accuracy measures for such forecasts relate the performance of the forecasts to actually observed outcomes. Let R_n be an indicator vector such that if the CAMEL rating is i (where i = 1,...,5), then the i^{th} element equals one and zero otherwise. For example, if bank n receives a CAMEL rating of 4, then R_n = [0; 0; 0; 1; 0].

The accuracy measure used here, known as the logarithmic scoring rule (LSR), examines how much weight the probability forecast places on the actual outcome. That is, if the CAMEL rating for a particular examination were 2, the LSR would assess the accuracy of the forecast only by looking at the probability it assigned to that outcome. The mathematical formula for the LSR is

$$LSR = -\frac{1}{N}\sum_{n=1}^{N} \log\left(\sum_{i=1}^{5} P_{in}{*}R_{in}\right),$$

where N is the number of banks for which forecasts are generated. Since R_{in} equals one only for the CAMEL rating actually observed, the LSR is simply the average of the negative, logged value of the probability forecast for the rating actually observed. LSR can take on values in the interval $[0,\infty]$ with smaller values implying greater accuracy.

The LSR measure permits model comparison by creating performance rankings. For example, if the LSR value for the probability forecasts from model A (denoted $P_{A,n}$) is smaller than that for the forecasts from model B (denoted $P_{B,n}$), then model A is said to be more accurate than model B. However, an important question is whether this observed difference in LSR values is statistically significant or just an artifact of the data sample. To examine this point, Diebold and Mariano (1995) propose several tests for determining whether the difference is statistically different from zero.

Generally, the null hypothesis under scoring rule g is $E[g(P_{An}, R_n)] = E[g(P_{Bn}, R_n)]$, or equivalently, $E[d_n] = E[g(P_{An}, R_n) - g(P_{Bn}, R_n)] = 0$. For LSR, d_n for a single observation is

$$d_n = -\log\left(\sum_{i=1}^{5} P_{Ani}{*}R_{ni}\right) + \log\left(\sum_{i=1}^{5} P_{Bni}{*}R_{ni}\right).$$

To examine this null hypothesis, we simply calculate the difference between the scores from our two models for each observation in the sample and regress it against a constant. If this coefficient is statistically different from zero, then the aggregate scores for the sample as a whole differ significantly; that is, the observed performance ranking is statistically significant.

This measure of forecast accuracy permits model comparison by creating performance rankings. For example, if the LSR value for the probability forecasts from model A is smaller than that for the forecasts from model B, then model A can be said to be more accurate than model B. However, an important question is whether this observed difference in LSR values is statistically significant or just an artifact of the data sample. To examine this point, Diebold and Mariano (1995) propose several tests for determining whether the difference is statistically different from zero. We use a simple variant of one of their suggestions (Box B). Specifically, we calculate the difference between the scores from our two models for each observation in the sample. We then regress this difference against a constant term and test to see whether the constant is statistically different from zero. This procedure is equivalent to testing whether the aggregate scores for the sample as a whole differ significantly.

The basic results of the out-of-sample analysis are presented in Table 3, which contains the comparison of LSR values for each year of the data set.[23] In the early years of the sample, the LSR values for the examination model are significantly smaller than those for the off-site model for examinations up to six to seven quarters old; that is, the difference between the two values is positive and significant. After 1991, this cutoff point increases to ten to twelve quarters. In other words, the results suggest that the private supervisory information contained in CAMEL ratings continues to provide useful information in predicting the current condition of a bank for six to twelve quarters after the previous examination. After this point, there appears to be little value in the information contained in the prior CAMEL rating.

Overall, the examination model generates more accurate forecasts than the off-site model up to a certain point. An alternative way to express this result is to examine the models' integer forecasts of the CAMEL ratings; that is, the expected CAMEL rating, rounded to the nearest integer, based on the models' probability forecasts. These forecasted CAMEL ratings can then be compared with the observed CAMEL ratings. For the 1990 data, the off-site model correctly predicted about 67 percent of the realized CAMEL ratings for banks that had lagged ratings up to six quarters old. The examination model improved this performance by correctly predicting roughly 75 percent of the realized ratings. However, for banks with older lagged CAMEL ratings, both models perform equally, with about 40 percent accuracy. For all the years in our sample, the off-site and examination models exhibit this difference in forecast performance before the cutoff point, but not after. Again, this result indicates that the private supervisory information in lagged CAMEL ratings from full-scope examinations decays over time and is not useful in predicting the current CAMEL ratings after a certain point.[24]

In fact, some of the results suggest that after a certain point, using lagged CAMEL ratings to predict current ones may actually be detrimental to producing accurate forecasts. In some instances, the score for the *off-site* model is significantly smaller than for the *examination* model, indicating that the former produces more accurate forecasts than the latter. For example, in Table 3, for observations in

Table 3
DIFFERENCES IN LOGARITHMIC SCORING RULE VALUES, SORTED BY TIME SINCE LAST EXAMINATION FOR THE ONE-YEAR-AHEAD FORECASTS OF CAMEL RATINGS
Full Sample

	(LSR1-LSR2)					
Quarters since Last Examination	1989/ 1990	1990/ 1991	1991/ 1992	1992/ 1993	1993/ 1994	1994/ 1995
1	0.167*	0.173	0.840	0.189*	0.297*	0.429*
2	0.223*	0.250*	0.119*	0.137*	0.184*	0.307*
3	0.218*	0.215*	0.138*	0.137*	0.202*	0.165*
4	0.160*	0.145*	0.162*	0.103*	0.146*	0.179*
5	0.152*	0.137*	0.101*	0.095*	0.153*	0.151*
6	0.179*	0.130*	0.090*	0.145*	0.278*	0.146*
7	0.029	0.069*	0.076*	0.077*	0.121*	0.129*
8	-0.045	0.046	0.025	0.082*	0.096*	0.097*
9-10	-0.104	-0.061	0.071*	0.077*	0.076*	0.109*
11-12	-0.177	-0.095*	0.037	0.083*	0.098*	0.053
13 or more	-0.212*	-0.157*	-0.030*	-0.027*	-0.073*	-0.073

Source: Authors' calculations.

Notes: Each figure gives the difference in values for banks examined in that year (column) whose lagged CAMEL ratings were of the corresponding age (row). The figures represent the difference between the LSR value for the off-site model (LSR1) and the LSR value for the examination model (LSR2). A positive (negative) value indicates that the examination (off-site) model produces a more accurate forecast than the off-site (examination) model. An asterisk indicates that the difference is significantly different from zero at the 5 percent level using the specified Diebold-Mariano test. The end of the shading indicates the point at which the difference between the LSR values is no longer statistically positive.

which the lagged CAMEL rating is thirteen or more quarters old, the off-site model has a significantly smaller LSR in all but one year in the sample, and thus it provides more accurate predictions of current CAMEL ratings in these years. These results imply that these aged CAMEL ratings add no value in assessing a bank's current condition.

Finally, the results suggest that there is significant variation over the sample period in the useful life of supervisory information from prior examinations. This variation may reflect changes in the condition of the U.S. banking industry over the sample period. In particular, the private supervisory information contained in CAMEL ratings appears to decay more rapidly during the early part of the sample period, when the U.S. banking industry was experiencing financial stress, than during the latter part of the sample period, when the industry experienced more robust performance. Because we would expect the condition of banks to change more rapidly during periods of financial stress, we would also expect a faster rate of information decay.

Our findings suggest that the rate of decay in private supervisory information is considerably faster for banks experiencing some degree of financial difficulty, regardless of the overall condition of the banking industry.

To explore our results further, we divided the data into subsets according to the initial financial condition of the bank. Specifically, for each year, we divided the data sample into observations with lagged CAMEL ratings of 1 or 2 (indicating little reason for supervisory concern at the time of the previous examination) and observations with lagged CAMEL ratings of 3, 4, or 5 (indicating moderate to severe degrees of supervisory concern). We then compared the LSR measures for our CAMEL forecasting models for each of these subsamples. These results are reported in Tables 4 and 5.

As Table 4 indicates, the results for the subsample with lagged CAMEL ratings of 1 or 2 are very similar to those for the overall sample. The results indicate that the lagged CAMEL ratings cease to provide useful information about the current condition of a bank after six to twelve quarters have elapsed and that this information decays faster in the early part of the sample, when the U.S. banking industry was experiencing financial distress. The similarity between these results and the overall results for the sample is not surprising, since the majority of observations (between 70 and 90 percent) have lagged CAMEL ratings of 1 or 2.

As indicated in Table 5, the findings for observations with lagged CAMEL ratings of 3, 4, or 5 are considerably different.[25] The point at which the lagged CAMEL rating ceases to provide useful information regarding the current CAMEL ratings is significantly earlier than it is for the overall sample.[26] The information in prior CAMEL ratings seems to be no longer useful after just three to six quarters. Further, the cyclical pattern that

Table 4
DIFFERENCES IN LOGARITHMIC SCORING RULE VALUES, SORTED BY TIME SINCE LAST EXAMINATION FOR THE ONE-YEAR-AHEAD FORECASTS OF CAMEL RATINGS
Subsample with Lagged CAMEL Ratings of 1 or 2

Quarters since Last Examination	(LSR1-LSR2) 1989/1990	1990/1991	1991/1992	1992/1993	1993/1994	1994/1995
1	0.150*	0.252*	0.082*	0.210*	0.158*	0.331*
2	0.147*	0.114*	0.165*	0.131*	0.164*	0.231*
3	0.145*	0.169*	0.112*	0.175*	0.137*	0.150*
4	0.119*	0.108*	0.168*	0.139*	0.136*	0.164*
5	0.133*	0.111*	0.134*	0.118*	0.125*	0.146*
6	0.126*	0.103*	0.111*	0.138*	0.283*	0.141*
7	0.043	0.048	0.073*	0.091*	0.160*	0.127*
8	-0.045	0.028	0.078*	0.096*	0.105*	0.091*
9-10	-0.058	-0.051	0.036*	0.085*	0.082*	0.109*
11-12	-0.207	-0.042	0.008	0.078*	0.111*	0.086
13 or more	-0.182	-0.109*	0.001	-0.013	-0.049*	-0.056

Source: Authors' calculations.

Notes: Each figure gives the difference in values for banks examined in that year (column) whose lagged CAMEL ratings were of the corresponding age (row). The figures represent the difference between the LSR value for the off-site model (LSR1) and the LSR value for the examination model (LSR2). A positive (negative) value indicates that the examination (off-site) model produces a more accurate forecast than the off-site (examination) model. An asterisk indicates that the difference is significantly different from zero at the 5 percent level using the specified Diebold-Mariano test. The end of the shading indicates the point at which the difference between the LSR values is no longer statistically positive.

was evident in both the overall sample and in the subsample with lagged CAMEL ratings of 1 or 2 does not emerge in these results. Taken together, these findings suggest that the rate of decay in private supervisory information is considerably faster for banks experiencing some degree of financial difficulty, regardless of the overall condition of the banking industry.

What do these results imply for the basic question motivating this article, namely, how often should a bank be examined? To answer this question, it is important to understand that the tests described above provide an upper-bound for the length of time that prior CAMEL ratings provide useful information about current bank conditions. That is, beyond the six-to-twelve-quarter range is where the lagged CAMEL rating contains *no* useful information about the current condition of a bank. In practice, supervisors would probably wish to examine a bank before this point, when the private information from the prior examination continues to have *some*, though diminished, value.

Table 5

DIFFERENCES IN LOGARITHMIC SCORING RULE VALUES, SORTED BY TIME SINCE LAST EXAMINATION FOR THE ONE-YEAR-AHEAD FORECASTS OF CAMEL RATINGS

Subsample with Lagged CAMEL Ratings of 3 to 5

	(LSR1-LSR2)					
Quarters since Last Examination	1989/ 1990	1990/ 1991	1991/ 1992	1992/ 1993	1993/ 1994	1994/ 1995
1	0.220	-0.143	0.089	0.142	0.725*	0.771*
2	0.345*	0.377*	0.071*	0.143*	0.215*	0.433*
3	0.365*	0.303*	0.173*	0.071*	0.374*	0.221*
4	0.294*	0.278*	0.142*	0.002	0.208*	0.305*
5	0.240*	0.275*	-0.005	0.020	0.321*	0.200*
6	0.568*	0.309*	0.000	0.112	0.193	0.359
7	-0.091	0.141	0.096	-0.072	-0.684	0.186
8	-0.049	0.136	-0.139	-0.060	-0.062	0.502
9-10	-0.383	-0.102	0.152*	-0.024	-0.042	0.133
11-12	0.223	-0.340	0.100	0.154	-0.106	NA
13 or more	-0.980	-0.450	-0.160*	-0.115*	-0.163	-0.391

Source: Authors' calculations.

Notes: Each figure gives the difference in values for banks examined in that year (column) whose lagged CAMEL ratings were of the corresponding age (row). The figures represent the difference between the LSR value for the off-site model (LSR1) and the LSR value for the examination model (LSR2). A positive (negative) value indicates that the examination (off-site) model produces a more accurate forecast than the off-site (examination) model. An asterisk indicates that the difference is significantly different from zero at the 5 percent level using the specified Diebold-Mariano test. The end of the shading indicates the point at which the difference between the LSR values is no longer statistically positive.

Finally, in thinking about the optimal time between examinations, the results suggest that this horizon may vary. During periods of financial stress in the banking industry, the quality of private supervisory information appears to decay faster than it does in more stable periods, suggesting that the optimal time between examinations may be shorter in times of stress. Further, the rate of information decay is markedly greater for banks that are themselves financially troubled, regardless of the state of the overall industry. This finding implies, rather sensibly, that it is desirable to examine troubled institutions more often than healthy ones, although the optimal examination interval for any particular bank will vary from the averages discussed here.[27]

ROBUSTNESS CHECKS

To examine the robustness of our results to the choices we made in setting up the analysis, we conducted two additional sets of tests. We examined the performance of our two models on out-of-sample observations, both to test the robustness of the results and to mirror more closely the actual situation faced by bank examiners. The approach we chose—year-ahead forecasts—is only one way of setting up such an out-of-sample test. As discussed in Granger and Huang (1997), out-of-sample analysis for models of this type can also be conducted by holding out a random part of the sample for a given year and using that holdout sample for the out-of-sample analysis.[28] We use this approach to test whether the results discussed above are due solely to the year-ahead forecast analysis.

Table 6 contains the results of our holdout sample prediction analysis.[29] For each year in the sample, we estimated the two models over a randomly selected 75 percent of the total sample. These estimated models for each year were then used to predict the CAMEL ratings on the remaining 25 percent of the sample. We again compared the accuracy of these predictions using the LSR measure.

The holdout sample prediction results are broadly similar to those for the year-ahead forecast analysis. The examination model exhibits better performance than the off-site model for observations with CAMEL ratings that are six to twelve quarters old; that is, the differences in

LSR values between the examination and off-site models are positive and significant for this portion of the holdout sample. The previously observed cyclical pattern is less evident, but the results nonetheless provide some indication that the information in lagged CAMEL ratings decays less rapidly in the latter years of the sample. The weaker cyclical pattern may be due to the considerably smaller number of out-of-sample observations available using this type of analysis. The smaller sample size reduces the power of the statistical tests to determine whether the accuracy measures for the two models differ significantly. Overall, however, the holdout sample results support the findings of the year-ahead forecasts, suggesting that our analysis is not overly sensitive to the structure of the out-of-sample analysis.

For the second set of robustness tests, we focus directly on the question whether the time between full-scope examinations can be treated as an exogenous variable in our two models. We have assumed that the models capture the relevant explanatory variables used by examiners in determining CAMEL ratings. However, it might be the case that in scheduling examinations, supervisors have additional information—not present in our empirical specifications—about the extent to which conditions at a bank have changed since the last examination. Using such information, supervisors might schedule more frequent examinations for banks whose financial conditions are less stable and less frequent examinations for those with more stable conditions. In that case, the time since the last examination would be an endogenous variable, rather than an exogenous one as we have assumed. That is, the time since the last examination may be a function of the current CAMEL rating that we are trying to model.

FDICIA's requirement for annual full-scope examinations seems reasonable, particularly for banks whose initial financial condition is troubled or when the banking system as a whole is experiencing financial stress.

Some preliminary evidence in favor of such endogeneity is presented in Table 7, which shows the cumulative distribution of the time since the last full-scope examination in percentage terms. The first column of the table presents the cumulative distribution for all observations aggregated across the seven years in the sample, while the other columns report the results for subsamples divided by current CAMEL ratings. Clearly, the time between examinations for banks with CAMEL ratings of 3 to 5 is shorter than it is for banks with ratings of 1 or 2. About 45 percent of banks with ratings of 1 or 2 had a lagged full-scope examination within four quarters, compared with almost 60 percent for banks with ratings of 3 to 5. Although this difference diminishes as the time between examinations increases (by eight quarters, the percentages are nearly equal), it may be the case that the time since the last examination is a function of the current CAMEL rating. The existence of such endogeneity might lead our

Table 6
DIFFERENCES IN LOGARITHMIC SCORING RULE VALUES FOR THE HOLDOUT SAMPLE, SORTED BY TIME SINCE LAST EXAMINATION
Full Sample

	(LSR1-LSR2)						
Quarters since Last Examination	1989	1990	1991	1992	1993	1994	1995
1	0.177	-0.040	0.226	0.175*	0.241*	0.229	0.238
2	0.224*	0.266*	0.228*	0.179*	0.280*	0.279*	0.150
3	0.237*	0.195*	0.123	0.177*	0.178*	0.242*	0.077
4	0.136*	0.168*	0.100*	0.142*	0.150*	0.113*	0.128*
5	0.104*	0.184*	0.134*	0.215*	0.137*	0.222*	0.141*
6	0.154*	0.237*	0.121*	0.072*	0.145*	0.142*	0.144*
7	0.057	0.121*	0.099*	0.059	0.093	0.106*	0.106*
8	0.070	0.135	0.005	0.017	0.049	0.069	0.108
9-10	-0.007	-0.175	0.053	0.071*	-0.008	0.058	0.028
11-12	0.077	-0.060	-0.038	0.017	0.045	0.144*	-0.085
13 or more	-0.057	-0.610	0.035	-0.008	-0.042	-0.083	-0.069

Source: Authors' calculations.

Notes: Each figure gives the difference in values for banks examined in that year (column) whose lagged CAMEL ratings were of the corresponding age (row). The figures represent the difference between the LSR value for the off-site model (LSR1) and the LSR value for the examination model (LSR2). A positive (negative) value indicates that the examination (off-site) model produces a more accurate forecast than the off-site (examination) model. An asterisk indicates that the difference is significantly different from zero at the 5 percent level using the specified Diebold-Mariano test. The end of the shading indicates the point at which the difference between the LSR values is no longer statistically positive.

empirical tests to overstate the length of time that a lagged CAMEL rating continues to provide useful information about the current condition of a bank.

To test for this possible endogeneity, we use two distinct methods. First, we use a logistic regression relating the probability that the CAMEL rating changes (either upward or downward) to the time between examinations. If the time between examinations were strictly endogenous, we would expect to find no significant relationship between these two variables: supervisors would schedule examinations at the point when conditions at the bank had changed sufficiently to warrant a change in the CAMEL rating. In contrast, if the time between examinations were exogenous, we would expect to see a positive relationship between the time since the last examination and the probability of a change in the CAMEL rating.

The results of this regression are reported in Table 8. Clearly, the coefficient on the time since the last examination is positive and significant for each year of the sample. Although the overall fit of the regressions is poor (the R^2 statistics are quite low), these results support the idea that the time since the last examination is not significantly endogenous.

Table 7

CUMULATIVE DISTRIBUTION OF TIME SINCE LAST FULL-SCOPE EXAMINATION

Complete Sample and Divided by Current CAMEL Rating

Quarters since Last Examination	Complete Sample (Percent)	Current CAMEL Rating of 1 or 2 (Percent)	Current CAMEL Rating of 3 to 5 (Percent)
1	1.0	0.9	1.4
2	8.5	6.7	16.2
3	20.8	17.5	35.5
4	47.1	44.7	58.0
5	64.1	62.3	71.8
6	73.3	72.5	76.8
7	78.3	77.8	80.3
8	82.7	82.6	83.3
9	85.8	85.8	85.8
10	87.8	87.8	87.7
11	89.4	89.4	89.3
12	90.8	90.8	90.9
13-14	93.3	93.1	94.2
15-16	95.4	95.1	96.7
17 or more	100.0	100.0	100.0
Memo: Number of observations	49,455	40,252	9,203

Source: Authors' calculations, based on data from the Board of Governors of the Federal Reserve System.

To explore this question further, we conducted a second test that explicitly attempts to control for the endogeneity of the time between examinations. We begin this test by modeling the time since the last examination as a function of variables that are correlated with it, but not with the current CAMEL ratings. The fitted values from this model should therefore be free of this possible endogeneity. By substituting these fitted values for the dummy variables for the actual time since the last examination in our two earlier models, we expect that the generated CAMEL rating forecasts will not be affected by any endogeneity between the time since the last examination and the current CAMEL rating. If the LSR results based on these modified models are found to be similar to those for the versions that do not control for potential endogeneity, then this finding would provide additional evidence that such endogeneity is not influencing our results.[30]

In technical terms, we model the time between examinations using an econometric technique known as hazard modeling.[31] The explanatory variables used in estimating the hazard models were the changes in the core balance sheet and income statement variables that form the basis of the off-site and examination models. Although the levels of these variables are significant determinants of current CAMEL ratings, it seems reasonable to assume that their lagged values, and thus the changes in their values, are exogenous. After the hazard models have been estimated, they can be used to generate predicted probabilities

Table 8

LOGISTIC REGRESSION RESULTS: PROBABILITY OF CAMEL RATING CHANGE AS A FUNCTION OF TIME SINCE LAST EXAMINATION

	1989	1990	1991	1992	1993	1994	1995
Constant	-1.096 *	-1.522 *	-1.456 *	-0.996 *	-1.032 *	-1.283 *	-1.470 *
	(0.057)	(0.078)	(0.073)	(0.042)	(0.035)	(0.039)	(0.052)
Time since last examination (months)	0.016 *	0.033 *	0.031 *	0.017 *	0.011 *	0.013 *	0.018 *
	(0.004)	(0.005)	(0.003)	(0.001)	(0.001)	(0.002)	(0.003)
R^2	0.002	0.010	0.026	0.013	0.008	0.005	0.004
Number of observations	6,998	4,306	3,980	8,324	8,998	8,837	8,012

Source: Authors' calculations.

Notes: The dependent variable equals 1 when the CAMEL rating changes (increases or decreases) and is zero otherwise. R^2 statistics are those derived for limited dependent variable models in Estrella (1998). An asterisk indicates that the coefficient is significantly different from zero at the 1 percent level.

Table 9
DIFFERENCES IN LOGARITHMIC SCORING RULE VALUES BASED ON FITTED HAZARD FUNCTIONS, SORTED BY TIME SINCE LAST EXAMINATION FOR THE ONE-YEAR-AHEAD FORECASTS OF CAMEL RATINGS
Full Sample

	(LSR1-LSR2)					
Quarters since Last Examination	1989/ 1990	1990/ 1991	1991/ 1992	1992/ 1993	1993/ 1994	1994/ 1995
1	0.159*	0.143	0.199*	0.141*	0.279*	0.391*
2	0.236*	0.242*	0.168*	0.059*	0.170*	0.295*
3	0.226*	0.199*	0.172*	0.073*	0.195*	0.179*
4	0.165*	0.146*	0.163*	0.114*	0.150*	0.187*
5	0.155*	0.118*	0.136*	0.100*	0.152*	0.155*
6	0.179*	0.125*	0.138*	0.134*	0.283*	0.189*
7	0.021	0.064*	0.100*	0.075*	0.144*	0.149*
8	-0.052	0.013	0.036	0.085*	0.095*	0.125*
9-10	-0.124*	-0.011	0.065*	0.075*	0.110*	0.136*
11-12	-0.185	-0.116*	0.031	0.096*	0.104	0.141*
13 or more	-0.199	-0.070	-0.017	0.009	-0.088*	-0.024

Source: Authors' calculations.

Notes: Each figure gives the difference in values for banks examined in that year (column) whose lagged CAMEL ratings were of the corresponding age (row). The figures represent the difference between the LSR value for the off-site model (LSR1) and the LSR value for the examination model (LSR2). A positive (negative) value indicates that the examination (off-site) model produces a more accurate forecast than the off-site (examination) model. An asterisk indicates that the difference is significantly different from zero at the 5 percent level using the specified Diebold-Mariano test. The end of the shading indicates the point at which the difference between the LSR values is no longer statistically positive.

that the time between examinations falls into specified ranges. We substituted these predicted probabilities for the dummy variables representing the actual time since the last examination.[32]

Table 9 presents the LSR comparison results for the off-site and examination models using the estimated survivor function for the examinations in each year. The results are quite similar to those reported in Table 3. In the early years of the sample, the LSR value for the examination model is less than the value for the off-site model for examinations up to six or seven quarters old. After 1991, this cutoff point increases to roughly nine to twelve quarters after the examination. The results for the two subsamples of CAMEL ratings (not reported in the tables) are similar to those in Tables 4 and 5. Thus, the out-of-sample forecast results do not appear to be sensitive to our attempts to control for the potential endogeneity of the time since the last examination. Based on these results, as well as on the logit results reported above, it does not appear that our conclusions are being driven by an endogenous relationship between the current CAMEL rating and the time since the previous examination.

CONCLUSION

This article examines the frequency with which supervisors should examine banks by assessing the decay rate of the private supervisory information gathered during full-scope examinations. Such information is costly to obtain since it can be gathered only during on-site examinations. Thus, the question of how quickly the information's value erodes has important implications for both supervisors and banks. The more quickly this information decays, the more frequently examinations need to take place in order for supervisors to have access to accurate information about the current condition of banks.

Our results suggest that CAMEL ratings cease to provide any useful information about the current condition of a bank after about six to twelve quarters. Thus, examinations should take place at least at this frequency, since supervisors would probably want to examine a bank while the information from the previous examination continues to have some value. Our results indicate that supervisory information tends to decay more rapidly for banks with weaker CAMEL ratings (3, 4, or 5). Thus, for these institutions, a somewhat shorter examination cycle may be justified. In this light, FDICIA's requirement for annual full-scope examinations seems reasonable, particularly for banks whose initial financial condition is troubled or when the banking system as a whole is experiencing financial stress. Of course, the optimal examination frequency for any particular bank can and will deviate from the average results presented here.

ENDNOTES

The authors thank Bob DeYoung, Fred Herriman, Larry Radecki, Marc Saidenberg, Phil Strahan, Joe Tracy, two anonymous referees, and participants at the 1998 meeting of the Federal Reserve System Committee on Financial Structure and Regulation and the Federal Reserve Bank of New York Banking Studies seminar for helpful comments and suggestions. Leslie DuPuy and Oba McMillan provided excellent research assistance in preparing the data set for this article.

1. An important qualification to this statement is that the verification of the accuracy of regulatory reports is one aspect of on-site examinations.

2. Note that our results reflect the average pattern of information decay across the examinations in the sample; the optimal examination timing for individual banks will differ from these averages.

3. Profits in the banking industry fell sharply through the mid-to-late 1980s, reflecting large loan losses in several lending sectors, including agriculture, energy, developing countries, and real estate. Profits, as measured by return on equity, did not return to pre-downturn levels until 1992. Failures also rose sharply during this period, reaching a high of more than 250 per year in the late 1980s (see Federal Deposit Insurance Corporation [1997]).

4. According to estimates by the Federal Deposit Insurance Corporation (FDIC), the field examination staffs of the three federal bank supervisory agencies—the FDIC, the Federal Reserve, and the Office of the Comptroller of the Currency (OCC)—and the fifty state banking supervisors totaled about 9,000 in 1994. For more information, see Federal Deposit Insurance Corporation (1997).

5. See Federal Deposit Insurance Corporation (1997, p. 428).

6. The exception is very small banks with supervisory ratings that indicate few, if any, significant supervisory concerns; these banks can be examined once every eighteen months.

7. For example, the Federal Reserve uses the Financial Institutions Monitoring System (FIMS) for this purpose (see Cole, Cornyn, and Gunther [1995] for details).

8. The formal name of the rating system is the Uniform Financial Institutions Rating System, although it is commonly known as the CAMEL rating system. In 1997, a sixth component was added, reflecting a bank's *S*ensitivity to market risk. The expanded rating system is known as the CAMELS rating system. Because our data sample extends only through 1995, none of the examinations in our sample includes this new component.

9. See Federal Financial Institutions Examination Council (1996) for a detailed description of the CAMEL rating system and an interpretation of the component and composite ratings.

10. Bank holding companies are examined separately from their bank subsidiaries. The BOPEC rating assigned at the conclusion of such an examination reflects the conditions of the holding company's *B*ank subsidiaries, *O*ther nonbank subsidiaries, *P*arent company, *E*arnings, and *C*apital adequacy.

11. We focus on the time between full-scope bank examinations, so banks in our sample may have had either a targeted or limited-scope examination between full-scope examinations. In such cases, supervisors will have had the opportunity to update some of their private information about the bank's condition. As discussed above, however, such examinations generally do not result in comprehensive assessments of a bank's condition. Therefore, by examining the time interval between full-scope examinations, we likely obtain the best indication of the time decay of the private supervisory information.

12. Note that there are a significant number of observations with prior examinations more than six quarters old, even in 1994 and 1995, well after the passage of FDICIA, which set an outside limit of eighteen months between examinations. About 75 percent of the observations have intervening, limited-scope examinations that occurred within six quarters of the current examination, suggesting the efforts made by supervisors to make a full transition to FDICIA's requirements. Furthermore, the relatively small number of observations during 1989 and 1990 for which the time between examinations is fairly long may partly reflect the source data used in constructing the data set. Because the source data contained increasingly sparse information on examinations before 1989, our data set for 1989 and 1990 excludes examinations of banks whose previous examinations were not recorded in the source data.

13. Technically, the statistical approach used is an ordered logit model. CAMEL ratings have discrete values, so a standard linear regression model—which assumes that the dependent variable is continuous—would be inappropriate. The ordered logit model is specifically designed to handle discrete dependent variables, such as CAMEL ratings, whose values are ordinally related (that is, 1 implies "strongest performance," while 5 implies "weakest performance"). See Maddala (1983) for a detailed discussion of ordered logit models.

14. We conduct our analysis on annual, cross-sectional data sets, as opposed to a panel data set, for two reasons. First, a simple likelihood ratio test rejects the null hypothesis that the model coefficients are constant across the years. Second, because examiners must allocate their

Note 14 continued
scarce resources over future examinations, we felt that focusing on year-ahead CAMEL rating forecasts generated from annual data sets would more closely mirror examiner behavior.

15. The R^2 statistic is the goodness-of-fit measure developed by Estrella (1998) specifically for limited dependent variable models. The statistic is roughly analogous to the R^2 statistic used in linear regressions because its value ranges between zero (for a model with no explanatory power) and one (for a model with complete explanatory power).

16. The information contained in the lagged CAMEL rating reflects both private supervisory information and past values of the public information on bank condition. To isolate the effects of the supervisory information, we also estimated a version of the examination model that controlled for the publicly available information component. In particular, we estimated an ordered logit model that regressed lagged CAMEL ratings on lagged values of the publicly available independent variables listed in Box A. From this model, we calculated a fitted value of the lagged CAMEL rating using the predicted probabilities that the rating was equal to a 1, 2, 3, 4, or 5. We then subtracted this fitted value from the actual lagged CAMEL rating. We interpret this residual as reflecting the information in the lagged CAMEL rating stemming just from the private supervisory information. We then substituted this residual for the actual lagged CAMEL rating in the ordered logit equation for the current CAMEL rating. The results of the subsequent out-of-sample forecast analysis were nearly identical to those for the examination model using the actual lagged CAMEL rating, suggesting that this variable primarily reflects private supervisory information.

17. Using mathematical notation, we can summarize the off-site model for a given year as $y_i = f\left(\gamma x_i + \sum_{j=2}^{15} \beta_j I(\text{lag})_{ij} + \varepsilon_i\right)$, where y_i is the current CAMEL rating for bank i; γ is the vector of coefficients on the independent variables x_i listed in Box A (except for the indicator variables for the time since last examination); the $I(lag)_{ij}$'s are the indicator variables corresponding to the time since the last examination for bank i; the β_j's are the corresponding coefficients; and ε_i is the error term. The examination model for a given year is

$$y_i = f\left(\gamma x_i + \sum_{j=2}^{15} \beta_j I(lag)_{ij} + \sum_{j=1}^{15} \theta_j {*} I(lag)_{ij} {*}\, lag\text{CAMEL}_i + \varepsilon_i\right),$$

where $lag\text{CAMEL}_i$ is the value of the lagged CAMEL rating for bank i from the previous examination and the θ_j's are the corresponding coefficients. The difference between the models is simply the inclusion of interacted, lagged CAMEL variables. The coefficients on the x_i variables, particularly on the core set of variables, do not significantly change when the interacted, lagged CAMEL variables are included in the specification.

18. We also estimated a constrained version of the examination model in which the coefficient on the lagged CAMEL rating does not vary with the time since the last examination. Based on a likelihood ratio test, these constraints are clearly rejected for every year in the sample. This finding indicates that there is meaningful variation in the coefficients on the lagged CAMEL ratings as the age of the rating increases. However, as the out-of-sample forecast results (discussed in the next section) were not significantly affected by these constraints, our primary results are not overly sensitive to the way in which the lagged CAMEL ratings enter the examination model.

19. Using this model specification, we assume that the time since the last full-scope examination is an exogenous variable; that is, it does not depend on the current CAMEL rating. A plausible argument can be made that the variable is endogenously determined, especially with respect to lower rated banks. Although we cannot conclusively prove that the variable is exogenous, we provide indirect evidence later on. We maintain the assumption throughout the analysis that follows.

20. The coefficient on lagged CAMEL ratings that are seventeen or more quarters old in the 1989 regression is an exception, since it is statistically significant.

21. To try to formalize this observation, we test the hypothesis that the coefficients on lagged CAMEL ratings that were twelve or more quarters old were smaller than the coefficients on lagged CAMEL ratings that were four quarters old. (We selected four quarters as being representative of relatively new CAMEL ratings, but the results are not sensitive to this choice.) For all cases, the coefficients on the older CAMEL ratings were smaller than those on the four-quarter-old CAMEL ratings, and these differences were statistically significant about half the time. In particular, in all but two of the sample years, at least half the coefficients on the older CAMEL ratings were significantly smaller. These results lend support to the more informal observation that the size of the coefficients tends to decrease as the age of the CAMEL ratings increases.

22. Estrella and Mishkin (1996) recommend using the logarithmic scoring rule to evaluate probability forecasts derived from models estimated using maximum likelihood estimation techniques (such as the ordered logit models used here). The reason for this is that the LSR mimics the weights used in maximizing the likelihood function to obtain the parameter estimates.

23. To maximize the power of the Diebold-Mariano test used in the analysis, we use a higher level of aggregation for the lagged CAMEL ratings. Specifically, we group into three categories all observations for which the prior examination occurred nine or ten quarters ago, eleven or twelve quarters ago, and thirteen or more quarters ago, rather than into the seven categories used in the ordered logit estimation.

24. We also estimated versions of our model in which we attempted to assess the impact of limited-scope examinations on these results. Overall, about 20 percent of the observations in the sample have limited-scope examinations between the full-scope examinations. The distribution of these observations is uneven across years and concentrated in 1991 to 1994, the period during which the supervisory agencies were in the midst of the transition to FDICIA. For this analysis, we substituted the time since the last limited-scope examination and its associated CAMEL rating for the time since the last full-scope examination and its CAMEL rating. The empirical results differ somewhat from the results presented in Table 3. The adjusted results suggest that the information contained in lagged CAMEL ratings decays within six to eight quarters, and the strong cyclical pattern in Table 3 is not evident. The difference in results may be attributable to the fact that the adjustment for limited-scope examinations reduces the number of observations with "old" lagged CAMEL ratings to the point where the statistical tests on this part of the sample have greatly diminished power. Alternatively, the results could reflect the fact that limited-scope examinations are not as in-depth as full-scope examinations and may not produce information of as high a quality. The difference in our results could reflect the fact that this lower quality information simply decays faster than the information derived from full-scope examinations. This interpretation is supported by the results presented in Table 3, which suggest that the information from lagged full-scope CAMEL ratings persists *even when there has been an intermediate limited-scope examination.* Based on this interpretation, we do not view the limited-scope results as undercutting our findings about the persistence of information from full-scope examinations.

25. The subsample of banks with lagged CAMEL ratings of 3, 4, or 5 makes up between 10 and 30 percent of the yearly samples. This smaller sample size reduces the power of the Diebold-Mariano tests upon which our results are based, especially for the reduced number of banks with older CAMEL ratings. However, the sample size for the figures just beyond the cut-over points (that is, the figures after which our inference is most relevant) remains large enough to permit valid inference.

26. Note that, for the 1992 results in Table 5, the LSR values for the examination model are lower than those for the off-site model up to four quarters since the last examination. There is no clear intuition why this is also the case at nine to ten quarters. This result is probably due to random noise in the data set.

27. In fact, the conclusion that financially troubled institutions should be examined more frequently is directly incorporated into FDICIA. For example, the previously noted exception for small banks is disallowed for banks with CAMEL ratings of 3, 4, or 5. In fact, the eighteen-month examination window is disallowed, regardless of CAMEL rating, for a number of reasons related to bank soundness and important changes in bank control. These additional criteria help ensure that potentially troubled banks are examined more frequently than healthy ones. Moreover, the FDICIA-specified intervals between examinations are meant to be outer limits; bank supervisors have the discretion to conduct more frequent examinations. The figures in Table 1 suggest that this discretion is often used.

28. Granger and Huang (1997) distinguish between forecasting, which involves estimating a model on a set of observations and then applying these estimates to observations from a future period, and prediction, which involves estimating a model on a subset of the observations from a given period and then applying these estimates to other observations from the same time period. In the discussion in the text, we use these terms in a manner consistent with these definitions.

29. Because we are no longer using the year-ahead forecast analysis, we have out-of-sample results for 1989.

30. Note that the use of these fitted values for the time since the last examination can be viewed, at least in spirit, as an instrumental variable estimation technique.

31. See Kiefer (1988) for a survey of hazard modeling. For our particular hazard model, we specify the baseline hazard function as a Weibull function, which allows the hazard rate (the probability that an examination occurs in a given quarter after the last examination) to increase or decrease as the time since the last examination increases. In our estimates, we found evidence that the hazard rate was increasing, suggesting that banks were more likely to be examined as the time since the last examination increased.

32. The hazard model results were used to create variables representing the probability of an examination occurring one, two, three, four, and five or more years after the lagged examination. Note that this is a higher level of aggregation than the one used in the results presented in Tables 3-5, where fifteen dummy variables were used. The reduction in the number of time variables was performed to facilitate the estimation of the ordered logit models used in the analysis. The results are not sensitive to this reduction.

REFERENCES

Barker, D., and D. Holdsworth. 1993. "The Causes of Bank Failures in the 1980s." Federal Reserve Bank of New York Research Paper no. 9325.

Berger, A. N., and S. M. Davies. 1994. "The Information Content of Bank Examinations." In PROCEEDINGS OF THE 1994 CONFERENCE ON BANK STRUCTURE AND COMPETITION, 615-28. Federal Reserve Bank of Chicago.

Berger, A. N., S. M. Davies, and M. J. Flannery. 1998. "Comparing Market and Supervisory Assessment of Bank Performance: Who Knows What When?" Board of Governors of the Federal Reserve System Finance and Economics Discussion Series Paper no. 1998-32.

Cole, R. A., B. G. Cornyn, and J. W. Gunther. 1995. "FIMS: A New Monitoring System for Banking Institutions." FEDERAL RESERVE BULLETIN, January: 1-15.

Cole, R. A., and J. W. Gunther. 1995. "A CAMEL Rating's Shelf Life." Federal Reserve Bank of Dallas FINANCIAL INDUSTRY STUDIES, December: 13-20.

———. 1998. "Predicting Bank Failures: A Comparison of On- and Off-Site Monitoring Systems." JOURNAL OF FINANCIAL SERVICES RESEARCH 13: 103-17.

DeYoung, R., M. J. Flannery, W. Lang, and S. Sorescu. 1998. "The Informational Advantage of Specialized Monitors: The Case of Bank Examiners." Federal Reserve Bank of Chicago Research Department Working Paper no. 98-4.

Diebold, F. X., and R. Mariano. 1995. "Comparing Predictive Accuracy." JOURNAL OF BUSINESS AND ECONOMIC STATISTICS 13: 253-64.

Estrella, A. 1998. "A New Measure of Fit for Equations with Dichotomous Dependent Variables." JOURNAL OF BUSINESS AND ECONOMIC STATISTICS 16: 198-205.

Estrella, A., and F. S. Mishkin. 1996. "The Evaluation of Out-of-Sample Probability Forecasts Based on Maximum Likelihood Estimates." Unpublished paper, Federal Reserve Bank of New York.

Federal Deposit Insurance Corporation. 1997. "An Examination of the Banking Crises of the 1980s and Early 1990s." In HISTORY OF THE EIGHTIES: LESSONS FOR THE FUTURE. Vol. 1, 1-86 and 428. Washington, D.C.: Federal Deposit Insurance Corporation.

Federal Financial Institutions Examination Council. 1996. "Uniform Financial Institutions Rating System." FEDERAL REGISTER 61, no. 245: 67021-9.

Gilbert, R. A. 1993. "Implications of Annual Examinations for the Bank Insurance Fund." Federal Reserve Bank of St. Louis REVIEW, January-February: 35-52.

Granger, C. W. J., and L.-L. Huang. 1997. "Evaluation of Panel Data Models: Some Suggestions from Time Series." University of California, San Diego, Department of Economics Discussion Paper no. 97-10.

Kiefer, N. M. 1988. "Economic Duration Data and Hazard Functions." JOURNAL OF ECONOMIC LITERATURE 26: 646-79.

Maddala, G. S. 1983. LIMITED-DEPENDENT AND QUALITATIVE VARIABLES IN ECONOMETRICS. Cambridge: Cambridge University Press.

Macro Markets and Financial Security

Stefano Athanasoulis, Robert Shiller, and Eric van Wincoop

Today, people have a rich set of investment options, ranging from low-risk money market instruments to high-risk growth stocks. They can choose to invest in mutual funds, hedge funds, and pension plans. They can hedge themselves with options and other derivatives while investing both at home and across the globe. Plenty of opportunities are available for diversifying their portfolios and avoiding excess exposure to sectoral or geographic risk. Nonetheless, there is good reason to believe that most people's wealth is not well diversified. For example, although investors can diversify through equity markets, corporate profits account for less than 10 percent of national income. That figure suggests that about 90 percent of an average person's income is sensitive to sectoral, occupational, and geographic uncertainty.

Shiller (1993) has proposed a new set of markets that could in theory provide much better diversification opportunities. These so-called macro markets would be large international markets trading, in the form of futures contracts, long-term claims on major components of incomes shared by a large number of people or organizations. For example, in a macro market for the United States, an investor could buy a claim on the U.S. national income and then receive, for as long as the claim is held, dividends equal to a specified fraction of U.S. national income. Such a claim is comparable to a share in a corporation, except that the dividend would equal a share of national income rather than a share of corporate profits. Such markets might exist for entire countries—the United States, Japan, and Brazil—or for regions—such as the European Union and North America. Even a market for claims on the combined incomes of the entire world could be formed. Prices would rise and fall in these markets as new information about national, regional, or global economies became available, just as

Stefano Athanasoulis is a visiting assistant professor at Yale University. Robert Shiller is the Stanley B. Resor Professor of Economics at Yale University. Eric van Wincoop is a senior economist at the Federal Reserve Bank of New York.

prices rise and fall in the stock market as new information about corporate profits is revealed.

The potential future importance of these markets is supported by the most basic principle of finance—diversification. People could use macro markets to hedge their own national income risks and to invest in the rest of the world. This investment strategy would reduce income growth uncertainty and lead to a more secure financial future.

We address several questions in this paper. First, how could macro markets be useful to the average person? Second, how large are the potential benefits from diversification if these markets were to be introduced and used optimally? Third, can existing financial markets achieve a similar degree of diversification when used optimally? Fourth, why don't these markets already exist?

How Would Investors Use Macro Markets?

The basic idea behind macro markets is a simple one. Consider the case of claims on national income. If macro markets existed for every country of the world, people could take short positions in their country's market, thereby hedging their own country's risk, and long positions in the markets of all other countries in proportion to each country's size, thereby completely hedging themselves. The short positions in their home country would exactly offset the long positions that they hold by virtue of living there, and the long positions in the world would mean that they were completely diversified. If everyone hedged risk in this way, it would all add up, that is, for every long in every country there would be a short, and demand would equal supply in each macro market. The dividends paid on the securities for each country would be paid by the people who live in that country and hold short positions. By definition, these people can always make the payments because they are earning the national income upon which the dividends are drawn.

Taking such positions in these markets is the best way for an individual to achieve diversification. After hedging, everyone earns a share of global income. It would be impossible for individuals to lower their risks any further. It is impossible for everyone to diversify away uncertainty about global income, because total income earned across all individuals equals global income itself.

Retail Institutions

Of course, most people are not accustomed to hedging. Thus, it would probably be unrealistic to expect the average person to hedge through macro markets without the assistance of intermediaries. Most people *are* familiar with insurance, and they readily buy insurance against other risks. Retail institutions, such as pension funds or insurance companies, could offer people contracts to hedge their aggregate income risk. These insurance companies and pension funds would trade in macro markets to sell off the risk incurred by writing the contracts in retail markets. These institutional investors would be hedging, much as institutions now hedge in stock index futures markets.

An Average Investor

We will now give an example of how these markets and retail institutions could serve the individual investor. Consider a person who earns income from wages and from returns on financial assets (such as stocks and bonds). The individual cares about the uncertainty of the future value of his or her *total wealth*, which is the sum of the future value of financial assets and the future value of "human capital." The value of human capital is equal to the present value of the stream of future wages earned by the individual. The value of the person's wealth can thus be written as

$$Wealth = PDV(\Pi) + PDV(W),$$

where PDV is present discounted value, Π represents the annual dividends and interest earned from financial assets, and W is wages plus noncorporate business income. Even if the individual were well diversified in the equity and bond markets, he or she would still be exposed to uncertainty associated with wages earned. Because wages plus noncorporate profits are at least nine times as great as corporate profits (in national income accounts), the largest component of wealth remains undiversified.

Let us further assume that the wealth of the individual is "average"—the value of the individual's financial assets is average and his or her wages are equal to the average wage rate in the country plus an idiosyncratic component. The idiosyncratic component of wages depends on individual-specific effort as well as a dose of good or bad luck. Insuring against the idiosyncratic component is impossible because of moral hazard problems. If an individual were insured against all uncertainty about future wages, he or she would have little incentive to work hard and to put effort into a successful career. Given these assumptions, the value of wages is written as $W = W_C + W_I$, where W_C is the average wage rate in the country, and W_I is the idiosyncratic component. The sum of the idiosyncratic component over all individuals is zero. Moral hazard problems do not apply to insuring oneself against uncertainty about W_C because the individual has little control over the average wage rate earned in the country as a whole.

The individual clearly gains by hedging in macro markets to the extent that less uncertainty surrounds the growth rate of world output than the growth rate of the home country's output.

We also assume that the individual invests only in domestic stocks and bonds and that he or she is well diversified domestically. The absence of international diversification is not far from current practice: Japanese and U.S. investors hold at least 90 percent of their equity portfolio in domestic assets.[1] Because the individual's financial assets are average, the dividends earned on these assets, Π, are equal to the per capita value of total corporate profits in the country. We can then write the individual's wealth as

$$Wealth = PDV(GDP) + PDV(W_I),$$

where GDP is per capita gross domestic product, which equals $\Pi + W_C$. Wealth is therefore equal to the present discounted value of future per capita GDP plus the present discounted value of the idiosyncratic component of wages. Macro markets can be used to insure the uncertainty associated with per capita GDP.

As a matter of simplification, assume that the expected future per capita GDP of the country in which the individual resides is equal to that for the world as a whole (GDP_W) and that the "riskiness" of the country's future GDP is average. We will be more precise about what that means in a moment. Insurance companies and pension funds can allow people to hedge uncertainty about the country's per capita GDP by offering a hedging instrument with a yearly payoff of $GDP_W - GDP$. As we explain below, the price of this hedging instrument is zero. Although the expected payoff is zero, the actual payoff can be both positive or negative. If it is negative, the individual must make a payment. If the hedging instrument is offered by a pension fund, the payment could be made through a debit on the individual's account at the pension fund. This contract is attractive to a risk-averse individual because he or she will lose on the hedging contract only when the domestic economy is doing unexpectedly well. The individual will receive positive payments from the contract when the economy's performance is unexpectedly poor. If the individual opts to use this instrument, his or her net wealth will be

$$Wealth = PDV(GDP_W) + PDV(W_I).$$

The individual clearly gains by hedging in macro markets to the extent that less uncertainty surrounds the growth rate of world output than the growth rate of the home country's output.

Notice that in our example the individual invests only in domestic financial assets, then hedges uncertainty about both domestic financial returns and domestic wages through the hedging instrument. This investment strategy is attractive because it avoids the need to make decisions about investment in foreign financial assets. The problem of asymmetric information means that domestic investors are at a disadvantage relative to foreign investors when evaluating foreign stocks and bonds. Foreign investors tend to be better informed about companies trading in their own stock markets, particularly in the case of smaller companies. They can therefore adjust their portfolio more rapidly than domestic investors as new information becomes available to them. Gehrig (1993) shows that investors are reluctant to invest abroad if foreign investors receive a more precise price signal about foreign stock returns than domestic investors. Asymmetric information is one of the most common explanations for the lack of observed international diversification in equity and bond markets. In macro markets, which are tied to aggregate incomes, asymmetric information is much less of a concern. Japanese investors are not likely to predict Japanese GDP growth rates more accurately than U.S. investors because the information needed to make such predictions is publicly available.

The diversification strategy outlined above is different from the type of diversification most investors are accustomed to. Most individual stock market investors diversify by investing their money in a wide basket of assets. With macro markets, diversification is achieved instead through a hedging contract.

Pricing in Macro Markets

So far we have left two issues unaddressed. First, the institutional investors that offer the hedging contract we just described will themselves be exposed to risk when offering the instrument. Second, we have yet to explain why the price of the contract will be zero. To understand how institutional investors will lay off the risk and what factors determine prices, we describe in more detail the macro markets on which the hedging instruments are based. These markets trade perpetual claims on a GDP index. Trade can take place either over the counter or on an exchange like the Chicago Board of Trade.

Existing theoretical research has laid out exactly what will determine prices in markets like these.[2] As with any asset, the price of a claim on a country's per capita GDP depends on two factors—expected payoff and risk. The expected payoff is the expected present discounted value of future per capita GDP. Risk is measured by the covariance between the present discounted value of a country's per capita GDP and the present discounted value of the world's GDP.

First consider a simple example in a symmetric world. Two countries have an equal number of residents. Assume that expected future per capita GDP is the same in both markets. If we also assume that the variance of the present discounted value of GDP is the same for both countries, then the covariance with the world claim will be identical for the two countries. Claims on the per capita GDP of both countries therefore will have the same price.

Let us say for the sake of simplicity that the only traders in these markets are pension funds, and let N be the size of the population in both markets. Domestic pension funds will sell $\frac{1}{2}N$ perpetual claims on domestic per capita GDP and buy $\frac{1}{2}N$ perpetual claims on foreign per capita GDP. Because these claims have the same price, the net cost will be zero. Foreign pension funds take the other side of the market. The per capita gross domestic product of the world, GDP_W, equals $\frac{1}{2}GDP + \frac{1}{2}GDP^*$, where GDP^* is foreign per capita GDP. Through their operations in the macro markets, domestic pension funds have effectively purchased N perpetual claims on $GDP_W - GDP$. Because the pension funds also sell N perpetual claims on $GDP_W - GDP$ to domestic individuals through the hedging instrument, domestic pension funds break even. The same is true for the foreign pension funds. The two countries have effectively agreed to swap a claim on half of each other's GDP. Under this arrangement, there is no cost or "insurance premium" to reducing risks. After risk sharing, the residents of both countries will hold claims on half the domestic country's per capita GDP plus half the foreign country's per capita GDP, which together add up to world per capita GDP. Residents' expected average income is the same as it was before, but the variability of income is lower.

Because people's exposure to national income risk differs, limiting trade in claims on a country's national income to the residents of that particular country would be beneficial.

So far everything in the example is very symmetric. Now suppose that the domestic country is much larger than the foreign country: its population N is a hundred times that of the foreign country. Accordingly, the covariance between domestic GDP and world GDP will be higher than the covariance between foreign GDP and world GDP, even if the variance of per capita GDP in both countries is the same. The price of a perpetual claim on the foreign country's per capita GDP will therefore be lower than the price of a claim on the domestic country's per capita GDP.

If the prices of claims on the per capita GDP of both countries were still equal—as they were when both countries had the same population—then people in the larger country would want to swap half their income for half the per capita income of the people in the smaller country. But there are not enough people in the smaller country to take the other side of these transactions. Therefore, the price of a perpetual claim on the foreign country's per capita GDP will be higher than the price of a claim on the domestic country's per capita GDP. Consequently, the people in the larger country will be discouraged from demanding so many claims on the foreign country, and market clearing can take place.

In more technical terms, a claim on domestic per capita GDP can be exchanged for α claims on world per capita GDP, with $\alpha < 1$. Through trade in macro markets, domestic pension funds will buy N claims on $\alpha(GDP_W) - GDP$ (with a net price of zero) and sell those claims as hedging instruments to domestic individuals. After the hedge, domestic residents have a perpetual claim on α times per capita world GDP. Foreign pension funds will take the other side of the market by selling N claims on $\alpha(GDP_W) - GDP$, which is equivalent to buying $N/100$ (the foreign population) claims on $\beta GDP_W - (GDP)^*$. Here $\beta = 101 - 100\alpha > 1$. Foreign pension funds will sell these claims as hedging instruments to foreign individuals, who will then own a perpetual claim on β times per capita world GDP. The higher price of a claim on the foreign country's output leads to larger claims on world per capita GDP by foreign residents after risk sharing.

In the example above, we have assumed for simplicity that all individuals within a country have the same exposure to their country's national income risk. In reality, some individual's income is more sensitive to national growth rates than other people's income. The optimal hedge position that an investor takes through pension funds or insurance companies depends on his or her exposure to national risk. Because people's exposure to national income risk differs, limiting trade in claims on a country's national income to the residents of that particular country would be beneficial. Although this limitation would eliminate international risk sharing, it would allow individuals to share their exposures to national income risk. Ultimately, through the appropriate retail institutions, those individuals with high exposure to national income risk could sell perpetual claims indexed to national income to those individuals with low exposure to national income risk.

The Potential Risk-Sharing Benefits

Individuals are exposed to many types of aggregate risk. The most common risks are specific to a sector (occupational risk), to an age cohort (demographic risk), or to a geographic area in which someone works (geographic risk). For example, an auto worker is subject to auto industry risk. A decline in demand for automobiles will affect the entire industry. Geographic risk can be linked to a specific neighborhood or to a whole continent. To measure the potential diversification benefits of macro markets, we restrict our analysis to national income risk, abstracting from other types of aggregate risk. Because we limit ourselves to national risk, the measure of hedgeable aggregate income risk derived in this section is lower than the level achievable through aggregate income markets generally.

Because individuals cannot diversify away global income growth uncertainty, we focus on country-specific growth, that is, the difference between a country's growth rate and the world growth rate. As explained in the previous section, macro markets allow individuals to eliminate the country-specific component of their income growth uncertainty. We now quantify the size of this uncertainty.[3]

A Regression Model of Country-Specific Growth Uncertainty

To identify country-specific growth uncertainty, we estimate the following regression for each horizon s:

$$g_{i,t,t+s} - g^{w}_{t,t+s} = \lambda_s'\left(z_{it} - z^{w}_{t}\right) + u_{i,t,t+s}.$$

The left-hand side of the equation represents country *i*'s growth in real per capita GDP from year t to $t+s$ minus global growth in real per capita GDP over the same period. The first term on the right-hand side of the equation is the predictable component of the deviation of country growth from world growth. This component depends on the relevant information set available to the market, which is captured by the vector z_{it}, in deviation from its global counterpart. The term u is the unpredictable component of the country-specific deviation from world growth. We also refer to country-specific growth uncertainty as residual risk.

We apply this regression for various horizons using panel data for the postwar period (1955-90) from the Penn World Tables and the Barro and Lee (1994) data set.[4] In our application, we consider two different sets of countries that engage in risk sharing (and therefore make up our artificial "world"): a set of twenty-one OECD countries and a more comprehensive set of forty-nine countries (see appendix). The OECD countries are of interest because they would likely be the first countries to experiment with macro markets. Their income risk, however, is likely to fall below that of developing countries. The larger set of forty-nine countries provides us with an estimate of the potential risk-sharing benefits in the event that a broader array of countries introduced macro markets. Because we have only one growth observation per country for long horizons, we are unable to estimate country-specific growth uncertainty for each country separately. Thus, the results from the regressions, which combine data from all the countries in the sample, reflect "average" growth uncertainty across countries.

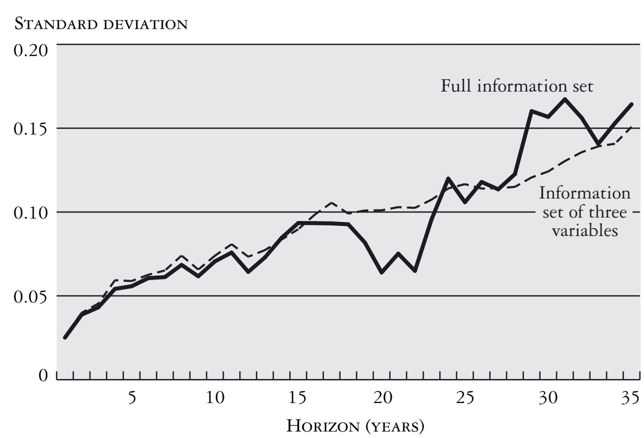
In choosing the variables that make up the information set, we draw on a large empirical and theoretical literature on economic growth.[5] Our base information set consists of thirteen variables: the log of per capita GDP; the most recent one- and five-year growth rates of per capita GDP; the most recent five-year population growth rate; the ratio of private consumption to GDP; the ratio of government consumption to GDP; the ratio of investment to GDP; openness as measured by exports plus imports as a fraction of GDP; gross enrollment ratios for primary, secondary, and higher education; the fertility rate; and life expectancy at birth.[6] We also consider a smaller information set consisting of the three variables with the most predictive power; that is, they led to the lowest estimated standard deviation of residual risk at a thirty-five-year horizon. For the set of forty-nine countries, these variables are the log of per capita GDP, the fertility rate, and the investment rate. For the OECD country set, the investment rate is replaced by enrollment in higher education.

Diversifiable Country-Specific Risk

Charts 1 and 2 show the standard deviation of residual risk as a function of the time horizon. For the base information set, the standard deviation of the growth rate at a thirty-five-year horizon is 16.4 percent for the set of OECD countries and 33 percent for the set of forty-nine countries. These numbers are very large, implying a 95 percent confidence interval of 66 percent for OECD countries and 132 percent for the forty-nine countries. The charts also show that the results for the smaller information set are almost the same as the results for the full information set. This similarity implies that adding more variables does not significantly help in predicting long-term growth rates.

Chart 1

Growth Uncertainty in the OECD Countries

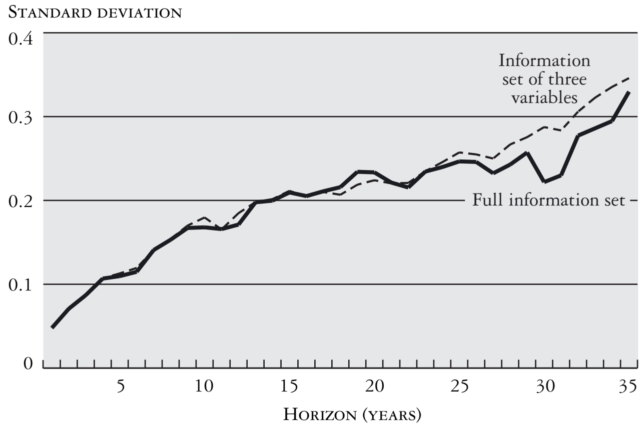


Source: Authors' calculations.

Notes: The chart shows the standard deviation of the unpredictable component of the difference between the per capita GDP growth of a representative OECD country and that of the world. The full information set used to predict growth consists of thirteen variables (see text). The information set of three variables consists of the log of per capita GDP, the fertility rate, and enrollment in higher education.

Chart 2

GROWTH UNCERTAINTY IN THE SET OF FORTY-NINE COUNTRIES



Source: Authors' calculations.

Notes: The chart shows the standard deviation of the unpredictable component of the difference between the per capita GDP growth of a representative country and that of the world. The full information set used to predict growth consists of thirteen variables (see text). The information set of three variables consists of the log of per capita GDP, the fertility rate, and the investment rate.

To get a better sense of the amount of uncertainty involved here, we perform a simple experiment. We take 10,000 draws from the distribution of residual risk for each country,[7] assuming that the draws are independent across countries and that each country's standard deviation of residual risk is the same. For the set of forty-nine countries, we use the results to compute the probability that per capita GDP of the best performing country will unexpectedly double, triple, quadruple, or quintuple relative to that of the worst performing country over the specified time horizon. The results are shown in Chart 3. The probability that the best performing country's per capita GDP doubles or triples relative to that of the worst performing country is practically 100 percent at the thirty-five-year horizon. The probability that the best performing country's per capita GDP quadruples or quintuples relative to that of the worst performing country is 81 percent and 44 percent, respectively. These results are striking. They suggest that, after controlling for the growth that had already been expected, per capita GDP of the best performing country is likely to rise by a factor of five relative to that of the worst performing country! Even at the short ten-year horizon, the probability that the per capita GDP of the best performing country would unexpectedly double relative to the per capita GDP of the worst performing country is 84 percent.

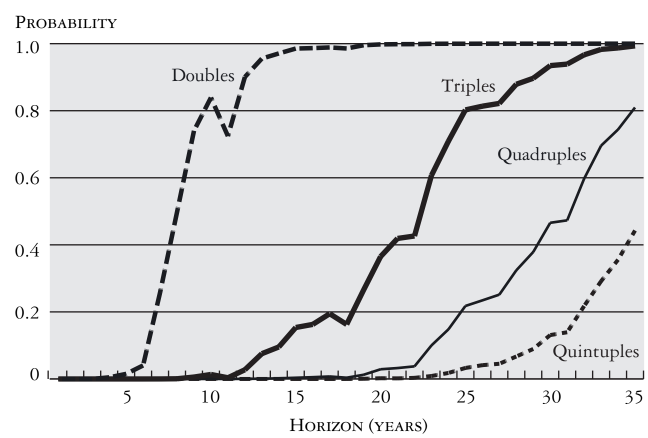
For the set of OECD countries, we report the probability that the per capita GDP for the best performing country rises by 30 percent, 50 percent, 70 percent, or 100 percent relative to that of the worst performing country (Chart 4). At a thirty-five-year horizon, the probabilities are 99.99 percent, 99.9 percent, 61 percent, and 13 percent, respectively. Although less spectacular, these numbers are still significant. Indeed, the best performing country's per capita GDP is likely to rise by 70 percent relative to the worst performing country's over a period of thirty-five years.

Because these figures only consider the very extremes, that is, the worst compared with the best performing countries, we also compute the probability that the unweighted average per capita GDP of the seven best performing countries doubles, triples, quadruples, or quintuples relative to the unweighted average of per capita GDP of the seven worst performing countries. For the set of forty-nine countries, at the thirty-five-year horizon the probabilities are 99.9 percent, 89.4 percent, 29 percent, and 3 percent, respectively. These results suggest that, contrary

Chart 3

PER CAPITA GDP: BEST PERFORMING COUNTRY VERSUS WORST PERFORMING COUNTRY

Forty-Nine Countries

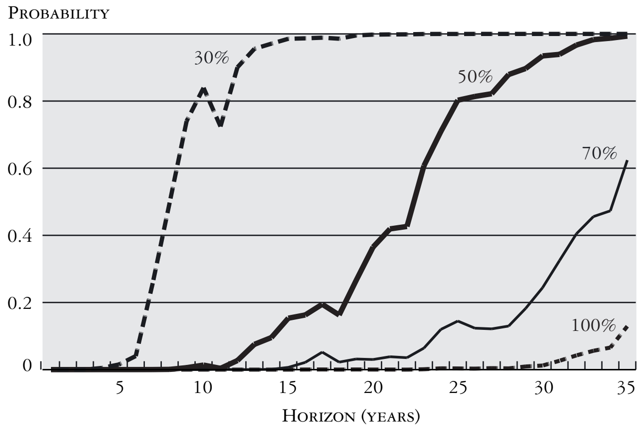


Source: Authors' calculations.

Notes: The chart shows the probability that the per capita GDP of the best performing country will unexpectedly double, triple, quadruple, or quintuple relative to that of the worst performing country. These probabilities depend on the growth horizon.

Chart 4

Per Capita GDP: Best Performing Country versus Worst Performing Country
OECD Countries



Source: Authors' calculations.

Notes: The chart shows the probability that the per capita GDP of the best performing country will unexpectedly rise by 30 percent, 50 percent, 70 percent, or 100 percent relative to that of the worst performing country. These probabilities depend on the growth horizon.

to expectation, the per capita GDP of the seven best performing countries as a group is likely to triple relative to that of the seven worst performing countries over thirty-five years. For the set of OECD countries, we find a probability of 88 percent that the unweighted average of per capita GDP of the three top-performing countries in the sample rises by 50 percent relative to that of the three worst performers. Note that in both of these cases we look at the best performing one-seventh and worst performing one-seventh of the countries in our sample.

To illustrate further that these numbers are not unrealistic, Chart 5 shows the expected deviation from world growth in 1955 for the thirty-five-year period 1955-90 (according to the information set of three variables) compared with the actual deviation from world growth over the same period. For the set of forty-nine countries, the best performing countries relative to the expectation in 1955 were Thailand and Japan. Several African and South American countries were the worst performers. Note that Thailand was expected to grow slightly less than Uruguay in 1955. In fact, however, Thailand's per capita GDP rose by a factor of 5.1 relative to that of Uruguay! Per capita GDP of the worst performing country in the sample, Nicaragua, dropped 22 percent over the period 1955-90. Some countries that are not in our sample performed even worse. Extreme cases include Nigeria, whose real per capita GDP declined 59 percent from 1976 to 1990, and Guyana, whose real per capita GDP dropped 59 percent from 1976 to 1990. For the world's poorest countries, hedging national income risks may truly be a matter of life and death for some citizens. In these countries, declines in national income have seriously harmed the quality of health care, nutrition, environmental protection, and law enforcement.

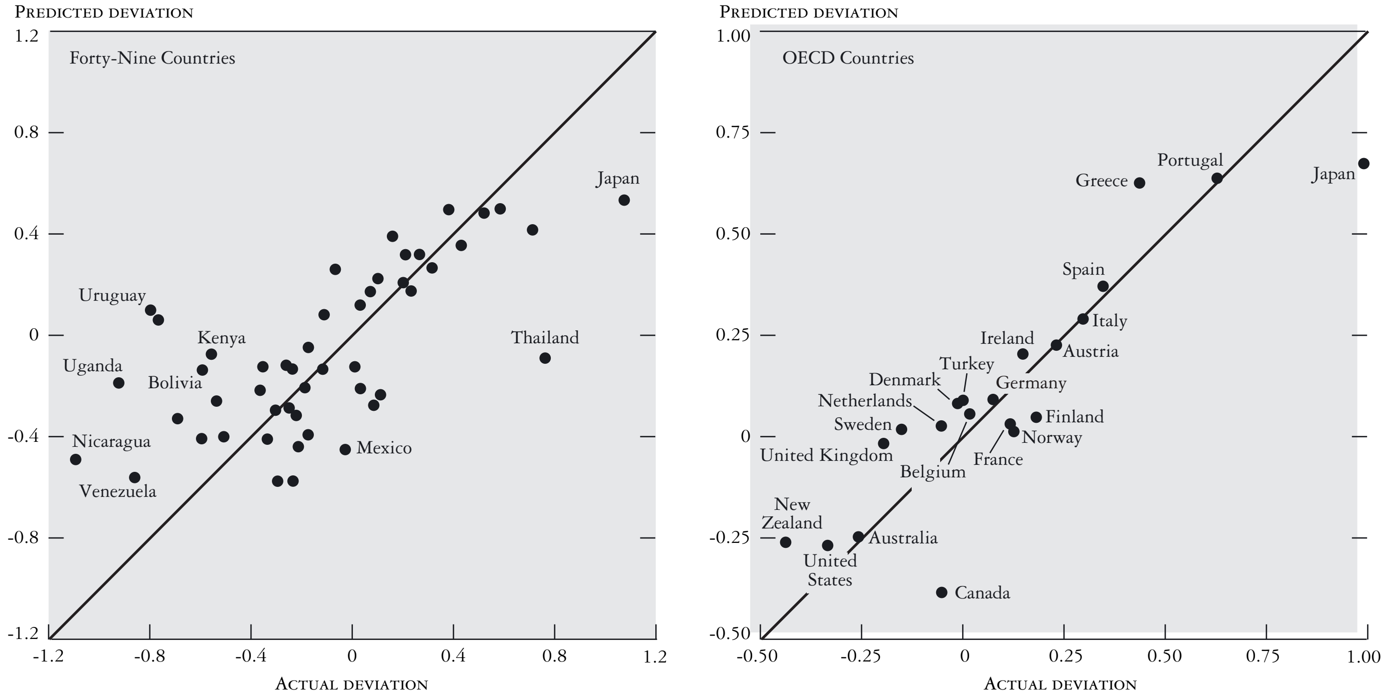
These results might leave the impression that only nations in Africa, South America, and East Asia are subject to large income shocks. Although these countries have experienced the most dramatic changes in per capita GDP during the past several decades, what matters today is uncertainty about *future* income. It is quite possible that over the next fifty years the biggest income surprises will come from other parts of the world. Large gains from risk sharing are therefore not necessarily limited to the set of countries that have faced the largest income shocks in recent years.

We see from Chart 5 that in our sample of OECD countries the best performing countries were Japan and Canada. In 1955, based on various indicators such as low investment, low school enrollment, high per capita income, and low recent growth rates, Canada was not expected to grow as fast as the average OECD country. Nonetheless, its growth rate turned out to be almost average. The worst performing countries were Greece, the United Kingdom, and New Zealand. Japan's per capita GDP grew 80 percent more than that of Greece, even though the two countries' expected growth rates were very similar in 1955. These results are suggestive of the significant uncertainty of relative performance among OECD countries. Of course, we caution against taking the results for individual countries too literally. The figures are somewhat sensitive to the precise information set and the countries considered. Nonetheless, this exercise provides a good sense of the degree of diversifiable uncertainty of future income.

Although our sample ends in 1990, a very recent and large growth surprise surfaced in Ireland. Ireland's economy stagnated during the first half of the 1980s. In 1987, its per capita GDP was 63 percent of Britain's. But

Chart 5

PREDICTED AND ACTUAL DEVIATION FROM WORLD GROWTH, 1955-90



Source: Authors' calculations.

Notes: These figures show the actual and predicted deviation of individual countries' per capita GDP growth from world per capita GDP growth. Here "world" is defined as the sum of the countries in the sample. Countries below the 45 degree line performed better than expected. Countries above the 45 degree line performed worse than expected.

only nine years later, in 1996, Ireland's per capita GDP surpassed Britain's. The economy expanded 10 percent in 1995 and 7 percent in 1996. Relative to expectations in the mid-1980s, this remarkable growth episode was clearly unexpected. Foreign direct investment contributed to growth, but even now it is hard to fully explain Ireland's spectacular growth performance.[8]

INDIVIDUAL-SPECIFIC RISK

In addition to aggregate income uncertainty, individuals must contend with income variations that are specific to their situation. Individual-specific risk cannot be shared through macro markets. Indeed, no institution can completely eliminate this type of risk because of moral hazard problems. How important are these individual risks? How much income variation is left after people have completely hedged their aggregate risks?

Fortunately, individual-specific income risk appears to amount to less than half of total income risk. Shiller and Schneider (1998), using 1968-87 U.S. data from the Panel Study of Income Dynamics, estimate the variance of income changes that are not under the control of individuals. They categorize individuals into seven occupational groupings according to objective factors such as retirement, employment, and educational status. They then compute an index of labor income for the United States for each grouping. The results show that between half and three-quarters of the variance of five-year income changes can be explained by the aggregate indexes. Most of people's income risk could therefore be managed through macro markets, assuming that they were opened not just on national incomes but, within that, on occupational incomes.

CAN EXISTING FINANCIAL MARKETS DO THE JOB?

In theory, existing financial markets could achieve most of the potential benefits from diversification if the aggregate return on domestic financial assets was highly correlated with the return of a claim on the present discounted value of aggregate income. This is the case when the return on human capital is highly correlated with the return on domestic financial assets. Consider an average individual whose current wealth consists of $900,000 in nontraded assets. Nontraded assets include both human capital and noncorporate business assets, but, for simplicity, here we will simply refer to both as human capital. An additional $100,000 of the individual's wealth is in financial assets, including pension funds. Now assume that the return on domestic financial capital is perfectly correlated with the return on domestic human capital. The individual can then achieve full diversification as follows. First, if the financial return has the same standard deviation as the human capital return, selling short domestic financial assets by $900,000 eliminates all domestic risk. After that, $1 million is invested globally ($100,000 of financial wealth plus the $900,000 of revenue from selling short domestic assets).

The correlation between the return on human capital and financial capital, however, is much smaller than one. Bottazzi, Pesenti, and van Wincoop (1996) compute this correlation using data for the years 1970-92 for OECD countries. The return on human capital is defined as the innovation in the present discounted value of wages divided by the current value of human capital.[9] The innovation is computed using the results from a vector autoregressive process for the wage rate and the profit rate or for the wage rate and a broad measure of return on domestic financial capital. A trend is extracted from both the wage rate and the profit rate. Three measures of the return on domestic financial capital are used: the profit rate (profits divided by the capital stock); the present discounted value of the profit rate, again using the results from the vector autoregressive process; and the weighted average of returns on stocks, long-term bonds, and short-term deposits (a broad measure of financial returns). Across countries, the average of the estimated correlation between the return on human capital and financial capital for the three measures is 0.26, -0.34, and -0.43, respectively—the correlations are all much smaller than one.

It is important to note that these correlations are based on wages and profits after extracting a trend. A common stochastic growth trend is likely to exist across countries.[10] Because such a common trend represents global risk, it cannot be shared among countries. Therefore, controlling for such a trend is appropriate for our purposes. It is useful to note, however, that the negative correlation for two of the measures is not inconsistent with a positive correlation between the "raw" returns on human capital and domestic assets. An improvement in global technology raises both profits and wages.

Macro markets would also allow individuals to invest in firms and companies that are not traded publicly.

There are many possible explanations for the absence of a strong positive correlation. First, shocks to the bargaining power of labor or a change in government can significantly affect the income distribution. Second, if wages are less flexible than prices, positive demand shocks will affect real wages and profits asymmetrically. Third, standard trade theory predicts that the wage rate and return to capital move in opposite directions in response to terms of trade shocks (Stolper-Samuelson).

An important question that we do not address is how much of the country-specific income growth uncertainty documented in the preceding section can be shared through existing financial markets. No research has yet been done to address that question. Nonetheless, the low correlations between the return on human capital and financial assets reported above suggest that macro markets have an important role to play in the diversification of aggregate income growth uncertainty, a role that existing financial markets cannot completely fill.

Macro markets would also allow individuals to invest in firms and companies that are not traded publicly. Stock indexes only include companies after they have become successful. But productivity growth is influenced by private firms and start-ups at least as much as by public companies. Thus, investment in stock indexes cannot capture the growth of these smaller companies. For an individual who wants to invest in a country because the fundamentals of the country are strong, buying a share of GDP would be more appropriate than buying a stock index.

Why Don't Macro Markets Exist?

If the potential benefits of aggregate income markets are so large, and the underlying risk management concepts are apparently so simple, why have they not already developed in the private sector? Surely, significant commissions could be earned if a large demand for these securities developed. Surely, there ought to be some niche for these securities somewhere in the world. And yet there is no evidence that markets like these have ever existed. In principle, macro markets would not be difficult to introduce. In 1997, the U.S. Treasury introduced inflation-indexed bonds. The only essential difference is that in macro markets the coupons would be indexed to a measure of aggregate income rather than to the consumer price index (CPI). It is important, therefore, to try and understand what barriers stand in the way of the creation of macro markets.

Not So Obvious

The first thing to note is that while the concept of risk management is very basic, the idea of markets that share income risks is not so obvious as to occur immediately to most people. The idea of markets in aggregate incomes is like other important inventions in the history of technology that have seemed extremely simple after they were implemented—simple, that is, from the vantage point of people viewing the final invention and not the idea that preceded it. For example, rejecting a proposal for investment in radio technology in the 1920s, David Sarnoff's Associates wrote, "The wireless music box has no imaginable commercial value. Who would pay for a message sent to nobody in particular?" Between 1939 and 1944, more than twenty companies rejected the idea of Chester Carlson, inventor of the Xerox machine, to copy a document on plain paper. Although the idea was considered useless at the time, today Rank Xerox Corporation earns annual revenues of about $1 billion, and it is hard to imagine life without the machine.

Establishing markets for long-term claims on flows of income aggregates is no more obvious than other recent financial innovations. Even the concept of national income itself is a relatively new invention that has been perfected over many years. Developed earlier in this century by Kuznets (1937), Stone (1947), and others, the concept of national income as we know it did not become widely accepted until after World War II.

Similarly, many risk management institutions that are now commonplace have gotten off to slow starts. For example, markets in foreign currency swaps—which now account for about half the gross turnover in the foreign exchange market—did not develop until the early 1980s. A futures market in stock price indexes also did not develop until 1982. An even more recent innovation is the creation of indexed bonds. Economists have been pointing out the dangers of long-term nominal contracting for more than a hundred years, and yet in the United States long-term debt has been almost exclusively nominal. Indexed federal government debt did not exist in the United States until 1997, and it still only accounts for less than 1 percent of the federal debt.[11] Brainard and Dolbaer (1971) have long pointed out the advantages of creating contracts that allow people to share occupational income risks, but serious discussion of such contracts has only just begun.

Establishing markets for long-term claims on flows of income aggregates is no more obvious than other recent financial innovations.

Potential for Failure

Not only do market innovations take a long time to start, they also often fail. Those who contemplate taking the time and effort to establish such markets may be deterred by past failures. A good example of such a failure is the CPI futures market, which bears some resemblance to the macro markets described here.

A CPI futures market allows an investor to hedge against a change in real income that occurs when nominal income is rigid and the price level changes. CPI markets were proposed in the 1970s by Lovell and Vogel (1974) at a time when U.S. inflation was high. The Lovell-Vogel proposal launched a discussion of the benefits of the CPI market, attracting endorsements from such prominent economists as Milton Friedman and Paul Samuelson. Despite this interest, it took a dozen years before the CPI market was established in the United States at the Coffee, Sugar, and Cocoa Exchange in 1985. Unfortunately, by the time the market was established, the inflation rate (as well as inflation uncertainty) had fallen to a fairly low level. As a result, the relatively short-term contracts had virtually no hedging function. Despite some early activity, the market was essentially dead by 1986.

The failure of the CPI futures market in the United States is often cited as evidence that the idea behind the market was flawed. A CPI market did succeed, however, in Brazil. The market started around the same time as in the United States, 1986, but inflation uncertainty was much higher in Brazil than in the United States. The Brazilian market flourished until it was shut down by the Brazilian government as an anti-inflation measure.[12] The lesson that can be learned from the CPI futures market is not that such markets cannot succeed but that they are slow to get started. Moreover, they must be started while the risks that the market is designed to manage are prominent.

Lack of Investor Awareness

It may be that people simply are not aware of long-term income growth uncertainty and the exposure of their own incomes to aggregate risk. Investors frequently emphasize short-term over long-term portfolio performance. One potential factor behind such a short-term focus is the agency problem associated with the delegation of financial market decisions. The difficulty in monitoring decisions carried out by an outside agency naturally leads to an overemphasis on easily observable short-term performance.

Individuals might not be aware of their exposure to aggregate income growth uncertainty because short-run fluctuations in their own income often appear to be independent of fluctuations in aggregate incomes. This narrow focus could lead them to underestimate the long-term correlation between individual income and aggregate income. Most people are probably not aware that over longer time intervals, individual's incomes tend to rise and fall with aggregate income. As we mentioned above, even at the relatively short five-year horizon, most of an individual's income growth uncertainty can be attributed to aggregate risk. Nonetheless, many people attribute these income fluctuations to their own efforts and abilities as well as to luck. This lack of awareness raises doubts about whether large-scale demand in macro markets would ultimately materialize, even though in principle the diversification benefits are high.

It may be that people simply are not aware of long-term income growth uncertainty and the exposure of their own incomes to aggregate risk.

Lack of Price History

We have yet to find a single example of a mutual fund that advertises a low or negative correlation of its returns with income aggregates as one of its selling points—even though finance theory suggests that such a correlation is one of the most important things to advertise. One explanation for the failure of mutual funds to advertise such a correlation is that claims on income aggregates have no market price and therefore no observable return. No one knows how volatile the price of aggregate income claims would be. Only the history of the income movements themselves is observable. Consider the case of investors who own corporate stock. If individuals could observe only dividend announcements and not the price, no one would know the amount of volatility present in stock prices.[13]

The Importance of Public Debate and Leadership

One reason aggregate income markets do not exist is that there has been very little public debate about the potential goals of such markets. Kennickell, Starr-McCluer, and Sundén (1996) find that friends and relatives are the most important source of financial advice. Others' actions clearly provide an important signal for most people. Thus, a broad consensus on the value of macro markets among financial advisors, writers, commentators, lawyers, regulators, and lawmakers is very important if risk management contracts are to be sold to the public. Historical evidence suggests that professional leadership is an important factor in making risk management institutions a success. Consider, for example, disability risk insurance. In the early part of this century, private disability insurance was available but the public showed little interest in it. Only through the work of economists—notably John R. Commons, a cofounder of the American Economic Association—did the state-government institution of Worker's Compensation become established in the United States in all but six states by 1920.[14] Since then, disability insurance has become common among private employers as well. Today, disability insurance is a well-established institution that is not exclusively governmental, even though relatively little disability insurance is sold directly to individuals by insurance companies.

A Public Goods Problem

Another reason why these securities may not exist is that market innovators typically capture a very small fraction of the benefits and almost all of the costs of introducing a new market. Financial instruments or ways of doing business usually cannot be patented. Evidence indicates that when a firm successfully issues a new financial product, a competitor typically introduces a similar product within a period of less than two or three months.[15] At the same time, the introduction of aggregate income assets requires substantial initial investments from the innovator, including data collection, publicizing the product, experimenting with different types of contracts, and educating the public on how to use these markets.

Measurement Problems

What these contracts should cash-settle on is a serious issue that poses significant measurement problems. Per capita income measures can change based on shifting demographics alone. One solution may be to keep track of the incomes of a large group of individuals. Changes in quality are also notoriously hard to measure. Beyond such measurement issues is the question of how to deal with revisions. Shiller (1993) advances the theory of index numbers to address these questions. He proposes several kinds of chain indexes that are relatively robust to revision problems, and adjustments to national income measures could be made along these lines. Attempts to generate labor income indexes that are less sensitive to the changing composition of the labor force are reported in Shiller and Schneider (1998). The standardization of the indexes is essential to creating liquidity in these markets. A related problem is that governments collect most of the data to compute these indexes. If countries sell short claims on their own income, which they should do for the purpose of risk sharing, governments have an incentive to underreport GDP. It is not immediately clear how to resolve the problem of underreporting, although similar problems have not stopped the development of markets in indexed bonds and CPI futures.

Historical evidence suggests that professional leadership is an important factor in making risk management institutions a success.

Problems of Enforcement

Enforceability may also be a significant obstacle. In the formation of macro markets, contract designers need to avoid incentives for investors to renege on contracts. Consider the hedging instruments discussed earlier, which yield an annual payoff of $GDP_W - GDP$. Domestic residents buy such securities from pension funds to eliminate their exposure to country-specific aggregate risk. But when per capita output in their own country unexpectedly grows faster than per capita world output, they lose on the contract. In order to guarantee their ability to pay, domestic residents must put up margin. These margin calls can be very large because the expected present discounted value of a country's per capita GDP can fluctuate widely. The amount of margin required shrinks as the margin is adjusted more frequently because at shorter time intervals the uncertainty about asset price changes is smaller. Nonetheless, as we saw in October 1987 and October 1997, sometimes very large asset price changes are observed even over very short periods of time. High levels of margin may push individuals who do not have sufficient liquid assets out of the market. One advantage of arranging these contracts through pension funds is that the money already invested in the fund can be applied as margin. Very young investors—whose pension accounts are still small—may not be able to fully diversify against aggregate income risk. This problem gradually improves as an investor gets older. Most middle-aged people have accumulated sufficient wealth to take full advantage of the option to hedge aggregate risk. But as an investor gets older, the horizon for hedging becomes shorter and the benefits from risk sharing decrease.

Macro Market Bubbles

An additional problem is that the price of the macro securities may be even more volatile than the underlying fundamentals. Asset price bubbles cannot be ruled out. An asset price bubble occurs when increasing optimism causes investors to bid up prices to unsustainable levels, eventually resulting in a bursting of the bubble and a sudden crash. By some accounts, bubbles are caused in part by individuals who overreact to past positive returns and flock into a bull market. Investors who enter the market because of excessive optimism typically choose to depart once they find that their optimism is unfounded and can cause a market to crash.

Stock market crashes have sometimes had significant repercussions on economic performance. The worldwide stock market crash of 1929, for example, appears to have triggered a public sense of great uncertainty and a desire to postpone expenditures until the economic outlook grew clearer (see Romer [1990]). This reaction may have been a factor in bringing on the Great Depression. The consequences of such price swings in macro markets, and safety measures to protect against such shocks, need to be considered and addressed.

Portfolio Management Problems

Finally, we would like to address a very practical question. Given the uncertainties surrounding a person's future income, future employment, and future career developments, how will he or she know what positions to take in these markets? In our earlier example, we assumed that the

In the end, almost all people are sensitive to the growth performance of the aggregate economy, no matter where or in what sector they work.

individual's wages are equal to the per capita wage rate plus an idiosyncratic component unrelated to aggregate risk. But in reality, some people's income is more exposed to the national business cycle than others'. This exposure depends on the location of someone's work as well as the sector in which he or she works. In general, the optimal positions in the aggregate income markets depend on how much one's future income is correlated with measures of aggregate income over long-term horizons. Depending on the sector and location of someone's work, information about long-term income fluctuations can be obtained from historical data. But what happens when someone moves to another part of the country or to another sector, or when someone changes careers altogether? Of course, every person's career has a significant idiosyncratic component. What is really needed, however, is a good estimate of the aggregate component of a person's future income that takes into consideration characteristics such as age, education, location, and the sector in which he or she works. Financial advisors can use this information, which can be obtained from longitudinal data sources, to determine optimal hedging strategies. Obtaining such measures of covariances is a very difficult task. We do not want to exaggerate this difficulty, however. Over longer horizons, which matter most for diversification purposes, people's incomes are more correlated than they are over short horizons. In the end, almost all people are sensitive to the growth performance of the aggregate economy, no matter where or in what sector they work.

Conclusion

We have outlined how macro markets can be beneficial to the average person interested in his or her long-term financial security. The introduction of such markets allows pension funds to offer a hedging instrument that can be used to reduce, or even eliminate, exposure to country-specific growth performance. We have found that the benefits of eliminating exposure to such country-specific risk are large. Over a period of thirty-five years, the per capita GDP of one industrialized country relative to that of another industrialized country could unexpectedly double. For a broader group of countries, the risks are much larger. While not documented in this paper, large gains are likely to be achieved by trading other forms of aggregate income claims, particularly those associated with occupational risks. We have also pointed out that existing financial markets are not a good substitute for macro markets that cash-settle on a measure of national income.

Given that macro markets can provide substantial improvements in long-term financial security—improvements that cannot be achieved in existing markets—it may seem peculiar that these markets have not yet developed. We offer several explanations for the absence of macro markets. Investors tend to be focused on short-term financial performance and may not consider the benefits of long-term financial security. Moreover, research has shown that for most people, friends and family represent the main source of financial advice. It is therefore unlikely that investors will consider the benefits of protecting themselves from country-specific risks until a broad consensus develops on the value of macro markets among financial advisors, writers, lawyers, the media, regulators, and lawmakers.

Before aggregate income contracts can be introduced, many practical hurdles must be overcome. Rules for settlement need to be developed, and decisions must be made about income measures, contract size, and margin requirements. Circuit breakers or other measures that deal with the possibility of sudden booms or crashes in the macro markets will be necessary. An array of regulatory and tax issues will need to be resolved. Perhaps most important, methods for evaluating the aggregate income risk exposure of individual households and businesses will need to be developed so that people will know how to use the markets. Given the costs of introducing such markets, it is also important to think about where the first markets should be created and whether initial markets should be for individual countries or for aggregates of countries.[16]

Some of the hurdles to a wide-scale use of macro markets could turn out to be too large. Margin requirements to enforce the contracts may be too big for many individuals. It may also be difficult to determine optimal exposure to aggregate income risk for individual people and to convince investors of the benefits of hedging this risk. Even if these markets are eventually introduced, they may be used more narrowly than has been suggested here. The presence of these obstacles, however, does not mean that we should avoid serious debate about the creation of aggregate income markets. Aggregate income growth uncertainty represents the largest macroeconomic risk incurred by households all over the world. The benefits from trading in macro markets are potentially very large. Factors that are essential to the start of such markets—including well-functioning financial exchanges, a sophisticated technology of trading, and the intellectual appreciation of the importance of risk management—are already in place. Eventually, portfolio managers and individuals could routinely hedge aggregate income risks in macro markets.

Appendix: Two Sets of Countries

We use two sets of countries in the regression analysis—a set of forty-nine countries and a smaller set of twenty-one OECD countries.

The forty-nine countries are Kenya, Mauritius, Uganda, Canada, Costa Rica, the Dominican Republic, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Trinidad and Tobago, the United States, Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, Venezuela, India, Japan, Pakistan, the Philippines, Sri Lanka, Thailand, Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Turkey, the United Kingdom, Australia, and New Zealand.

The twenty-one OECD countries are Canada, the United States, Japan, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Turkey, the United Kingdom, Australia, and New Zealand.

ENDNOTES

The authors thank Phil Strahan and two anonymous referees for many useful comments and suggestions.

1. See Kang and Stulz (1997), French and Poterba (1991), and Tesar and Werner (1994, 1997).

2. See Shiller and Athanasoulis (1995) and Athanasoulis and Shiller (1997); for related work, see also Demange and Laroque (1995) and Allen and Gale (1994).

3. The country-specific growth uncertainty can also be transformed into a measure of welfare gains from international risk sharing. See Athanasoulis and van Wincoop (1999) and van Wincoop (1994, 1996, 1999).

4. See Athanasoulis and van Wincoop (1997) for details on the estimation procedure. For each horizon *s*, we use data for all non-overlapping intervals with that length, starting with the most recent interval ending in 1990.

5. See Barro and Sala-i-Martin (1995) and Levine and Renelt (1992).

6. We experimented with additional variables: political instability; terms of trade growth over the past five years; percentage of primary, secondary, and higher education attained; the most recent one-year and five-year growth rates of private consumption; and the investment rate averaged over the past five years. None of these variables improved predictive power substantially.

7. The residual risk is based on the three variables that have the most predictive power.

8. See *The Economist*, May 17-23, 1997, pp. 21-4, for a discussion of Ireland's recent growth.

9. The wage rate is the average real wage per employee using national data on employee compensation divided by the number of employees and the consumer price index.

10. Plenty of evidence suggests that technological convergence occurs across industrialized countries, leading to a common stochastic growth trend.

11. See Shiller (1997) for a discussion of public resistance to indexation.

12. Similar markets were reintroduced twice in the late 1980s. However, each time they were eventually shut down by the government.

13. This problem is not insurmountable. Initial public offerings face the same problem.

14. See Moss (1995).

15. See Tufano (1992).

16. Shiller and Athanasoulis (1995) find that a U.S.-Japan swap of national incomes may be the best single contract to recommend, with a U.S.-Europe swap being important as well. Athanasoulis and Shiller (1997) find that an important market to develop early would be a market for the entire world, a market that would trade claims on the aggregated incomes of all countries.

REFERENCES

Allen F., and D. Gale. 1994. FINANCIAL INNOVATION AND RISK SHARING. Cambridge: MIT Press.

Athanasoulis, S. G., and E. van Wincoop. 1997. "Growth Uncertainty and Risksharing." Federal Reserve Bank of New York STAFF REPORTS, no. 30 (October).

Athanasoulis, S. G., and R. J. Shiller. 1997. "The Significance of the Market Portfolio." NBER Technical Working Paper no. 209.

Barro, R. J., and J. Lee. 1994. "Sources of Economic Growth." CARNEGIE ROCHESTER CONFERENCE SERIES ON PUBLIC POLICY 40: 1-46.

Barro, R. J., and X. Sala-i-Martin. 1995. ECONOMIC GROWTH. New York: McGraw-Hill.

Bottazzi, L., P. Pesenti, and E. van Wincoop. 1996. "Wages, Profits and the International Portfolio Puzzle." EUROPEAN ECONOMIC REVIEW 40, no. 2 (February): 219-54.

Brainard, W., and F. T. Dolbaer. 1971. "Social Risk and Financial Markets." AMERICAN ECONOMIC REVIEW 61, no. 2: 360-70.

Davis, E. Philip. 1995. PENSION FUNDS: RETIREMENT-INCOME SECURITY AND CAPITAL MARKETS: AN INTERNATIONAL PERSPECTIVE. Oxford: Clarendon Press.

Demange, G., and G. Laroque. 1995. "Optimality of Incomplete Markets." JOURNAL OF ECONOMIC THEORY 65, no. 1 (February): 218-32.

French, K. R., and J. M. Poterba. 1991. "Investor Diversification and International Equity Markets." AMERICAN ECONOMIC REVIEW 81, no. 2 (May): 222-6. Papers and Proceedings of the 103rd Annual Meeting of the American Economic Association, December 1990.

Gehrig, T. P. 1993. "An Information Based Explanation for the Domestic Bias in International Equity Investment." SCANDINAVIAN JOURNAL OF ECONOMICS 95, no. 1: 97-109.

Kang, J., and R. M. Stulz. 1997. "Why Is There a Home Bias? An Analysis of Foreign Portfolio Equity Ownership in Japan." JOURNAL OF FINANCIAL ECONOMICS 46, no. 1: 3-28.

Kennickell, A. B., M. Starr-McCluer, and A. E. Sundén. 1996. "Financial Advice and Household Portfolios." Working paper, Board of Governors of the Federal Reserve System.

Kuznets, S. 1937. NATIONAL INCOME AND CAPITAL FORMATION. New York: National Bureau of Economic Research.

Levine, R., and D. Renelt. 1992. "A Sensitivity Analysis of Cross Country Growth Regressions." AMERICAN ECONOMIC REVIEW 82, no. 4: 942-63.

Lovell, M. C., and R. C. Vogel. 1974. "A CPI-Futures Market." JOURNAL OF POLITICAL ECONOMY 81, no. 1: 9-12.

Merton, R. C. 1983. "On Consumption-Indexed Public Pension Plans." In Z. Bodie and John B. Shoven, eds., FINANCIAL ASPECTS OF THE UNITED STATES PENSION SYSTEM, 259-90. NBER and University of Chicago Press.

Moss, D. A. 1995. SOCIALIZING SECURITY: PROGRESSIVE ERA ECONOMISTS AND THE ORIGIN OF AMERICAN SOCIAL POLICY. Cambridge: Harvard University Press.

Romer, C. 1990. "The Great Crash and the Onset of the Great Depression." QUARTERLY JOURNAL OF ECONOMICS 105, no. 3: 597-624.

Shiller, R. J. 1993. MACRO MARKETS: CREATING INSTITUTIONS FOR MANAGING SOCIETY'S LARGEST ECONOMIC RISKS. New York: Clarendon Press.

———. 1997. "Public Resistance to Indexation: A Puzzle." BROOKINGS PAPERS ON ECONOMIC ACTIVITY, no. 1: 159-211.

Shiller, R. J., and S. Athanasoulis. 1995. "World Income Components: Measuring and Exploiting International Risk Sharing Opportunities." NBER Working Paper no. 5095.

Shiller, R. J., and R. Schneider. 1998. "Labor Income Indices Designed for Use in Contracts Promoting Income Risk Management." REVIEW OF INCOME AND WEALTH 44, no. 2: 163-82.

Stone, R. 1947. "Definition and Measurement of the National Income and Related Totals." In Subcommittee on National Income Statistics of the League of Nations Committee of Statistical Experts, Studies and Reports on Statistical Methods no. 7, MEASUREMENT OF NATIONAL INCOME AND THE CONSTRUCTION OF SOCIAL ACCOUNTS, 21-116. Geneva: United Nations.

Tesar, L. L., and I. M. Werner. 1994. "International Equity Transactions and U.S. Portfolio Choice." In J. A. Frankel, ed., THE INTERNATIONALIZATION OF EQUITY MARKETS. Chicago: University of Chicago Press.

———. 1997. "The Internationalization of Securities since the 1987 Crash." Working paper, University of Michigan.

References (*Continued*)

Tufano, P. 1992. "'Financial Innovation and First Mover Advantage." JOURNAL OF APPLIED CORPORATE FINANCE 5: 83-7.

van Wincoop, E. 1994. "Welfare Gains from International Risksharing." JOURNAL OF MONETARY ECONOMICS 34, no. 2 (October): 175-200.

———. 1996. "A Multi-Country Real Business Cycle Model with Heterogeneous Agents." SCANDINAVIAN JOURNAL OF ECONOMICS 98, no. 2: 233-51.

———. 1999. "How Big Are the Potential Welfare Gains from International Risksharing?" JOURNAL OF INTERNATIONAL ECONOMICS 47, no. 1: 109-35.

Exchange Rates and Profit Margins: The Case of Japanese Exporters

Thomas Klitgaard

Exporters must make a pricing decision whenever exchange rates change. A rise in the yen's value, for example, forces Japanese exporting firms to decide how much to alter the prices seen by their foreign customers. At one extreme, they could lower the yen price of their exports so that the dollar price of their sales to the United States would remain unaffected. The yen's rise would then have no impact on the price and volume of U.S. imports from Japan, but it would have adverse consequences for Japanese profit margins. At the other extreme, Japanese firms could keep the yen price of their exports unchanged so that the yen's rise would be completely passed through to U.S. consumers in the form of higher prices in dollar terms. The profit margins of Japanese exporters would then remain unchanged, but the volume of sales would drop, leading to a lower *level* of profits.

This article examines how Japanese exporters are responding to the conflicting objectives of maintaining stable profit margins and stable export sales when the value of the yen fluctuates. We find that Japanese firms tend to strike a balance between these goals. The firms' foreign customers do see exchange-rate-driven changes in prices, but the firms moderate the extent of these changes by altering their profit margins. For Japanese exporters producing industrial machinery, electrical machinery, and transportation equipment, our analysis suggests that a 10 percent rise in the yen leads to a roughly 4 percent decline in export margins (relative to the margins on goods sold in Japan) when other factors are held constant. That is, exporters in these industries pass on more than half of any change in the yen to the price seen by their foreign customers and absorb the remainder by adjusting profit margins on their foreign sales.

We also address other key issues related to the behavior of profit margins. For example, the short-run response of profit margins to exchange rate movements appears to be most pronounced in the transportation equipment and electrical machinery industries. One explanation

Thomas Klitgaard is a senior economist at the Federal Reserve Bank of New York.

for this is that these two industries have a greater tendency to invoice their goods in foreign currency terms so that prices, when measured in yen terms, respond automatically to exchange rate swings. In addition, the direction of the yen's movement is found to have no effect on how willing Japanese exporters are to use profit margins to stabilize prices in foreign markets. Firms are just as likely to raise profit margins when the yen depreciates as they are to cut margins when the yen appreciates. Finally, an examination of profit margins since the beginning of the Asian currency crisis in mid-1997 reveals that only firms in the electrical machinery industry have altered their pricing behavior. The higher than expected profit margins on exports from this industry, however, do not appear to be related to how these firms are responding to yen movements.

The consequences for the U.S. economy from a change in the yen's value depend on how much of any change is passed through by Japanese exporters to the prices seen by their U.S. customers.

Exchange Rates and Export Prices

The consequences for the U.S. economy from a change in the yen's value depend on how much of any change is passed through by Japanese exporters to the prices seen by their U.S. customers.[1] Chart 1 shows that by 1995, prices of imports from Japan were up roughly 20 percent from 1991, the first year for which data are available. But if Japanese export prices had remained unchanged in yen terms (meaning a full pass-through of changes in the yen), then U.S. import prices of Japanese goods would have tracked the dollar/yen index. The U.S. import price of Japanese goods did not rise by nearly as much as the dollar/yen exchange rate in the first half of the 1990s because Japanese firms lowered their export prices in yen terms. When the yen started to depreciate in early 1995, the price of U.S. imports from Japan did not fall as much as the currency rate did because Japanese firms took this opportunity to lift their yen export prices back up.

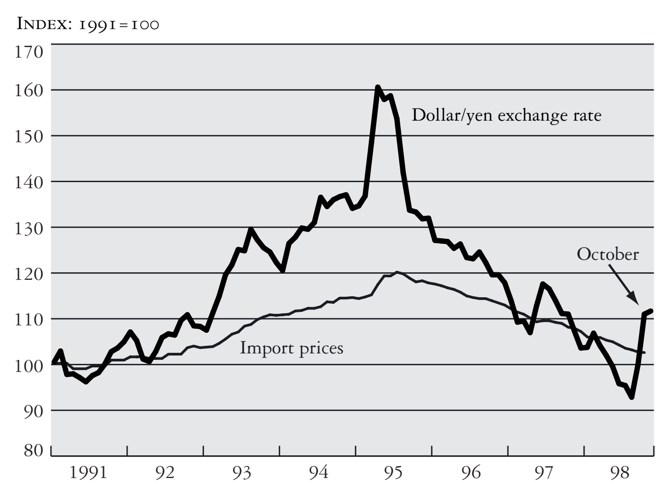
There are two possible explanations for how Japanese firms are able to offset yen movements. One is that the yen has a significant impact on production costs. For example, a stronger yen lowers the cost of imported inputs. The drop in production costs, in yen terms, then makes it easier for firms to lower yen export prices (Box A). The second explanation is that Japanese exporters absorb part of the yen's movement into their profit margins, an action that reduces the profit on each item sold when the yen appreciates and raises the profit margin when the yen depreciates. It is this latter explanation that will now be explored.

The Yen's Impact on Profit Margins

Studies that focus on profit margins are derived from the belief that firms, having found a profit-maximizing price

Chart 1

U.S. Import Prices for Japanese Goods



Sources: U.S. Department of Labor, Bureau of Labor Statistics; Federal Reserve Bank of New York.

Box A: The Yen's Impact on Input Costs

The yen affects the price of imported goods used by Japanese exporters. When the yen appreciates, the yen price of imports such as oil and other commodities, which are typically contracted in dollar terms, drops. During the yen's rise from 1990 to 1995, for example, import prices fell 25 percent, suggesting that exporters were significantly aided in their efforts to reduce the yen price of their goods during this period through lower production costs.

The data, though, suggest that any help for Japanese exporters trying to stabilize the foreign currency price of their goods is limited. For example, from 1990 to 1995, the raw material price of fuel in the Japanese wholesale price index fell 25 percent, in line with the overall import price index. Over the same period, the intermediate price of energy—the price seen by manufacturers—fell by a more modest 7 percent. Indeed, prices for all inputs (both imports and domestically produced goods) declined by roughly 5 percent for the industrial machinery, transportation equipment, and precision equipment industries during the yen's sharp rise in first half of the decade (see table). Input prices of electrical machinery fell more sharply, by 13 percent, but this was due in part to the strong deflationary trend in this industry.

The modest change in input costs for manufacturers, relative to the size of the yen's movements, was also evident when the yen depreciated from 1995 to 1998. Import prices rose 18 percent, and the wholesale price of raw material fuel jumped 15 percent. Nevertheless, the intermediate price of energy was up 2 percent and the index of input prices for Japan's four major exporting industries fell in the range of 0 to 4 percent.

In general, input prices seen by Japanese firms are much more stable than exchange rates, a factor that limits how much exchange-rate-induced swings in production costs can help Japanese exporters offset currency movements. The more modest the impact the yen has on production costs, the more the efforts to offset the yen's impact on the foreign currency price of Japanese exports must be accomplished by varying profit margins.

Percentage Change in Input Prices

Industry	1990-95	1995-98
Industrial machinery	-3.5	-0.3
Electrical machinery	-12.0	-3.9
Transportation equipment	-3.7	-1.6
Precision equipment	-5.7	-1.6

Source: Bank of Japan.

Notes: Percentage changes are not annualized. Data for 1998 are through October.

in each market based on supply and demand conditions, will attempt to keep their goods close to this price. This idea of pricing to market means that firms are inclined to absorb currency swings into their profit margins in order to stabilize prices seen by foreign customers.[2]

Data on profit margins of exports and domestic sales are not available separately, but we can construct a good proxy variable since Japanese wholesale price data contain two separate price indexes. One is for prices charged by Japanese firms to their foreign customers and the other is for prices charged to their domestic customers. Changes in the ratio of the two indexes can be interpreted as measuring changes in firms' relative profit margins.[3] That is, if export prices are falling relative to prices charged to domestic customers, then the markup over costs for exports has fallen relative to the markup for domestic sales, assuming the same production costs for both sets of goods.

A comparison of export prices and domestic prices reveals that profit margins on exports relative to domestic sales varied substantially over the 1990s. Table 1 shows that the prices Japanese firms charged domestic customers in Japan's four major exporting industries were stable from 1990 to 1995,[4] with the exception of those for electrical machinery, which declined 13 percent (in line with the deflationary trend of prices in this industry). During the

Table 1
PERCENTAGE CHANGE IN JAPANESE GOODS' EXPORT AND DOMESTIC PRICES IN JAPAN'S FOUR MAJOR EXPORTING INDUSTRIES

	1990-95		1995-98	
Industry	Export Prices	Domestic Prices	Export Prices	Domestic Prices
Industrial machinery	-6.7	-1.1	10.6	0.0
Electrical machinery	-35.0	-12.7	2.6	-14.9
Transportation equipment	-8.9	-1.9	23.3	-1.7
Precision equipment	-9.8	-1.3	13.0	-0.9

Source: Bank of Japan.

Notes: Percentage changes are not annualized. Data for 1998 are through October.

same period, export prices dropped 7 to 10 percent for all industries except electrical machinery, which fell 35 percent. As a consequence, there were substantial differences in the markup between domestic and foreign sales in 1995 compared with markup differences in 1990.

Divergences also occurred when the yen started falling in 1995. From 1995 to 1998, domestic prices continued on the same path, with prices being stable except for electrical machinery prices, which fell 15 percent. But the behavior of export prices was very different. Prices for industrial machinery were up 11 percent and those for transportation equipment were up 23 percent. Even electrical machinery, with its strong deflationary trend, saw prices rise 3 percent during this period.

These data show that Japanese firms charge different prices for their Japanese and foreign customers when the yen changes value. Firms try to stabilize prices as seen by their customers in both foreign and domestic markets, which means that their relative profit margins rise and fall with the yen. The next section explains why a profit-maximizing strategy leads firms to vary markups across markets.

PROFIT MAXIMIZING WITH EXCHANGE RATE CHANGES

A useful model for understanding why profit margins vary between domestic and foreign markets was developed by Marston (1990). Firms in this model are assumed to have some control over their prices because of product differentiation or some other market imperfection. Manufacturers produce goods locally but sell them in both domestic and export markets. These firms charge P_h (in yen terms) in the domestic market and P_x (in foreign currency terms) in export markets. We assume that imperfect arbitrage between markets allows prices to differ in each market. Therefore, firms can take advantage of the profit-maximizing strategy of setting prices according to each market's demand characteristics.

To illustrate, we think of prices as markups over the same marginal costs:

(1) $$P_h = M(P_h/P, Y)MC \text{ and}$$
$$P_x/S = N(P_x/P^*, Y^*)MC.$$

The exchange rate, S, is foreign currency per yen, P is the general price level, and Y is income. The asterisk represents foreign variables. The destination-specific markup functions M and N depend on price elasticities of demand in each market and how these elasticities change with prices.[5] A number of factors—such as consumer tastes, the substitutability with competing products, and the firm's market share—dictate these demand characteristics. Any gap between P_h and P_x/S reflects differences between the markup functions M and N.

An exchange rate movement alters profit margins on exports because firms know that letting prices automatically rise when their currency falls reduces the demand for their goods.

Essentially, the model says that with common production costs, differences in prices for any particular market are based on marginal revenue calculations made by the firm, which in turn are dictated by the responsiveness of demand to changes in prices. Any negative relationship between demand and prices implies a negative relationship between prices and markups.[6] An exchange rate movement alters profit margins on exports because firms know that letting prices automatically rise when their currency falls reduces the demand for their goods. As a result, export prices, in foreign currency terms, do not adjust one-to-one with exchange rate changes.

Production costs and income also affect relative profit margins when consumer demand characteristics differ across markets. For example, an increase in input costs such as energy would push a firm to raise prices in both markets, but not necessarily by the same amount. The relative change in the two prices depends on how customers in the two markets react to higher prices. The impact of income on profit margins is determined by differences in demand elasticities with respect to prices and to income. The dependence on differences in demand characteristics across markets means that the model does not require any particular direction for relative margins to move with changes in production costs or income.

In sum, Marston's model suggests that the yen and a set of other variables influence the relative markup measured by the ratio of the export price index to the price index for Japanese goods sold in Japan. We will now use the model to evaluate empirically the impact of exchange rate changes on these margins across Japan's four major exporting industries.

The Long-Run Response of Profit Margins to Changes in the Yen

Marston's model suggests the following empirical specification:

$$p_{xt} - p_{ht} = \beta_0 + \beta_1(s_t + p^*_t - p_t) + \beta_2(c_t - p_t) + \beta_3 y_t + \beta_4 y^*_t + u_t, \tag{2}$$

where p_{xt} is the yen price of exports, p_{ht} is the price of Japanese goods sold in Japan, s_t is the exchange rate, p_t is the overall wholesale price index, c_t is production costs, and y_t is income. An asterisk designates a foreign variable. All variables are in log levels. The coefficients on the real exchange rate, real production costs, and real output measures are dictated by the demand characteristics faced by each industry in both foreign and domestic markets. The only sign implied by the model is that the coefficient on the real exchange rate, β_1, is negative.

Dynamic ordinary least squares regressions are a statistically efficient method for estimating the long-run response of relative export prices to each of these variables (Box B). Table 2 shows that the real yen index has a significant impact on relative profit margins. In particular, the estimates for β_1 are around -0.4 for industrial machinery, electrical machinery, and transportation equipment, three industries that together make up 70 percent of Japanese exports. This means that a 10 percent appreciation of the yen, all else being kept constant, decreases yen export prices by 4 percent relative to prices charged by Japanese firms to their domestic customers over the long run.[7] The fourth industry, precision equipment, has profit margins that respond much more modestly to the yen, with relative export prices falling only 2 percent for every 10 percent

Table 2
DYNAMIC ORDINARY LEAST SQUARES REGRESSION RESULTS

	Industrial Machinery	Electrical Machinery	Transportation Equipment	Precision Equipment
Constant	3.11	2.08	-0.06	5.93
	(75.1)	(9.6)	(1.4)	(10.4)
Real exchange rate	-0.38	-0.37	-0.46	-0.21
	(37.2)	(11.5)	(22.9)	(6.9)
Japanese industrial production	-0.46	0.37	—	-0.76
	(23.6)	(8.9)		(9.3)
Foreign industrial production	0.19	-0.37	0.50	—
	(11.8)	(4.9)	(22.7)	
Real unit labor costs	—	—	—	-0.29
				(4.1)
Real input costs	—	0.84	-0.71	—
		(7.7)	(4.6)	
Adjusted R^2	.99	.99	.95	.93
Sum of squared errors	.01	.04	.03	.14
ADF statistic	-5.5	-4.2	-5.7	-4.7
Error-correction coefficient	-0.11	-0.13	-0.19	-0.10
	(2.9)	(3.5)	(4.4)	(3.6)

Source: Author's calculations.

Notes: The sample period extends from January 1981 to June 1997. Variables are in log-level form. The regressions follow the dynamic ordinary least squares method as described in the article. The t-statistics in parentheses are based on Newey-West adjusted standard errors.

The Campbell and Perron method is used to select lags for the ADF statistic: The regression is initially run with twelve lags; if the twelfth lag is insignificant, then the lag is deleted and the regression is run again. This process is repeated until the last lag is significant at the 5 percent level. The critical values for the ADF depend on the number of variables in the regression and the sample size. For all regressions but electrical machinery, the critical values are -3.8 (5 percent) and -4.4 (1 percent). Because the electrical machinery regression has one more variable, the critical values are -4.2 (5 percent) and -4.7 (1 percent). The data are described in the appendix.

BOX B: ESTIMATING LONG-RUN AND SHORT-RUN PRICING-TO-MARKET BEHAVIOR

Japanese pricing-to-market behavior is estimated in two stages. The first stage calculates the long-run relationship between the variables in equation 2 with a dynamic ordinary least squares (DOLS) regression. The second stage measures the short-run behavior of relative export prices using an error-correction (EC) regression. The two regressions are connected since the difference between actual values and fitted values in the DOLS regression is used in the EC regression to predict relative export price movements.

The DOLS approach, as used by Stock and Watson (1993), modifies basic ordinary least squares estimation techniques by including both leads and lags of the first difference of all explanatory variables. These additional regressors are necessary because estimates in a single equation model can be biased by endogeneity among the variables. (In our regressions, four leads and lags were used, with the longest leads and lags eliminated if they were statistically insignificant. If the fourth lead or lag was significant, then additional leads or lags were added. The coefficients on the first difference variables are of no economic interest and are therefore not listed in Table 2.)

Certain conditions must be satisfied when conducting a DOLS regression:

- The level data of all variables must be nonstationary, while the first differences of the variables must be stationary. Essentially, a variable is stationary if its unconditional expected value and standard error do not change over time. The data used here satisfy this requirement.
- The DOLS regression results must yield stationary residuals. The test statistic is the augmented Dickey-Fuller test, listed as ADF in Table 2.
- The DOLS residual must be significantly correlated with subsequent monthly changes in relative export prices.

Short-run behavior is evaluated using an EC regression. In this regression, all the variables are converted into first differences, which, since the variables are in logs, is similar to converting the data into percentage changes. The change in relative export prices is then regressed against the lag of the residual from the DOLS regression, u_t, lags of changes in the explanatory variables, and lags of changes in relative export prices.[a]

$$\Delta(p_{xt} - p_{ht}) = a_1 + b_1 u_{t-1} \sum_{i=0}^{n} c_i \Delta(s_{t-i} + p_{t-i} - p^*_{t-i})$$
$$+ \sum_{i=1}^{n} f_i(\Delta(c_{t-i} - p_{t-i})) + \sum_{i=1}^{n} g_i \Delta y_{t-i} + \sum_{i=1}^{n} h_i \Delta y^*_{t-i}$$
$$+ \sum_{i=1}^{n} j_i \Delta(p_{xt-i} - p_{ht-i}) + e_i \ .$$

The residual is included since it should help predict how relative export prices will change in subsequent months. That is, any divergence of the relative export price index from its long-run value should tend to disappear. A negative and statistically significant residual is evidence that the DOLS coefficients do indeed represent a long-run relationship between the variables. The estimates of b_1, listed as error-correction coefficients in Table 2, are all negative and statistically significant, ranging from –0.10 for precision equipment to –0.19 for transportation equipment.

The EC regression estimates will be used to illustrate how relative export prices respond in the short run to a change in the yen's value.

[a] The contemporaneous value of the real yen is included to capture the automatic change in relative export price from Japanese exporters invoicing their goods in foreign currency terms.

rise in the yen. Apparently, this industry has foreign customers that are relatively insensitive to price changes.

The model discussed above makes no restrictions on the sign or the magnitude of the other coefficients since these estimates reflect reduced-form representations of differences in demand characteristics between markets. Of note is the fact that the electrical machinery regression calculates a sign on its foreign industrial production, domestic income, and input costs that is the opposite of the signs estimated for the other industries.

With these results, it is now possible to address issues about how Japanese profit margins respond in the short run to the yen, whether the response depends on the direction of the yen's movement, and whether markup behavior changed during the Asian currency crisis.

The Short-Run Response of Profit Margins to Changes in the Yen

It is likely that the short-run response of relative profit margins to changes in the yen differs from the long-run responses in Table 2. For example, if firms list the prices of their exports in dollar terms, one might see export prices, in yen terms, overshoot their long-run value. Alternatively, export prices might be slow to respond if firms contract out their prices in yen terms.

Chart 2 lists the response of relative export prices to a 10 percent rise in the real yen. (See Box C for details on how these values were calculated.) Essentially, the chart plots the month-to-month change following a rise in the yen. The reactions of profit margins for firms producing electrical machinery and transportation equipment are similar. For transportation equipment, there is an immediate 7 percent drop in export prices relative to prices charged to Japanese customers. Profit margins on exports are consequently squeezed sharply in the first few months, while foreign customers see relatively modest increases in the price of Japanese goods. Firms in this industry, though, do not view a drop in markups of this magnitude as being in their best interest. They pull export prices up shortly thereafter so that within twelve months the decline in relative export prices is near the long-run response of roughly 4.5 percent (represented by the horizontal lines in the chart). The overshooting of yen export prices suggests that a substantial portion of Japanese exports in both the electrical machinery and transportation equipment industries are invoiced in foreign currency terms, making the initial reaction of export prices to the yen more a passive response than a strategic choice by firms to reduce profit margins.

Exporters of industrial machinery and precision equipment are much less prone to overreact to a yen appreciation, implying that they are less likely to set contract prices in dollar terms. Firms producing industrial machinery adjust their margins to their long-run levels almost immediately, while those exporting precision equipment shift their margins to near their desired long-run levels within four months.

Symmetry

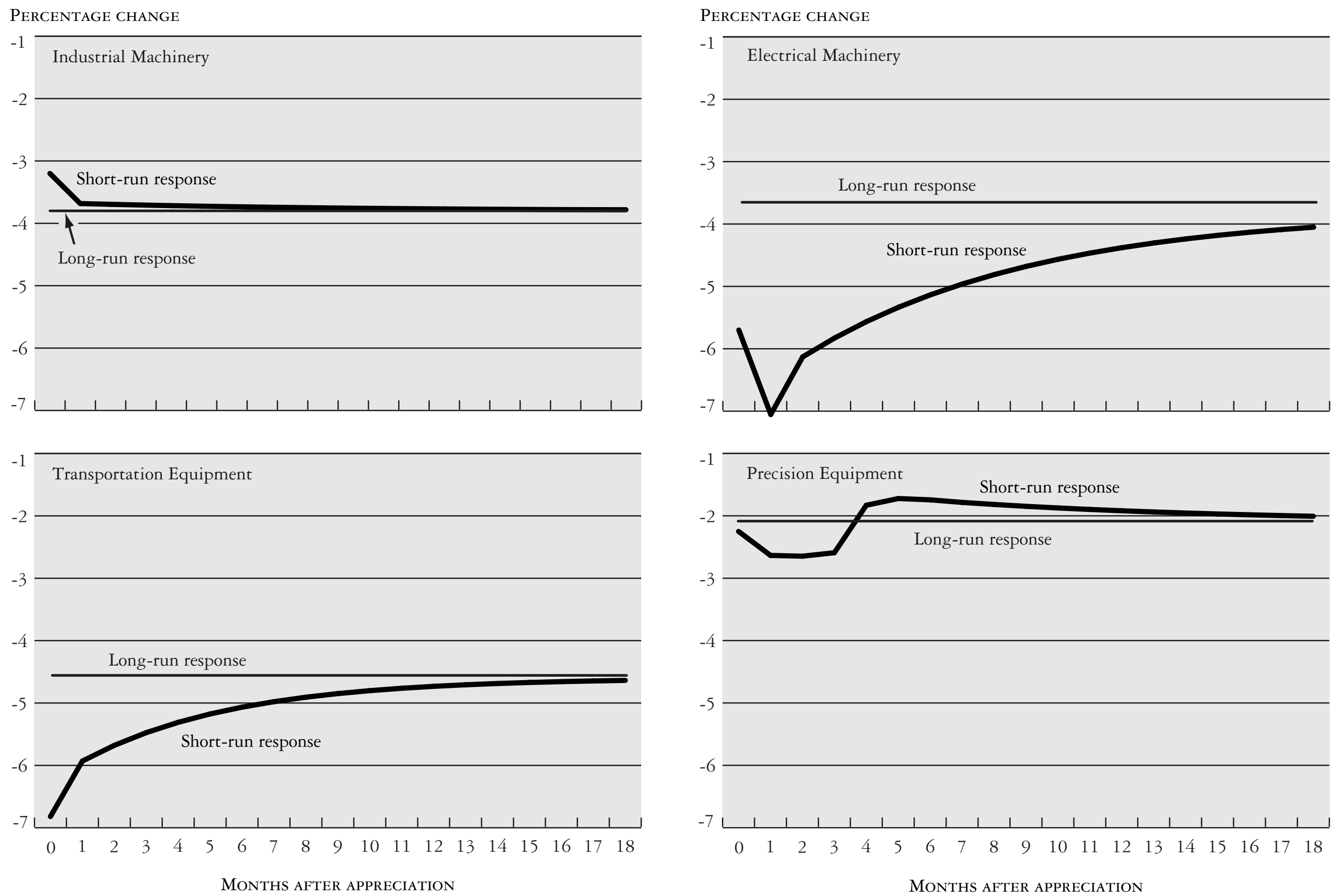
When Japanese firms cut the yen price of their exports to offset a yen appreciation, they are sacrificing profit margins to protect sales. However, when the yen depreciates, do these firms raise the yen price of their exports to the same extent in order to build their margins, or do they view a weak yen as an opportunity to gain market share by forgoing profits?

Since the beginning of the floating rate period, there have been only two episodes in which the yen experienced a prolonged depreciation: November 1988 to April 1990 and May 1995 to April 1997. To address the issue of how firms respond to a yen depreciation, we created an event dummy variable with a value of 1 for these two episodes and a value of 0 for all other months. This dummy variable was then multiplied by the exchange rate series to create an additional variable for the regression. The statistical significance of the dummy exchange rate variable is a simple test of whether the exchange rate coefficient depends on the direction of the yen's movements. The extent of any change is measured by adding the two exchange rate coefficients.

In all four regressions, the responses of relative export prices to the yen were found to be the same regardless of the yen's direction, with the size of the coefficients on the dummy variables being too small to change the values in Table 2. Japanese exporters of electrical machinery, industrial machinery, transportation equipment, and precision equipment respond in the same fashion regardless of the yen's direction. When the yen falls, the firms allow the foreign currency price of their exports to fall. The drop in prices seen by foreign customers is not as great as the yen's slide because Japanese firms strive to keep prices near other competing market prices. As a consequence, a falling yen helps profit margins to the same extent as a rising yen hurts them.

Chart 2

SHORT-RUN RESPONSE OF RELATIVE EXPORT PRICES TO A 10 PERCENT YEN APPRECIATION



Source: Author's calculations.

Note: Calculations are explained in Box C.

THE ASIAN CURRENCY CRISIS

The recent currency crisis in Asia created considerable uncertainty for Japanese exporters. The crisis started in early 1997 when the Thai government spent heavily to prop up its financial sector. Efforts to defend the currency against speculators failed, and the bhat was eventually allowed to float in early July. Pressures on exchange rates and government reserve holdings subsequently spread throughout the region, leading Indonesia, Malaysia, South Korea, and Thailand to suffer major devaluations while Singapore and Taiwan experienced significant but relatively modest declines.

The crisis complicated decision making for Japanese firms. In particular, exporters faced increased uncertainty over the level at which these currencies would eventually settle and the extent of their Asian customers' decline in demand for Japanese goods. In addition, Japanese firms faced greater uncertainty about the pricing and availability of imported components and materials from their Asian suppliers.

Because the Asian crisis occurred at the end of our sample period, it is possible to investigate the change in markup behavior by looking at the out-of-sample performance of the regressions estimated above. A successful forecasting of relative export prices would provide evidence that the markup behavior of Japanese firms was unaffected by the turmoil in Asia. A poor forecasting performance would

BOX C: SIMULATING SHORT-RUN DYNAMICS

The dynamic ordinary least squares (DOLS) regression results in Table 2 represent estimates of the long-run relationship between the export-weighted real yen index and the index of export prices relative to the index for domestic wholesale prices. The error-correction (EC) equation specified in Box B estimates how month-to-month changes in relative export prices are related to all variables considered and to the difference between the current level of relative export prices and the value suggested by the coefficients in Table 2. A combination of the DOLS and EC regressions can be used to illustrate how a permanent change in the real yen works to affect relative export prices over time. Specifically, the DOLS regression defines u so that the EC regression can be used to track monthly changes. (The coefficients from the EC regression appear in the table.)

A spreadsheet can show how relative export prices would react to a permanent 10 percent increase in the real export-weighted yen index. All other variables are assumed to be unchanged in this exercise. In the first month, the yen affects the change in relative export prices through the contemporaneous yen coefficient in the EC regression. In the second month, the yen works through the lagged yen variable (if it was statistically significant) and through the lagged residual value from the DOLS regression. This value is nonzero if the change in relative export prices in the first month did not push it to its long-run value dictated by the DOLS regression. The process continues, with the change in relative export prices in each month added to the change in the previous month to arrive at the results in Chart 2.

ERROR-CORRECTION REGRESSION COEFFICIENTS

	Industrial Equipment	Electrical Machinery	Transportation Equipment	Precision Equipment
Yen	-0.32 (25.8)	-0.57 (25.4)	-0.68 (31.7)	-0.22 (7.6)
Yen (-1)	-0.04 (3.4)	-0.16 (6.5)		
Yen (-2)		0.05 (2.0)	0.05 (2.1)	
Yen (-4)				0.07 (2.8)
Relative export price (-1)				0.18 (2.7)
Real input prices (-3)		-0.30 (2.0)		
Real unit labor cost (-1)				0.03 (2.8)
Japanese industrial production (-3)	-.05 (2.5)			
EC (-1)	-0.11 (3.6)	-0.13 (3.5)	-0.19 (4.4)	-0.10 (3.9)
R^2	.81	.81	.84	.29
Sum of squared errors	.003	.012	.012	.019

Source: Author's calculations.

Notes: The sample period extends from January 1981 to June 1997. The t-statistics are based on Newey-West adjusted standard errors. The regressions were originally run as specified in Box C and then all statistically insignificant variables were dropped. The EC variable is the time series of residuals from the regression results in Table 2. The number in parentheses next to a variable name represents the lag of the variable.

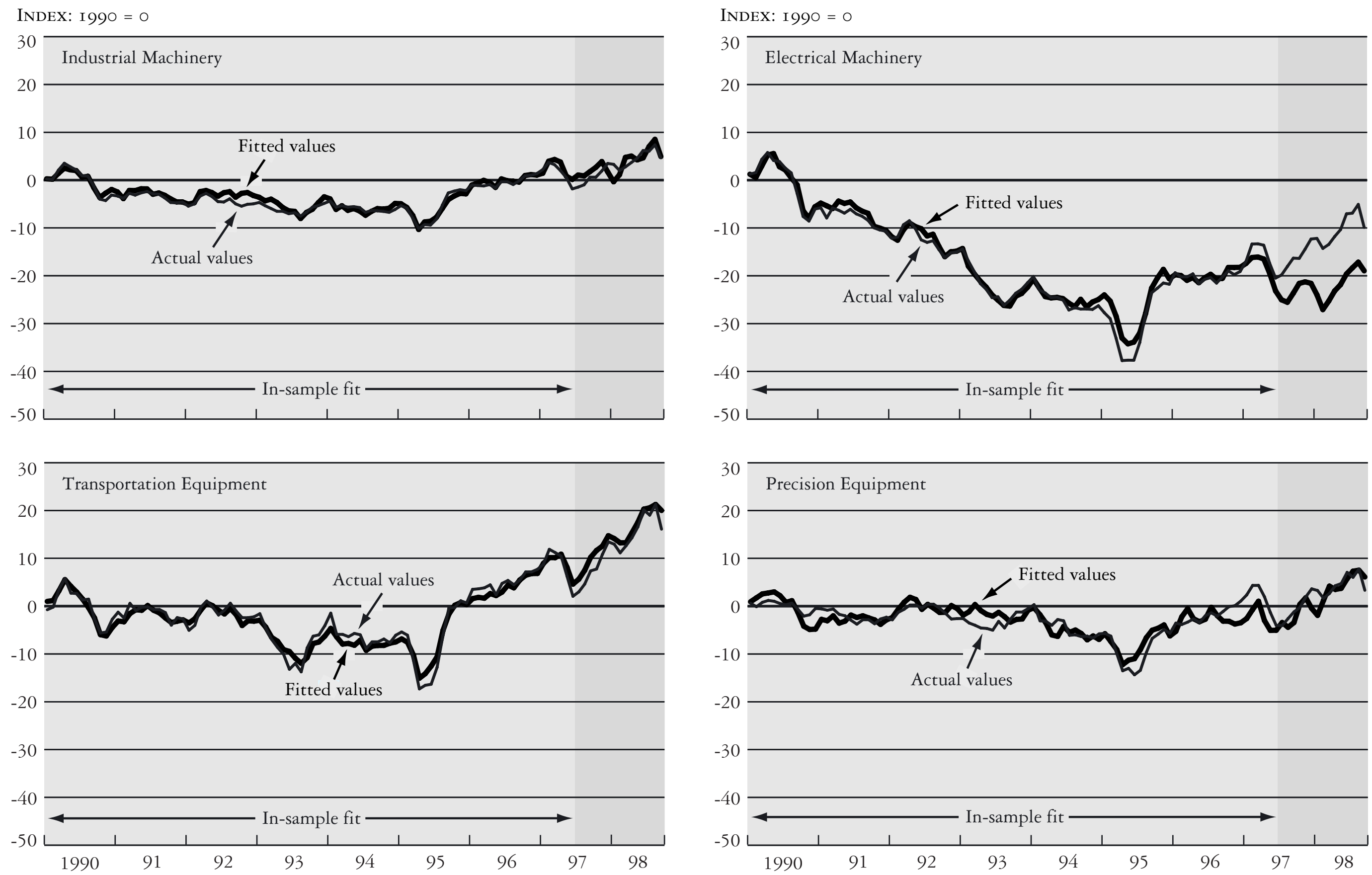
suggest that markup behavior changed in the aftermath of the crisis. Chart 3 plots the percentage difference between the prices Japanese firms charged their foreign and domestic customers, with the difference set to zero in 1990. (A decline reflects the falling of export prices relative to domestic Japanese prices.) For industrial machinery, transportation equipment, and precision equipment, the behavior of relative export prices appears to have been unaffected by the Asian currency crisis. All three industries raised their export prices relative to domestic prices after June 1996. In doing so, they took advantage, in a predictable way, of the yen's weakness against the dollar and European currencies to boost their profit margins on exported goods relative to goods sold in Japan.

For electrical machinery, however, relative export prices quickly went off track, with the gap widening considerably during the height of the crisis at the end of 1997. Specifically, Japanese firms in this industry raised relative export prices—an action that boosted margins on exported goods—while our regression predicted no such increase. The gap remained essentially unchanged during the first nine months of 1998.

Chart 3

Behavior of Relative Export Prices since the Beginning of the Asian Crisis

Percentage Gap between Export and Domestic Price Indexes



Sources: Bank of Japan; author's calculations.

Notes: Data are through September 1998. Fitted values are based on regressions in Table 2. Fitted values in shaded areas are out-of-sample forecasts starting in July 1997. A decline represents a fall in export prices relative to Japanese domestic prices.

The source of our regression's failure to predict an increase can be found by again using dummies to look for changes in coefficient estimates during the crisis period. We ran our regression through September 1998 with four additional variables that were calculated by multiplying each variable by a dummy. The dummy is defined as equal to 0 before July 1997 and equal to 1 for the rest of the sample. The results show that the uncertainty faced by exporters did not change the specific response of relative export prices to the yen. Instead, the coefficients on the other three variables all moved to explain why relative export prices increased during the crisis period. Such instability suggests that firms were uncertain about the demand characteristics of their foreign and domestic customers. Nevertheless, this uncertainty did not appear to have caused Japanese firms to alter their willingness to absorb exchange rate swings into their profit margins.

Conclusion

Japanese firms adjust the yen prices of their exports when the yen's value changes, a strategy that makes profit margins an important channel through which exchange rates affect Japan's economy. We find that in three of the four industries examined, the firms aggressively shield their foreign customers from price swings by allowing the profit margins on exports to fall 4 percent (relative to margins on goods sold in Japan) for every 10 percent appreciation of the yen.

Our findings also reveal that in Japan's electrical machinery and transportation equipment industries, the short-run responses of profit margins to changes in the yen are significantly greater than the long-run responses. This behavior is likely due to the fact that many exports in these two industries are denominated in foreign currency terms, making the change in export prices and profit margins an automatic, proportional response to changes in the yen. Moreover, the response of profit margins to changes in the yen is not found to depend on the direction of the yen's movements. Firms are as aggressive at raising export prices and building up profit margins after a favorable yen shift as they are at reducing profit margins after the yen moves against them.

Finally, the instability observed in pricing behavior in the wake of the Asian currency crisis seems to be limited to Japan's electrical machinery industry. Our forecasts predicted that firms in this industry should not have raised profit margins on exports relative to domestic sales as much as they did. Our findings suggest that firms in this industry have not changed the way in which they adjust profit margins in response to yen movements. Rather, the recent instability in markup behavior stems from changes in how exporters respond to other relevant variables.

Appendix: Data Sources

Japanese price data, at the wholesale level, are available for Japanese exports and for goods made by Japanese firms that are sold in the Japanese domestic market. The price indexes, published by the Bank of Japan, are Laspeyres, with weights altered every five years. These price data are significantly better than unit value export price indexes, which are simply the value of exports divided by the number of items shipped with no adjustments made for changes in the quality and composition of exports over time. The Bank of Japan is also the source of price indexes for inputs purchased by firms in each industry. The real unit labor cost measure is calculated using indexes for wages and productivity for each industry available in the *Monthly Statistics of Japan*. These indexes are seasonally adjusted using the X-11 (multiplicative) command in the EViews 3 software package.

Foreign variables are indexes constructed using data from the United States, Canada, Germany, the United Kingdom, France, the Netherlands, Korea, Taiwan, Hong Kong, and Singapore. The weights are based on the share of Japanese exports to each country, by each industry, in 1990. The data on industrial production, wholesale prices, and exchange rates are gathered from the International Monetary Fund, Data Resources International, Datastream, the Federal Reserve Bank of New York, and country sources.

Industrial production data from the four Asian countries are seasonally adjusted using the X-11 (multiplicative) command in EViews 3.

ENDNOTES

1. Examples of pass-through studies on Japan include Athukorala and Menon (1994), Loopesko and Johnson (1988), Marston (1990), and Tange (1997).

2. The pricing-to-market concept goes back at least as far as Krugman (1987). Pricing-to-market studies on Japanese exporters include Marston (1990), Ohno (1991), Khosla (1991), and Gagnon and Knetter (1995).

3. The ratio of the two indexes cannot tell you how the profit margin for exports compares with the margin for goods sold in Japan, since by construction both indexes are equal in the base year.

4. The four industries made up approximately 75 percent of Japanese exports in 1997. Industrial machinery and electrical machinery were each 24 percent, transportation equipment was 22 percent, and precision equipment was 5 percent.

5. Marston's model also has markups influenced by the derivative of marginal cost with respect to output.

6. The special case is when foreign customers do not respond at all to prices.

7. This result is similar to those found in other empirical studies. A survey of the literature in this field done by Goldberg and Knetter (1997) reported that measurements of pricing to market tend to be around 50 percent. They cite, as an example, the results reported by Marston (1990).

References

Athukorala, P., and J. Menon. 1994. "Pricing to Market Behavior and Exchange Rate Pass-Through in Japanese Exports." Economic Journal 104: 271-81.

Campbell, J., and P. Perron. 1991. "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots." In O. Blanchard and S. Fisher, eds., NBER Macroeconomics Annual 1991: 141-200.

Dickey, D. A., and W. A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Series with a Unit Root." Journal of the American Statistical Association 74: 427-31.

Gagnon, J. E., and M. Knetter. 1995. "Markup Adjustment and Exchange Rate Fluctuations: Evidence from Panel Data on Automobile Exports." Journal of International Money and Finance 14, no. 2: 289-310.

Goldberg, P., and M. Knetter. 1997. "Goods Prices and Exchange Rates: What Have We Learned?" Journal of Economic Literature, September: 1243-72.

Khosla, A. 1991. "Exchange Rate Pass-Through and Export Pricing: Evidence from the Japanese Economy." Journal of the Japanese and International Economies 5: 41-59.

Krugman, P. 1987. "Pricing-to-Market When the Exchange Rate Changes." In S. W. Arndt and J. D. Richardson, eds., Real Financial Linkages among Open Economies. Cambridge: MIT Press.

Loopesko, B., and R. A. Johnson. 1988. "Realignment of the Yen-Dollar Exchange Rate: Aspects of the Adjustment Process in Japan." In R. C. Marston, ed., Misalignment of Exchange Rates: Effects of Trade and Industry, 105-44. Chicago: University of Chicago Press.

Marston, R. C. 1990. "Pricing to Market in Japanese Manufacturing." Journal of International Economics 29: 217-36.

Ohno, K. 1991. "Export Pricing Behavior of Manufacturing: A U.S.-Japan Comparison." IMF Staff Papers 36, no. 3: 551-79.

Stock, J., and M. Watson. 1993. "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems." Econometrica 61, no. 4: 783-820.

Tange, T. 1997. "Exchange Rates and Export Prices of Japanese Manufacturing." Journal of Policy Modeling 19, no. 2: 195-206.

The Federal Reserve Bank of New York
Economic Policy Review

Single copy subscriptions to the Economic Policy Review (ISSN 0147-6580) are free. Multiple copies are available for an annual cost of $12 for each additional subscription. Checks should be made payable in U.S. dollars to the Federal Reserve Bank of New York and sent to the Public Information Department, 33 Liberty Street, New York, NY 10045-0001 (212-720-6134). Single and multiple copies for U.S. subscribers are sent via third- and fourth-class mail.

Economic Policy Review subscribers also receive the Bank's Annual Report.

Library of Congress Card Number: 77-646559

www.ny.frb.org/rmaghome