EMPIRICAL ASSET PRICING AND INVESTMENT STRATEGIES

Krister Ahlersten

STOCKHOLM · SCHOOL OF ECONOMICS

Empirical Asset Pricing
and
Investment Strategies

Krister Ahlersten

AKADEMISK AVHANDLING

Som för avläggande av ekonomie doktorsexamen
vid Handelshögskolan i Stockholm
framläggs för offentlig granskning
onsdagen den 26 september 2007, klockan 14.15
i sal Torsten, Handelshögskolan, Sveavägen 65

EMPIRICAL ASSET PRICING
AND
INVESTMENT STRATEGIES

STOCKHOLM SCHOOL OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute

EFI Mission

EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution that works independently of economic, political and sectional interests. It conducts theoretical and empirical research in the management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics. At EFI, the researchers select their projects based on the need for theoretical or practical development of a research domain, on their methodological interests, and on the generality of a problem.

Research Organization

The research activities at the Institute are organized into 21 Research Centres.
Centre Directors are professors at the Stockholm School of Economics.

EFI Research Centre:	*Centre Director:*
Management and Organisation (A)	Sven-Erik Sjöstrand
Centre for Entrepreneurship and Business Creation (E)	Carin Holmquist
Public Management (F)	Nils Brunsson
Information Management (I)	Mats Lundeberg
Centre for People and Organization (PMO)	Andreas Werr (acting)
Centre for Innovation and Operations Management (T)	Christer Karlsson
Centre for Media and Economic Psychology (P)	Guje Sevón
Centre for Consumer Marketing (CCM)	Magnus Söderlund
Centre for Information and Communication Research (CIC)	Per Andersson (acting)
Marketing, Distribution and Industrial Dynamics (D)	Björn Axelsson
Centre for Strategy and Competitiveness (CSC)	Örjan Sölvell
Centre for Business and Economic History (BEH)	Håkan Lindgren
Accounting and Managerial Finance (B)	Johnny Lind
Centre for Financial Analysis and Managerial Economics in Accounting (BFAC)	Kenth Skogsvik
Finance (FI)	Clas Bergström
Centre for Health Economics (CHE)	Bengt Jönsson
International Economics and Geography (IEG)	Mats Lundahl
Economics (S)	Paul Segerstrom
Economic Statistics (ES)	Anders Westlund
Law (RV)	Johnny Herre
Centre for Tax Law (SR)	Bertil Wiman

Chair of the Board: Professor Carin Holmquist
Director: Associate Professor Filip Wijkström

Address

EFI, Box 6501, SE-113 83 Stockholm, Sweden • Website: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

EMPIRICAL ASSET PRICING
AND
INVESTMENT STRATEGIES

Krister Ahlersten

STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics 2007

ISBN NR 978-91-7258-726-7

KEYWORDS: Asset Pricing, Decoupling, Equity Premium, Investment Strategies, Portfolio Choice, Portfolio Optimization, Predictability, Regime-Switching Models, Smooth Transition Regressions, Structural Breaks, Time-Series Models

PRINTED BY:
Elanders Gotab, Stockholm 2007

DISTRIBUTED BY:
EFI, The Economic Research Institute
Stockholm School of Economics
P O Box 6501, SE-113 83 Stockholm, Sweden
www.hhs.se/efi

To Klara

Acknowledgments

A number of people have helped to bring this thesis to fruition. First and foremost, I thank my two supervisors, Peter Englund and Timo Teräsvirta, who have "supervised" me quite differently. Peter and I have engaged in innumerable and lengthy discussions on the workings of the economy and what effects it might have on the price of financial assets. Fortunately, I overcame my intellectual hubris and learned a great deal from someone whose knowledge of this subject is vastly superior to mine. Timo and I met punctually once a week to discuss the problems and obstacles I had encountered since our last meeting. Each time, I was impressed by his breadth of knowledge spanning the arts and sciences, and his extraordinarily deep knowledge of econometrics. The combination was appreciated, to say the least.

Martin Dufwenberg was my supervisor at Stockholm University. Martin's enthusiasm for practical research was contagious, and very important to me during the first phase of my thesis work. Martin now lives in Arizona and is most likely considered indispensable to the research world there. Rumor also has it that he has abandoned his shorts and is now into blazers, but I find that hard to believe.

On a few occasions I have assisted Torbjörn Sällström on the portfolio course at IFL. Torbjörn is a brilliant teacher and has a great intuitive understanding of the subject. I have been very privileged to be able to learn from him. I am also grateful to Dick van Dijk, who generously gave me several program files that I have used as the basis for my own program routines. They saved me a lot of time and also steepened my learning curve for programming.

Throughout the Ph.D. program, I have benefited from countless discussions with colleagues and friends at the Stockholm School of Economics and Stockholm University. Together we have played, argued, and theorized. I am very grateful to Andrei Simonov, Birgit Strikholm, Bo Larsson, Clas Bergström, Håkan Blomdahl, Iñaki Rodríguez Longarela, Jesper Roine, Johan Söderström, Marcus Salomonsson, Mikael Elhouar, Otto Elmgart, Rickard Sandberg, Roland Nilsson, and Sven Skallsjö.

I have also been fortunate to have spent some time at the Swedish Central Bank in Stockholm, and at the European Central Bank in Frankfurt. At the Swedish Central Bank, Hans Dillén and Peter Sellin gave generously of their time and experience; in Frankfurt, I was lucky to have Manfred Kremer, Henrik Amilon and Magnus Andersson as workmates and playmates.

Marita Rosing and Anneli Sandbladh have smoothly and professionally resolved all the administrative problems, and Maria De Liseo has read the manuscript and improved the English considerably.

Finally, I am grateful to Bankforskningsinstitutet and the Ljungberg Foundation for providing the research funds for this work.

Stockholm, August 20, 2007
Krister Ahlersten

Contents

Introduction and Summary

This thesis examines a number of topics related to portfolio choice, asset pricing, and strategic and tactical asset allocation. The first two papers treat the predictability of asset returns. Since at least the mid-1980s until quite recently, the conventional wisdom has been that it is possible to predict the return on, for example, an index of stocks. However, a series of recent papers have challenged this conventional wisdom. I answer this challenge and show that it is possible to predict returns if structural changes in the underlying economy are taken into account. The third paper examines the comovement between stocks and bonds. I show how it is possible to improve the composition of a portfolio consisting of these two asset classes by taking into account how the comovement changes over time. All three papers are self-contained and can therefore be read in any order.

The first paper is entitled "Structural Breaks in Asset Return Predictability: Can They Be Explained?" Here I investigate whether predictability has changed over time and, if so, whether it is possible to tie the change to any underlying economic variables. Dividend yield and the short interest rate are often used jointly as instruments to predict the return on stocks, but several researchers present evidence that the relation has undergone a structural break. I use a model that extends the conventional structural breaks models to allow both for smooth transitions from one state to another (with a break as a special case), and for transitions that depend on a state variable other than time. The latter allows me to directly test whether, for example, the business cycle influences how the instruments predict returns. The results suggest that this is not the case. However, I do find evidence of a structural change primarily in how the instruments predict returns for large firms. The change differs from a break in that it appears to be an extended non-linear transition during the period 1993–1997. After the change, the short rate does not predict returns at all. Dividend yield, on the other hand, is strongly significant, and the return has become more sensitive to it.

In the second paper, "Restoring the Predictability of Equity Returns," I take another perspective on predictability and structural shifts. Several recent papers have questioned the predictability of equity returns, potentially implying serious negative consequences for investment decision-making. With return data including the 1990s, variables that previously predicted returns, such as the dividend yield, are no longer significant and results of out-of-sample tests are often weak. A possible reason is that the underlying structure of the economy has changed. I use an econometric model that allows for regime shifts over time as well as due to changes in a state variable, in this case the price-earnings ratio. This makes it possible to separate influences from these two sources and to determine whether one or both sources have affected return predictability. The results indicate that, first, a structural change occurred during the 1990s, and, second, that the unusually high level of price earnings in the late 1990s and early 2000s temporarily affected predictability at the 12-month horizon.

In the third paper, "Coupling and Decoupling: Changing Relations between Stock and Bond Market Returns," I investigate stock-bond comovement. The correlation between stocks and bonds has changed dramatically over the last ten years, introducing a new type of risk for portfolio managers, namely, correlation risk. I use GARCH estimates of stock volatility, simple regressions, and regime-switching econometric models to assess whether level of volatility, or changes in volatility, can be used to explain some of the changes in comovement in seven different countries. As regards volatility level, strong support is found in almost all countries to suggest that high volatility predicts lower, or negative, comovement. I argue that this can be evidence of a market-timing type of behavior. As for changes in volatility, the results are more mixed. Only for the U.S. market do I find strong support to conclude that large changes tend to coincide with lower, or negative, comovement. This could be evidence of a flight-to-quality (or cross-market hedging) type of behavior.

Structural Breaks in Asset Return Predictability: Can They Be Explained?

ABSTRACT. Dividend yield and the short interest rate are often used jointly as instruments to predict the return on stocks, but several researchers present evidence that the relation has undergone a structural break. I use a model that extends the conventional structural breaks models to allow both for smooth transitions from one state to another (with a break as a special case), and for transitions that depend on a state variable other than time.

The latter allows me to directly test whether, for example, the business cycle influences how the instruments predict returns. The results suggest that this is not the case. However, I do find evidence of a structural change primarily in how the instruments predict returns for large firms. The change differs from a break in that it appears to be an extended non-linear transition during the period 1993–1997. After the change, the short rate does not predict returns at all. Dividend yield, on the other hand, is strongly significant, and the return has become more sensitive to it.

1. Introduction

Conventional wisdom in the academic asset pricing literature has long held that stock returns are to some extent predictable. The literature dates back at least to Fama and French (1988), and Campbell and Shiller (1988), who showed that dividend yield is a significant predictor of the return on a stock index.[1] Campbell and Shiller also showed that there is a theoretical connection between dividend yield, return, and dividend growth–a connection that has convinced many people that returns not only happen to be predictable, but that they must be so.

In this context, predictability has become widely accepted, so much so that Cochrane (2001) writes about predictability as if it were a stylized fact. But the question has once again become controversial. The new debate is based primarily on two empirical results. The first is that several researchers (Bossaerts and Hillion (1999) and Goyal and Welch (2003 and 2007)) have shown that while there might be predictability in-sample, there seems to be none out-of-sample. The second result is that even in-sample predictability, in ordinary linear models, has shown a decreasing level of significance over the 1990s and

[1] Other important papers in this literature include Rozeff (1984), Shiller (1989), Cochrane (1992), Hodrick (1992), Lamont (1998), and Lettau & Ludvigson (2001).

early 2000s (Lettau and Ludvigson (2001), Ang and Bekaert (2007), and Lettau and van Nieuwerburgh (2006)).

By and large, one can take two different positions on these new findings. One might regard the earlier results, with significant predictions, as a statistical artefact, and that with more data the return showed its true face: that it is not predictable. But one might also believe that the new results are misleading, and that a deeper analysis will show that returns are still predictable. One line of research arguing for this latter case tries to demonstrate that the results can be explained by structural breaks. Pesaran and Timmermann (1995), Viceira (1996), Pastor and Stambaugh (2001), Paye and Timmermann (2006), and Lettau and van Nieuwerburgh (2006) show that there are good reasons to believe that the returns process has undergone one or several structural breaks. Their results differ slightly, depending on sampling frequency, method, and purpose, but several of them find a structural break somewhere around the early or mid-1990s.[2] Lettau and van Nieuwerburgh also show that if one "corrects" the data (i.e., adjusts the mean) for a structural break in 1991, then the dividend yield, for example, is still an in-sample significant predictor of returns.

While these papers have found evidence of instability and structural breaks, they do not attempt to explain the breaks, in terms of underlying variables, by means of an econometric model. Furthermore, most of them do not allow for other types of economic changes other than instantaneous breaks. However, it is likely that a structural break is closely connected to changes in the underlying economy, such as changes in the business cycle, and also that structural changes in the economy are not instantaneous. In this paper, I therefore show how to model and test whether or not a certain variable, for example inflation, explains a structural break, and I test this model for a string of different macroeconomic and financial variables. To study the robustness of the results, I do this not only on a broad index, but also on the ten decile portfolios sorted on size. Using this strategy, I show that none of the analyzed variables can explain the structural breaks. This, in turn, can be interpreted as indicating that the breaks occur because of unique, one-off, events and not because of recurring changes in underlying variables. In a way, this confirms the results in previous studies, but from quite a different starting point.

In the real economy, one would often think of a structural change as extending over a period of time. During that period, the economy is in transition from one state to another—a transition period that could take several years. The model I use therefore extends the set of possible outcomes regarding structural breaks to incorporate smooth shifts. This allows me to estimate the length of the transition period from the data, with an instantaneous break as a special case, and more accurately determine the timing of the event. My results for the deciles of small and large firms are systematically different. While there is ample evidence of at least one break for the deciles of large firms, there is little if any evidence for the deciles of small firms.

---

[2] For a univariate regression of dividend yield on returns, Paye and Timmerman find a break in Dec 1994, and Lettau and van Nieuwerburgh find one in 1991. In a Bayesian framework, Pastor and Stambaugh find a break in the estimated risk premium in about 1992 (read from Figure 1 in their paper).

2. Background and Literature

In the early financial literature, the idea that returns are predictable was often rejected, as predictability was believed to contradict the efficient markets hypothesis. Later research has shown, however, that this need not be the case. For example, a certain amount of predictability is consistent with a time-varying risk premium. Thus, to gain some intuition about why the risk premium would vary over time, it is useful to consider the Consumption Capital Asset Pricing Model (CCAPM). A popular version of the model states that[3]

$$\begin{aligned} E\left[R^i - R^f\right] &= \beta_{i,\Delta c}\lambda_{\Delta c}, & (2.1) \\ \lambda_{\Delta c} &= \gamma \cdot var\left(\Delta c\right), & (2.2) \end{aligned}$$

where E is the expected value operator, R^i is the return on financial asset i, R^f is the risk free rate of return, Δc is consumption growth, $\beta_{i,\Delta c}$ is the beta from a regression of R^i on Δc, and γ is the coefficient of relative risk aversion. The CCAPM is typically rejected in empirical studies, but it can be informative about where to look for explanatory variables, how to determine possible sources of time variation, and how to interpret variables in the context of predictability: If risk aversion, γ, or consumption risk, var(Δc), changes, the risk premium on an asset should, ceteris paribus, also change. Assets with a high beta should also yield a higher expected return, but since we are dealing with indices here, I will not discuss this property.[4] A standard way of interpreting a predictive variable in the CCAPM context, then, is that it is related either to risk aversion or consumption risk, or, with reference to the standard CAPM, to market risk. Through any of these channels, it is then also related to returns. Early approaches along these lines were the attempts to tie stock market variance to returns (French, Schwert and Stambaugh (1987), Poterba and Summers (1988), Nelson (1991), and Glosten, Jagannathan and Runkle (1993)). The results from this literature have, however, been quite inconsistent, and the links from the variance of both consumption and market risk to returns are generally considered to be weak. The empirical predictability literature has therefore largely left both the CCAPM strategies and the variance-to-return strategies (Cornell (1999)).[5]

The most influential papers in the early predictability literature are probably Fama and French (1988), and Campbell and Shiller (1988). Fama and French show that the dividend yield is a significant predictor of returns of a stock index in an OLS regression, and also that the predictions become better and better, in terms of higher R^2's, the longer the return horizon is. Campbell and Shiller start with Gordon's (1962) growth model,

$$P_t = \frac{D_{t+1}}{r-g} \Longleftrightarrow \frac{D_{t+1}}{P_t} = r - g, \tag{2.3}$$

[3] This version assumes power utility and uses a first order Taylor expansion. See, e.g., Cochrane (2001) for a derivation.

[4] Note that the beta is not such an important component when examining the index, as we do not have different assets with different betas.

[5] A recent attempt to tie volatility directly to returns is that by Ghysels, Santa-Clara and Valkanov (2005).

expand it to allow r and g to vary over time, and then derive an expression for the dividend yield:

$$\log\left(\frac{D_{t-1}}{P_t}\right) \approx E_t\left[\sum_{j=0}^{\infty}\rho^j\left(r_{t+j} - \log\left(\frac{D_{t+j}}{D_{t+j-1}}\right)\right)\right] + constant. \tag{2.4}$$

Here D_t is the dividend at time t, P is the price, r is the discount rate, and ρ is a constant. In (2.3) the discount rate, r, and the dividend growth, g (that corresponds to $\log\left(\frac{D_{t+j}}{D_{t+j-1}}\right)$ in (2.4)), are constant, so the dividend yield must also be constant. The expression (2.4), on the other hand, allows for time variation in the dividend yield and states that it is the "expected discounted value of all future one-period 'growth-adjusted discount rates'".[6] This means that the dividend yield *must* predict either returns or growth, possibly both. Campbell and Shiller also show empirically that dividend yield predicts returns but not growth. Variation in the dividend yield must then, more or less, completely spill over to returns.

For a long while, these results have been considered very convincing, and Cochrane (2001) writes about asset return predictability as if it were a stylized fact (see also Cochrane (2006)). Several researchers have expanded the literature and found a number of variables, both financial and macroeconomic, that are able to predict returns. Apart from the dividend yield, the short interest rate, earnings, the term premium, the default premium and different measures of market risk belong to the most popular and thoroughly analyzed variables. Furthermore, variables such as inflation, industrial production, and monetary growth are also frequently used. However, the results have been criticized for being unstable over time, and for being the product of data mining.

In the late 1990s, the predictability literature began to attract criticism. Several researchers argue that while evidence might exist of in-sample predictability, as shown by Campbell and Shiller and Fama and French, this does not mean that a real life practitioner would actually be able to predict returns. The arguments center around out-of-sample tests. Bossaerts and Hillion (1999) start with a set of linear models with different combinations of predictive variables. Then they select models based on different statistical criteria, such as the adjusted R^2, in order to find the "best" models. This strategy selects some models with in-sample predictability, but none of the "best" models have any out-of-sample predictive ability. So an investor who chooses the "best" model will not be able to use it successfully. Goyal and Welch (2003 and 2007) could be said to work in a similar spirit. They compare out-of-sample predictions from linear models using several different predictive variables, and show that almost none of these do better than the historical mean of asset returns.[7] Furthermore, they show that the predictability in previous papers, such as Fama and French (1988), depends mostly on extreme events during 1973 and 1974.

[6] See function (6) in their paper. The notation has been changed for consistency.

[7] For a contrary view on the out-of-sample evidence of Goyal and Welch's, see Campbell and Thompson (2005). They argue that if one imposes some reasonable restrictions on the predictions, such as nonnegativity, the out-of-sample results are in favor of predictability.

Another problem for the predictability literature, perhaps an even more serious one than the out-of-sample problems, is that the degree of significance of the coefficient estimates for variables such as the dividend yield has been decreasing since the works of Fama and French, and Campbell and Shiller, to the extent that they no longer predict returns (Lettau and van Nieuwerburgh (2006), Ang and Bekaert (2007), and also Goyal and Welch (2007)). This could mean that the two seminal papers of Fama and French, and Campbell and Shiller depend on a statistical accident.

Nevertheless, it is far from self-evident that insignificant results from in-sample or out-of-sample linear regressions are proof of no predictability. As shown in, for instance, Stock and Watson (1996), macroeconomic time series and interest rates often experience structural breaks. If the returns process has experienced one or several structural breaks, this is going to affect the significance levels of predictive regressions, both in-sample and out-of-sample, provided that both pre-break and post-break data are used. Furthermore, if a break in the returns process occurred during the 1990s, then the results from the critics of the predictability literature are precisely what one would expect, even if returns are actually predictable: a drop in in-sample significance levels in the 1990s and weak out-of-sample predictability, as the out-of-sample period typically chosen is the late 1990s and early 2000s.

Over the last few years, several researchers have modeled the returns process with structural breaks. Pastor and Stambaugh (2001) use a Bayesian model where the agents update their beliefs about the level of the risk premium given historical returns, and the possibility that a structural change has occurred. In addition, they make several different assumptions about risk aversion, volatility and transition periods, and estimate the corresponding implied time series of the "Bayesian" risk premium for the period 1887–1998. The results change considerably depending on which assumptions the authors make, but they find evidence of a break in about 1992.

Paye and Timmermann (2006) use a linear model with several explanatory variables on ten different national stock markets, and show that the processes have undergone Can They Be Explained most of the variables and on most of the markets. The breaks, however, are localized to different points in time for different variables and different markets. They find, for example, a break in 1974/75 on the U.S. market, which they speculate might be related to the oil crisis. Lettau and van Nieuwerburgh (2006) find a break in the relation between dividend yield and returns in 1991. They also show that if one "corrects" for that break, i.e., if one adds a constant to the dividend yield after the break so that the mean is the same as before the break, then the variable is still a significant predictor of returns in an in-sample OLS regression, and the R^2 increases from 0.024 to 0.086.[8] Viceira (1996) also finds a structural break in the returns process, but he does not find a break in how the dividend yield predicts returns at the one-month horizon. This might be because his data series ends in 1995, which is very close to the date when others have found a break to occur. Thus, his test might lack power due to an insufficient number of observations after the break.

[8] This is for a return horizon of one year.

These last few papers presenting evidence of structural breaks represent a large leap ahead in the predictability literature. But as already mentioned, they do not attempt to tie down the breaks with any underlying variables in an econometric model, and they only model instantaneous breaks, not transitions.[9] In the following section, I will discuss several reasons for structural changes in the returns process, and describe an econometric model that has the ability to overcome these problems.

3. Motivation and Modeling

As we have seen, several researchers have found structural breaks in the returns process. A structural break is a potential source of serious problems. If, for instance, data is used from both before and after the break in a linear model with constant parameters, the predictions will be biased and the t-values might give a misleading picture about the existence of predictability.

To say that the data-generating process has undergone a structural break is not very informative in itself. The challenges are to date the break in time, or tie it to a variable other than time, and to model the connections between explanatory and explained variables, given the break. How we model the break, in turn, depends on how we look upon different possible economic explanations of the break. There are many theoretical reasons why a structural break might occur. Specific macro events, such as institutional changes, changes in regulations, taxes or fees, and changes in monetary policy, or shocks to the economic system, such as oil price shocks, might cause structural breaks in the returns process. Other macroeconomic reasons could be that the economy changes over a period of time due, for example, to increased globalization, increased use of computers, and increased efficiency, or that people learn more about financial economics. There could also be more or less sudden changes in market risk, risk aversion, or market sentiments in addition, perhaps, to changes in other psychological aspects of financial behavior.

A structural break explained by t is probably what most people associate with the term "structural break", i.e., before and after a certain point in time the process is different. Some of the possible explanations mentioned above fit rather well into such a description. If, for instance, a new tax is implemented on a certain day, it could be reasonable to date a break to exactly that day. But many of the possible explanations do not fit the description of a break at all. For instance, globalization could hardly be described as a break. To obtain a reasonable dating of these types of explanations, the model must allow for the possibility that the "break" is not a sharp break, but a smooth shift from one state to another. And that the process, between the beginning and the end of the shift, is in transition. In the following, I will mostly use the term "structural shift" instead of "structural break" to emphasize that the model I use allows for smooth transitions, with a break as a special case.

Another central question is whether it is possible to tie a structural shift to any underlying variable. A structural shift explained by t is tied to time, but time in itself

[9] Pastor and Stambaugh account for a sort of transition periods in their Bayesian model. They, however, do not attempt to connect these to any underlying variable.

does not explain anything. However, there are theoretical reasons that link shifts to underlying variables. Campbell and Cochrane (1999) argue that the business cycle affects risk aversion. In their model, a representative agent has a time-varying habit consumption level (or subsistence level). As he gets closer to the habit level, which is what happens during a recession, he becomes more and more risk averse to avoid the danger of falling below it. As his risk aversion rises, he demands a higher premium to invest in risky assets, and because of this effect there is a nonlinear structure in the authors' model for returns. In the present context, we would then expect it to be possible to tie structural shifts in returns to the business cycle. In a similar vein, it is perhaps also likely that risk aversion could vary with factors such as market sentiment, behavioral effects, or price bubbles, to name just a few. Note that since the business cycle changes back and forth over time, structural shifts that are tied to it would, most likely, also shift back and forth over time, possibly many times. This makes the model here potentially different from traditional structural break models in which the number of breaks is often very low.

Furthermore, Perez-Quiros and Timmermann (2000) argue that small firms with less collateral should be more affected by tighter credit market conditions than large, better collateralized, firms. During a recession, one would then observe a sort of flight-to-quality, as the risk associated with small firms increases more than the risk associated with large firms, and investors demand higher compensation for bearing this extra risk. Consequently, the should exist both time-series effects (the business cycle affects the risk premium over time) and cross-sectional effects (small firms are more affected that large firms). The authors use data on returns on the ten size-sorted decile portfolios from the Center for Research in Securities Prices (CRSP) to study effects from size, and an extended version of the Hamilton (1989) Markov switching model to study effects over time. The model states that the economy, at a given time, is in one of two different states. Within each state, returns evolve as a linear combination of several risk factors. The state is not observed, but is inferred from the data, and one can estimate a time series of probabilities of being in a certain state, as well as the Markov probabilities of shifting from one state to the other.

The probabilities of shifting are, in the model, dependent on the value of a business cycle indicator. However, the corresponding parameters mostly turn out insignificant, so it is unclear whether a connection to the business cycle really exists. Small firms, on the other hand, often do get betas that are larger in absolute value than those for large firms, indicating that they indeed are more sensitive to the risk factors. The arguments and the results from Perez-Quiros and Timmermann motivate me to use the same ten decile portfolios and a similar business cycle indicator as they do. In contrast to their paper, the approach in this paper allows me to directly test if several different indicators, both macro and financial variables, have any explanatory value for structural shifts in returns. Moreover, doing this for all ten decile portfolios allows me both to detect the presence of any systematic size effects and to determine the timing of structural shifts.

These types of nonlinear effects, where a shift is associated with an underlying variable, could, in a way, be modeled as structural shifts in time. For example, assume that risk aversion actually depends on the business cycle, and that the returns process during the high-inflation economy of the 1970s and early 1980s was therefore different from that of the other periods. This could be modeled with two structural shifts: one in the mid-1970s and another one, in the opposite direction, in the beginning of the 1980s. But such a model would not take the explanatory value of inflation into account, and it would not produce correct predictions in the future if the high-inflation economy were to return. Not until well after this had occurred would such a model indicate that a third shift had taken place. Furthermore, the model would not motivate anti-inflationary policy work.

To incorporate this type of causality from underlying variables into the model, I expand the notion of structural shift to include not only shifts explained by t, but also shifts explained by some other state variable, z_t. A structural shift explained by the state variable is similar to that explained by t, but the process is different depending on whether the state variable obtains values higher or lower than a certain level. To continue the previous example, say that the returns process is different depending on whether inflation is "normal" or unusually high. Then we have a structural break in the returns process that is explained by the level of inflation. If, furthermore, the transition is not instantaneous but smooth, we have a structural shift. Since z_t explains the transition from one state to another, it is often referred to as a transition variable.

There is, however, one more complication. Suppose we find some sort of structural shift that can be explained by a transition variable, z_t. Assume, for instance, that the returns process experienced a structural shift in 1975 as a consequence of the stagflation economy. If the returns process did not shift back when the stagflation economy disappeared, the shift could be explained by both z_t and t. In a way, z_t has caused the shift, but it has then failed to cause another shift when it moves in the opposite direction. In that case, z_t is not relevant for explaining future returns, and since it is the returns process we are interested in here, I will treat such a case as a shift that is explained by t.[10] Consequently, it is important to control for a shift explained by t when we look for shifts explained by other variables. A candidate variable must explain a shift better that t does, or explain it even if we have another shift that is explained by t. In Section 4.3, I will show that the results we obtain with and without this assumption are very different.

In the next section, I will describe how we can model and test for structural shifts in any underlying variable, both controlling and not controlling for t. The model will also allow for a smooth shift from one state to another, and the speed, or duration, and timing of the shift will be estimated from the data.

3.1. The STR and TV-R Models. To make econometric sense of how to tackle the problem, I start with the following linear model:

$$R_{t+1} = \boldsymbol{\beta}'\mathbf{X}_t + \varepsilon_{t+1}, \tag{3.1}$$

where R is the variable to be predicted, and $\mathbf{X}$ is a set of variables, including a constant, believed to predict R. I will discuss the selection of variables in $\mathbf{X}$ later in Section 4.2. If the process has undergone a structural break explained by time, the model (3.1) can be

[10] It might be the case, of course, that the variable only cases shifts in one direction. e.g., that stagflation has an impact when it enters, but not when it leaves. In such a case, it is of paramount interest for policy work, but this possibility is not studied here.

written as follows:

$$R_{t+1} = \begin{cases} \boldsymbol{\beta}_1'\mathbf{X}_t + \varepsilon_{t+1}, & t \leq \tau \\ \boldsymbol{\beta}_2'\mathbf{X}_t + \varepsilon_{t+1}, & t > \tau \end{cases}, \tag{3.2}$$

where the break occurs between τ and $\tau+1$. The only difference between before and after the break is that the parameter vector has changed from $\boldsymbol{\beta}_1$ to $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$. We may think of this as two regimes or states; first we are in state 1 and then, after the break, we are in state 2. A more compact way to write (3.2) is to use an indicator function, $I_t(t > \tau)$, that takes the values 0 or 1 depending on its argument:

$$R_{t+1} = \boldsymbol{\beta}_1'\mathbf{X}_t(1 - I_t(t > \tau)) + \boldsymbol{\beta}_2'\mathbf{X}_t I_t(t > \tau) + \varepsilon_{t+1}. \tag{3.3}$$

To expand (3.3) to incorporate smooth transitions, we substitute a continuous transition function, $G(t)$, that is bounded on $[0, 1]$ for the indicator function, $I_t(t > \tau)$. Note that since we are studying structural breaks or shifts from one state to another, G must be monotonically increasing or decreasing in t. In this paper, I will only use the logistic function, which is S-shaped and is one of the most popular choices in the literature. (See the top two graphs of Figure 3 for typical shapes of the logistic function.) We can then write the function as

$$R_{t+1} = \boldsymbol{\beta}_1'\mathbf{X}_t(1 - G(t)) + \boldsymbol{\beta}_2'\mathbf{X}_t G(t) + \varepsilon_{t+1} \tag{3.4}$$

$$G(t) = [1 + \exp(-\gamma\{t - c\})]^{-1}, \quad \gamma > 0. \tag{3.5}$$

This is the Time-Varying Regression (TV-R) model of Teräsvirta (1994) and others. The parameters γ and c in G are estimated from the data, and also lend themselves easily to an intuitive interpretation. If $t = c$, then $G = 0.5$, which means that we are halfway through the transition. The location parameter c thus indicates the midpoint of the transition. For higher values of t, the parameter vector $\boldsymbol{\beta}_2$ will become more and more dominant, and vice versa for lower values of t. The parameter γ controls the speed of the transition, so that small positive values will correspond to a slow transition, and high values to a rapid one. In Figure 3, the top graph has a higher value of γ than the middle graph has. As $\gamma \to \infty$, $G \to I$, i.e., very high values of γ give us an (almost) instantaneous transition, and (3.3) is consequently nested in (3.4)-(3.5).

As expressed in (3.4)-(3.5), we have a model with a structural shift explained by t, with a smooth transition between two states. But we could, of course, use the same model for structural shifts explained by another variable. Substituting the variable of our choice, z_t, for t, yields the Smooth Transition Regression (STR) model:

$$R_{t+1} = \boldsymbol{\beta}_1'\mathbf{X}_t(1 - G(z_t)) + \boldsymbol{\beta}_2'\mathbf{X}_t G(z_t) + \varepsilon_{t+1} \tag{3.6}$$

$$G(z_t) = [1 + \exp(-\gamma\{z_t - c\})]^{-1}, \quad \gamma > 0. \tag{3.7}$$

The question then arises of how to decide whether or not a given transition variable is adequate. Note that if $\gamma = 0$, then $G = 0.5$ regardless of the value of the transition variable (t in (3.5), z_t in (3.7)). This means that one can test the existence of a structural shift with respect to a certain transition variable by testing the null hypothesis $\gamma = 0$ against

$\gamma > 0$. Luukkonen, Saikkonen and Teräsvirta (1988) develop a test for this hypothesis, and their procedure is briefly described in the appendix.[11]

3.2. The TV-STR Model. We now have a model that allows for a smooth structural shift explained by one underlying variable. But, as I argued earlier, we need to control for a structural shift explained by t. To be accepted, a variable must be significant even if we control for such a shift. We therefore need an extended version of the STR model.

Lundbergh, Teräsvirta and van Dijk (2003) expand (3.4) in yet another dimension, so that we get 2×2 different states with smooth shifts between them all:

$$\begin{aligned} R_{t+1} &= [\beta_1'\mathbf{X}_t(1-G_1(t)) + \beta_2'\mathbf{X}_t G_1(t)][1-G_2(z_t)] \\ &\quad [\beta_3'\mathbf{X}_t(1-G_1(t)) + \beta_4'\mathbf{X}_t G_1(t)]G_2(z_t) + \varepsilon_{t+1}, \qquad (3.8) \\ G_1(t) &= [1+\exp(-\gamma_1\{t-c_1\})]^{-1}, \quad \gamma_1 > 0, \qquad (3.9) \\ G_2(z_t) &= [1+\exp(-\gamma_2\{z_t-c_2\})]^{-1}, \quad \gamma_2 > 0. \qquad (3.10) \end{aligned}$$

The variable that is believed to explain a structural shift in the process is z_t. Both G_1 and G_2 are logistic functions, but with different arguments, t and z_t respectively, and different parameters: c_1 and γ_1 for G_1, and c_2 and γ_2 for G_2. This is the Time-Varying Smooth Transition Regression (TV-STR) model of Lundbergh, Teräsvirta and van Dijk.

To test if a suggested transition variable is adequate, the authors expand the linearity test used for (3.4)-(3.5) and (3.6)-(3.7), so that one can test (3.8)-(3.10) with a similar procedure. This procedure is also briefly described in the appendix.

Following Lundbergh, Teräsvirta and van Dijk, I use the following strategy to determine whether z_t and/or t are appropriate transition variables.[12] First, I test a linear model against the entire model (3.8)-(3.10). This amounts to testing H_0: $\gamma_1 = \gamma_2 = 0$ against H_1: $\gamma_1 > 0$ and/or $\gamma_2 > 0$. If the null hypothesis is rejected, I continue with two sub-tests, one for each transition variable. The first test is H_{01}: $\gamma_1 = 0|\gamma_2 > 0$ against H_{11}: $\gamma_1 > 0|\gamma_2 > 0$, and the second test is H_{02}: $\gamma_2 = 0|\gamma_1 > 0$ against H_{12}: $\gamma_2 > 0|\gamma_1 > 0$.

We can interpret the sub-tests in the following way: Given that we have accepted H_1, i.e., that there is a structural shift explained by either t, z_t, or both, can we do without one of the variables? If we reject both H_{01} and H_{02}, we conclude that we cannot, and we stay with the full TV-STR model with shifts in both variables. If we reject H_{01} but not H_{02}, we retain the model that only has a structural shift in t (i.e., the TV-R model (3.4)-(3.5)). On the other hand, if we reject H_{02} but not H_{01}, we retain the model that only has a structural shift in z_t (i.e., the STR model (3.6)-(3.7)).

4. Empirical Results and Discussion

4.1. Data. I use both macro variables and financial variables from the U.S. market in the empirical investigation. I have chosen variables that are among the most frequently used variables in the literature, and with definitions that are as close as possible to those

[11] Note that the t-value of γ cannot be given its usual interpretation, since the model is not identified under the null. This applies to Tables 8 and 12.

[12] See the "Specific-to-General-to-Specific Approach" of their paper.

that others have used, noting that some variations exist in the literature. These variables have all been shown to significantly predict returns in some studies. They can also be connected to variables in the CCAPM, (2.1)-(2.2). In other words, they are believed to covary with consumption risk or risk aversion, or to be related to market risk or to the business cycle.

- $\Delta_{12}CLI$ and Δ_6CLI are two measures of the growth rate in the Composite Leading Indicator that is published by The Conference Board. This is similar to the one used in Perez-Quiros and Timmermann (2000).
- Inflation is the yearly change in the Consumer Price Index.
- $\Delta_{12}IP$ (IP) is the yearly change in Industrial Production.
- The Default Premium (DP) is the difference in yields between AAA and BAA rated bonds.
- The Term Premium (TP) is calculated as the difference in yields between a five-year bond and a one-month T-bill.
- The Dividend Yield (DY).
- The Earnings to Price Ratio (EP).
- $\Delta_{12}M1$ is the growth rate in narrow money, M1.
- Realized Volatility is a measure of monthly volatility calculated from daily returns.
- The return on a one-month T-bill (Tb).

Descriptions of how the variables were calculated can be found in the appendix. I have chosen to measure returns and growth rates using log changes, and financial ratios using log differences. The sample period is February 1960–December 2005. The starting date is a result of CLI not being available before that date.[13] All explanatory variables are plotted in Figures 1a-c. To highlight how dependent most variables are on the business cycle, NBER recession periods are also indicated in each figure.

All variables are sampled at the monthly frequency and all regressions are at the one-month horizon.[14] There are several reasons for studying this horizon. It is the shortest horizon at which we have access to all data, particularly the macro data. It is also a decision horizon used by many practitioners in tactical asset allocation. It is also a frequently used horizon in academic studies, for example, in the classic study by Fama and French (1988)[15]. Ang and Bekaert (2007) also find that "predictability is mainly a short-horizon, not a long-horizon, phenomenon."

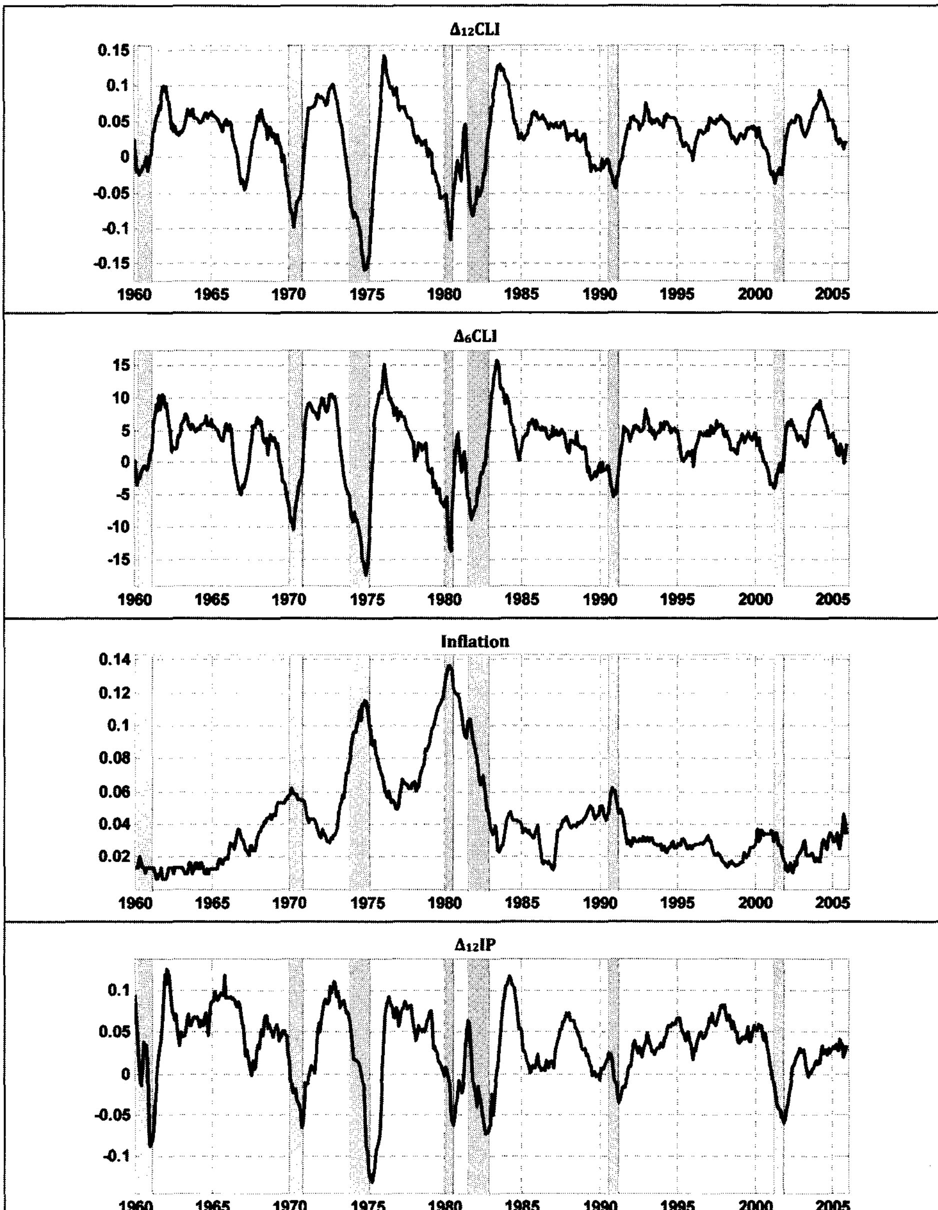
Summary statistics for all variables are collected in Tables 1 and 2. One presumably problematic feature of most series is the high level of autocorrelation. I will correct t- and p-values for autocorrelation by using the Newey and West (1987) covariance matrix with one lag for deciles 1–9, since these have a rather obvious first-order autocorrelation. As

[13] The first observation in this series is for January 1959, which is known in February 1959. After taking yearly differences, the first observation is known in February 1960.

[14] The extreme observation of October 1987 has been deleted as the type of models used in this paper are sensitive to outliers.

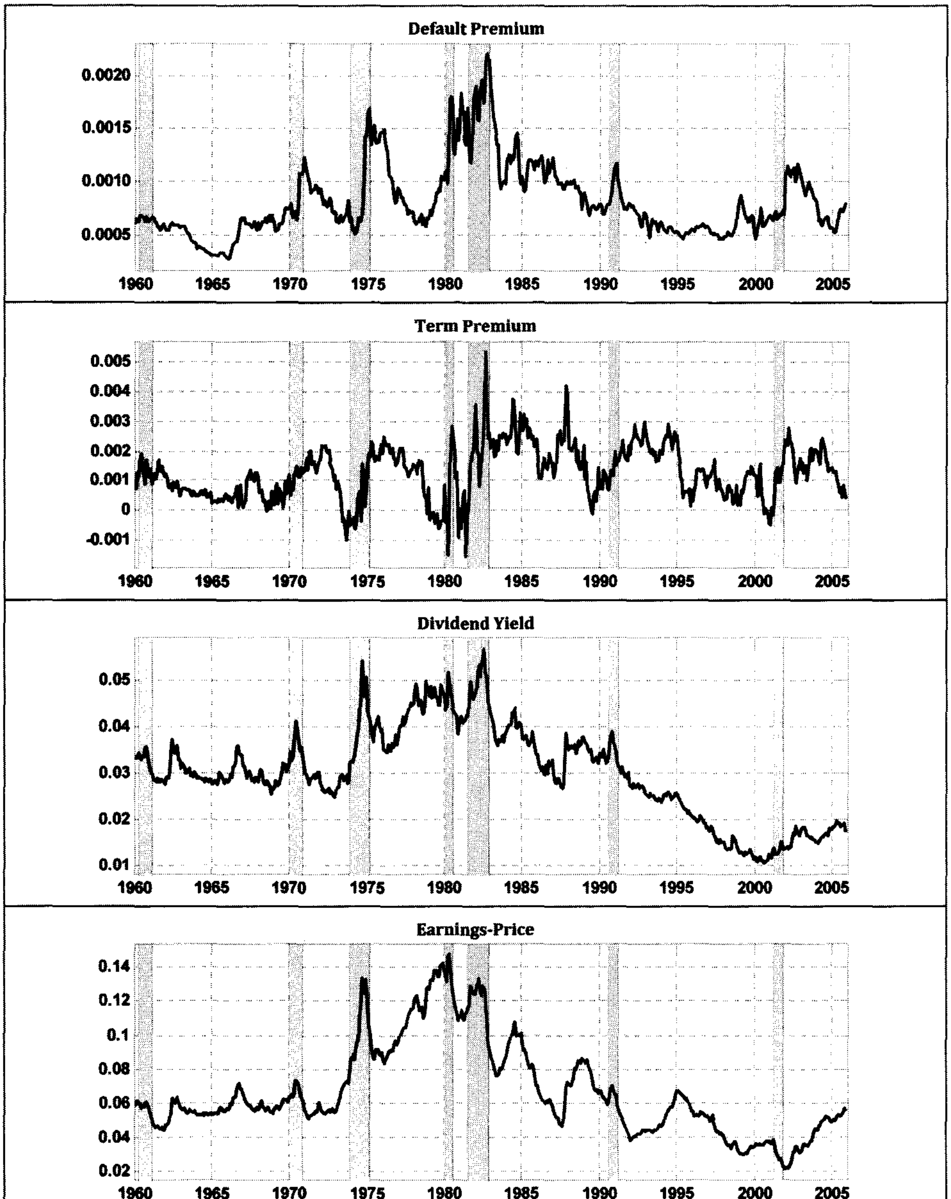
[15] Fama and French use several different sampling frequencies.

Figure 1a
Independent Variables



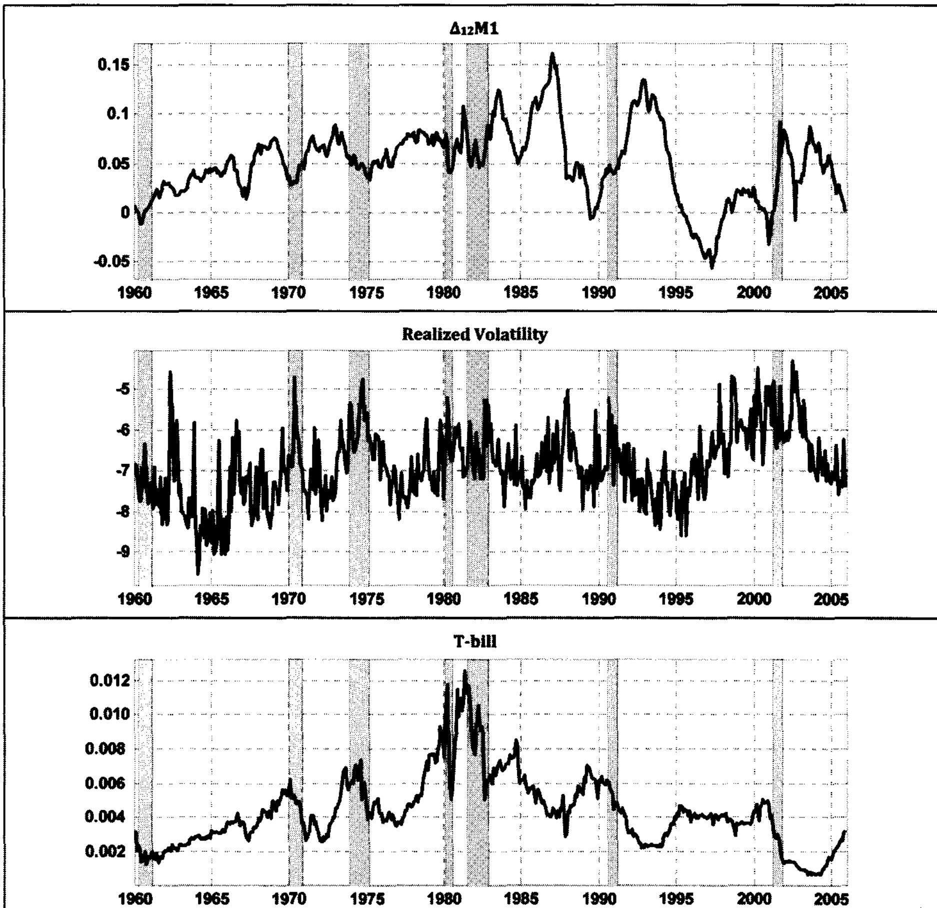
For abbreviations and information about the construction of the variables, see the appendix. Data is sampled monthly from Feb 1960 to Dec 2005.

Figure 1b
Independent Variables



For abbreviations and information about the construction of the variables, see the appendix. Data is sampled monthly from Feb 1960 to Dec 2005.

Figure 1c
Independent Variables



For abbreviations and information about the construction of the variables, see the appendix. Data is sampled monthly from Feb 1960 to Dec 2005.

Table 1
Summary Statistics for the Independent Variables

	Δ_{12}CLI	Δ_6CLI	Infl.	Δ_{12}IP	DP	TP	DY	EP	Δ_{12}M1	Real. Vol.	Tb
mean	0.0256	2.70	0.0413	0.0317	0.000817	0.00128	0.0300	0.0670	0.0496	-6.90	0.00431
min	-0.158	-17.3	0.00669	-0.132	0.000266	-0.00156	0.0105	0.0214	-0.0564	-9.54	0.000659
max	0.142	15.7	0.136	0.126	0.00221	0.00535	0.0565	0.147	0.161	-4.31	0.0126
std	0.0494	5.23	0.0277	0.0450	0.000350	0.000911	0.0101	0.0279	0.0378	0.882	0.00212
skewness	-0.927	-0.863	1.40	-0.794	1.32	0.204	0.0194	0.963	0.0160	0.219	0.992
kurtosis	4.37	4.35	4.51	3.86	4.83	3.56	2.47	3.17	3.26	3.13	4.48
JB, p-val	**0**	**0**	**0**	**0**	**0**	**0.0045**	**0.037**	**0**	0.478	0.094	**0**
ρ_1	0.978	0.970	0.993	0.967	0.971	0.875	0.988	0.994	0.981	0.701	0.972

"JB, p-val" is the p-value for the Jarque-Bera statistic for normality; "ρ_1" is the in-sample first autocorrelation.
The construction of the variables is described in the appendix. The data is sampled monthly from Feb 1960 to Dec 2005.

Table 2
Summary Statistics for the Dependent Variables

	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Index
mean	0.0140	0.0105	0.00865	0.00830	0.00766	0.00743	0.00719	0.00725	0.00689	0.00504	0.00546
min	-0.219	-0.225	-0.232	-0.221	-0.230	-0.216	-0.208	-0.202	-0.192	-0.155	-0.162
max	0.540	0.348	0.328	0.350	0.316	0.292	0.275	0.245	0.221	0.174	0.161
std	0.0775	0.0668	0.0629	0.0606	0.0586	0.0575	0.0557	0.0524	0.0492	0.0418	0.0428
skewness	1.34	0.591	0.293	0.300	0.00648	-0.0601	-0.187	-0.273	-0.250	-0.166	-0.253
kurtosis	9.48	6.08	5.67	5.85	5.32	4.93	4.81	4.47	4.25	3.89	3.83
JB, p-val	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1.1e-12**	**1.4e-9**	**4.5e-5**	**2.4e-5**
ρ_1	0.223	0.257	0.233	0.225	0.198	0.187	0.148	0.155	0.132	0.00182	0.0413

"JB, p-val" is the p-value for the Jarque-Bera statistic for normality; "ρ_1" is the in-sample first autocorrelation.
The dependent variables are excess returns (in excess of the risk-free rate) on portfolios sorted on size, where decile 1 is the 10% smallest companies and decile 10 is the 10% largest. The index is the value-weighted NYSE/AMEX/NASDAQ index from CRSP.
The construction of the variables is described in the appendix. The data is sampled monthly from Feb 1960 to Dec 2005.

there is little evidence of autocorrelation in decile 10 and the index, I will use the White (1980) covariance matrix for the tests involving those portfolios.[16]

Sample correlations for all variables are collected in Table 3. The explanatory variables have, unsurprisingly, quite low correlation with the predicted variables. In most cases, the explanatory variables also have low correlations with each other. The most notable exceptions are the two CLI variables (that attempt to measure the same thing), and EP and DY (that are also closely connected through the payout ratio). The predicted variables are, as can be expected, highly correlated with each other. The index and decile 10 are almost identical. This is because the index is value weighted and consequently dominated by the largest firms, which are exactly the firms included in decile 10. Each decile portfolio is also highly correlated with the ones closest to it in size, but the correlation diminishes with the increasing difference in size.

4.2. Finding Basic Predictive Variables. In Section 3.1, I postponed the problem of choosing the variables in **X** for (3.4), (3.6), and (3.8). Choosing them is not an easy task, as a large number of variables have been suggested in the literature. I begin by illustrating the problem of asset return predictability empirically. In Table 4, we see t-values for β_1 in the regression

$$R^e_{t+1} = \beta_0 + \beta_1 z_t + \varepsilon_{t+1}, \tag{4.1}$$

where R^e_{t+1} is the excess return on the index, and z_t is one of the explanatory variables. For the full period, IP, DP, TP, and Tb are significant, but the subsample tests show that it is primarily during the first period that there is a consistent linear predictive relation. During the second subsample period, none of the variables predicts index returns in a univariate regression. It is noteworthy that DY is the only variable whose coefficient estimate has quite high t-values in both subsample periods, although not sufficiently high to be significantly different from zero. A preliminary conclusion from these regressions is that asset return predictability is both weak and unstable.[17]

However, theoretical reasons do exist for preferring certain predictive variables to others. Remember the two versions of Gordon's growth model, (2.3) and (2.4). They imply that the DY *must* predict returns or growth (see Section 2). Since there is little evidence that it predicts growth, it is reasonable to use the DY, which is probably the most frequently used predictor of returns in the literature.

To continue the search for variables, Table 5 lists t-values for β_1 and β_2 from the bivariate regression

$$R^e_{t+1} = \beta_0 + \beta_1 DY_t + \beta_2 z_t + \varepsilon. \tag{4.2}$$

[16] The choice does not influence the conclusions.

[17] Lettau and van Nieuwerburgh (2006) study instability in linear regressions further. See also Ang and Bekaert (2006).

Table 3
Correlation Matrix of all Variables

	Δ_{12}CLI	Δ_6CLI	Infl.	Δ_{12}IP	DP	TP	DY	EP	Δ_{12}M1	Real. Vol.	Tb	Dec 1	Dec 2	Dec 3	Dec 4	Dec 5	Dec 6	Dec 7	Dec 8	Dec 9	Dec 10
Δ_6CLI	0.94																				
Infl.	-0.60	-0.58																			
Δ_{12}IP	0.57	0.36	-0.33																		
Def. Prem.	-0.22	-0.12	0.50	-0.55																	
Term Pr.	0.29	0.37	-0.17	-0.18	0.40																
Div. Yield	-0.34	-0.31	0.66	-0.18	0.52	0.10															
EP	-0.39	-0.40	0.82	-0.15	0.53	-0.01	0.90														
Δ_{12}M1	0.23	0.23	0.18	-0.05	0.34	0.33	0.34	0.22													
Real. Vol.	-0.32	-0.33	0.23	-0.30	0.32	0	-0.07	0.03	-0.10												
T-bill	-0.37	-0.39	0.72	-0.11	0.50	-0.20	0.64	0.75	0.17	0.16											
Decile 1	-0.02	0.02	0	-0.15	0.11	0.07	0.01	0.02	0.02	0.02	-0.11										
Decile 2	-0.01	0.03	-0.01	-0.15	0.10	0.08	0.01	0.02	0	0.02	-0.13	0.95									
Decile 3	-0.01	0.02	-0.01	-0.15	0.10	0.08	0.02	0.03	-0.02	0.04	-0.12	0.93	0.98								
Decile 4	-0.02	0.02	-0.01	-0.15	0.10	0.09	0.04	0.04	-0.02	0.04	-0.13	0.90	0.96	0.98							
Decile 5	-0.02	0.01	-0.01	-0.14	0.11	0.10	0.04	0.05	-0.02	0.05	-0.11	0.86	0.94	0.96	0.98						
Decile 6	-0.03	0	0	-0.15	0.11	0.11	0.06	0.06	-0.02	0.06	-0.10	0.83	0.91	0.94	0.96	0.98					
Decile 7	-0.02	0	-0.01	-0.13	0.11	0.11	0.06	0.06	-0.03	0.07	-0.10	0.77	0.86	0.91	0.94	0.96	0.98				
Decile 8	-0.03	-0.01	0	-0.14	0.12	0.12	0.08	0.07	-0.02	0.06	-0.09	0.72	0.83	0.88	0.91	0.94	0.97	0.98			
Decile 9	-0.02	0	-0.02	-0.13	0.12	0.14	0.08	0.06	-0.02	0.06	-0.10	0.66	0.77	0.83	0.87	0.90	0.94	0.96	0.98		
Decile 10	0.01	0.02	-0.06	-0.09	0.08	0.14	0.06	0.03	-0.04	0.03	-0.09	0.49	0.60	0.66	0.70	0.75	0.78	0.82	0.86	0.90	
Index	0	0.02	-0.05	-0.10	0.09	0.14	0.07	0.04	-0.04	0.04	-0.09	0.56	0.67	0.73	0.77	0.81	0.85	0.88	0.91	0.94	0.99

For abbreviations and the construction of the variables, see the appendix. Note that the index and the ten decile portfolios are lagged one month w.r.t. the other variables.

Table 4
Univariate Predictability for the Index

	Δ_{12}CLI	Δ_6CLI	Infl.	Δ_{12}IP	DP	TP	DY	EP	Δ_{12}M1	Real. Vol.	Tb
All	0.01	0.35	-0.86	**-2.28**	**2.04**	**3.15**	1.29	0.78	-0.92	0.77	**-2.07**
1st half	-0.14	0.27	-0.58	**-2.62**	1.88	**3.18**	1.76	0.75	**-2.04**	0.42	**-2.23**
2nd half	-0.19	-0.20	-0.01	0.09	0.91	0.65	1.53	1.66	-0.07	0.30	-0.17

The table reports t-statistics of β from the regression $R^e_{t+1} = \alpha + \beta z_t + \varepsilon_{t+1}$, where R^e_t is the index excess return, and z_t is the variable listed in the heading.
The full sample is from Feb 1960 to Dec 2005, the first half is from Feb 1960 to Dec 1982, and the second half is from Jan 1983 to Dec 2005.
Standard errors are calculated using White's (1980) heteroscedasticity-consistent covariance matrix.

Table 5
Bivariate Predictability for the Index

	Δ_{12}CLI	Δ_6CLI	Infl.	Δ_{12}IP	DP	TP	DY	EP	Δ_{12}M1	Real. Vol.	Tb
All	1.45/0.47	1.56/0.80	**2.72/-2.41**	0.97/**-2.10**	0.45/1.46	1.03/**3.03**	-	1.53/-0.99	1.68/-1.59	1.35/0.87	**3.13/-3.89**
1st half	**2.20**/0.97	**2.44**/1.41	**3.31/-2.92**	0.89/**-2.06**	0.88/0.75	1.31/**2.92**	-	**3.18/-2.86**	**2.27/-2.71**	1.90/-0.67	**4.29/-4.67**
2nd half	1.58/-0.37	1.57/-0.37	1.84/-1.09	1.53/0.05	1.30/-0.14	1.52/-0.18	-	0.25/0.65	1.88/-0.99	1.93/0.95	**2.13**/-1.79

The table reports t-statistics of β_1/β_2 from the regression $R^e_{t+1} = \alpha + \beta_1 DY_t + \beta_2 z_t + \varepsilon_{t+1}$, where R^e_t is the index excess return, DY_t is the dividend yield, and z_t is the variable listed in the heading.
The full sample is from Feb 1960 to Dec 2005, the first half is from Feb 1960 to Dec 1982, and the second half is from Jan 1983 to Dec 2005.
Standard errors are calculated using White's (1980) heteroscedasticity-consistent covariance matrix.

There are some qualitative changes as compared to Table 4, but the only dramatic change is found for DY in combination with Tb.[18] This combination is significant, or near-significant, for all the periods listed. Ang and Bekaert (2007) also note that the DY is particularly significant when used in combination with the short rate at short horizons.[19]

This is, of course, only one of many possible ways to select the components of **X**. Other methods of empirical statistics include using all variables, eliminating variables until all are significant at some standard significance level ("general-to-specific"), or expanding the model one variable at a time as long as new variables are significant ("specific-to-general"). The strategy chosen here has the advantage of being parsimonious and well connected to basic financial theory. DY and Tb are also among the variables most frequently used in the predictability literature and are, for this reason, of particular interest in the present context.

4.3. Testing for Shifts with Different Transition Variables. With **X** defined, we can turn to the main questions of this paper: Are there structural shifts in the process in which **X** predicts returns, and if so, can we explain these by underlying macroeconomic and financial variables? I will arrive at an answer in four steps. First, I test the null hypothesis of no structural shift against the alternative of a structural shift explained by only one transition variable. The main reason for this is to show how easy it is to be misled to believe that several underlying variables explain structural shifts, and how important it is to control for a shift explained by t. Second, using the sub-tests described in Section 3.2, I test for a structural shift explained by one transition variable and t simultaneously. Third, in Section 4.4, I estimate the chosen models, and fourth, in Section 4.5, I expand the model chosen for the index to arrive at an empirically useful model for predicting the index with DY and Tb. This is a model that also allows us to study the structural shift in more detail with respect to timing and duration.

To begin with, Table 6 reports p-values for $H_0 : \gamma = 0$ (i.e., "no shift") in (3.6)-(3.7) for each of the decile portfolios and the index, and with each of the explanatory variables as transition variable, z_t.[20] Several variables provide low p-values, indicating that we should reject the no-shift hypothesis for those variables. For IP and DP, we reject the null for all deciles, and for t, DY, EP, and M1, we reject it for most of them. It seems as though we have found structural shifts explained by quite a few underlying variables.

Next, I shall find out how controlling for structural shifts with t as explanatory variable changes the results. Note that there are now several different possible outcomes. We may or may not reject $H_0 : \gamma_1 = \gamma_2 = 0$ in (3.8)-(3.10). If we do not, we accept the hypothesis that there is no structural shift. If, on the other hand, we do reject H_0, we proceed to the sub-tests, testing the hypotheses H_{01} and H_{02}, to see whether we can do without either the transition variable, z_t, or t.

[18] Several researchers have examined the predictability of returns by the short rate: Fama and Schwert (1977), Campbell (1987), and Shiller and Beltratti (1992) among others.

[19] To mitigate data snooping concerns, they confirm this using data from the UK, France, and Germany.

[20] Note that with t as transition variable, the STR model (3.6)-(3.7) becomes the TV-R model (3.4)-(3.5).

Table 7 contains p-values for H_0 as well as the sub-hypotheses H_{01} and H_{02}. The last row in the table contains a rough conclusion for each variable along the decile dimension, so that deciles for which the conclusion is the same are grouped together. As an example, for $\Delta_{12}CLI$ we accept the hypothesis that there is no structural shift for deciles 1–4 (we do not reject H_0), and reject it (we reject H_0 and H_{01}, but not H_{02}) for deciles 5–10 and the index.

As can readily be seen from the last row of Table 7, the only variable that is accepted is t.[21] Furthermore, it is primarily for the deciles of large firms that we reject the null hypothesis of no structural shift. For the deciles of small firms, there is very little evidence of structural shifts. For some in-between deciles the evidence is mixed.[22] In several cases we reject H_0 in favor of the TV-STR model, but as we do not reject either H_{01} or H_{02}, it is unclear what explains the shift. I have not examined these cases any further.

It seems reasonable to conclude that we cannot link structural shifts in the way DY and Tb predict returns to any of the variables used in this paper, and that time is the best "explanation" we can offer. One way to interpret this result is that structural changes in the predictability of short horizon asset returns are not caused by recurrent economic events, such as the business cycle, high/low inflation, time-varying risk or risk aversion (as measured by the default premium), or earnings. Instead, they seem to be caused by unique, one-off events that possibly extend over a period of time. The arguments presented by Campbell and Cochrane (1999), and Perez-Quiros and Timmermann (2000) that the business cycle can affect returns, either because it affects risk aversion (Campbell and Cochrane) or because it tightens the conditions on the credit markets (Perez-Quiros and Timmermann), do not appear to hold in the present context (see Section 3). This is also the case when I use the two business cycle indicators ($\Delta_{12}CLI$ and Δ_6CLI) that have the additional benefit of hindsight (see the appendix).

4.4. Analyzing the Shifts Explained by t. We now have ample evidence for a structural shift explained by t for the deciles of large firms, some evidence for the mid-range, and little or no evidence for the deciles of small firms. I therefore estimate the TV-R model (3.4)-(3.5) for deciles 5–10 and the index. Table 8 presents parameter estimates. The first column contains estimates for all parameters. Many of them are insignificant, primarily in state 2. One possible explanation could be that we have fewer observations in that state; the estimated structural shift occurs in about 1995, leaving approximately 130 observations in state 2. But the low significance levels also cast doubt on whether all components of $\mathbf{X}$ actually predict returns after the shift. The t-values for $\widehat{\beta}_{2,Tb}$ are very small for almost all deciles, and the second column therefore presents estimates with the restriction $\beta_{2,Tb} = 0$. The restriction has very small effects on the other estimates.

For deciles 9–10 and the index, I find that returns are predictable in both states of the world. In state 1, both DY and Tb predict returns, and in state 2, DY does so.

[21] To be exact, for IP in decile 10 we actually reject (with a p-value of 0.049) that there is no shift with the state variable as transition variable. As this is the only exception, I have disregarded it.

[22] t is accepted for at least ten out of eleven variables from decile 9 and up, and for seven out of eleven variables in decile 8. For deciles 5, 6, and 7, t is only accepted for between two and four variables, and for deciles 1-4 for none. Note that for t to be accepted, we must reject both "No TV-STR" *and* "No t."

Table 6
LM Tests for Linearity against the STR Model

	t	Δ_{12}CLI	Δ_6CLI	Infl.	Δ_{12}IP	DP	TP	DY	EP	Δ_{12}M1	Real. Vol.	Tb
Decile 1	0.103	0.534	0.427	0.118	**0.004**	**0.003**	0.408	0.058	**0.031**	**0.046**	0.403	0.085
Decile 2	0.099	0.463	0.379	0.076	**0.010**	**0.005**	0.407	**0.036**	**0.012**	**0.044**	0.292	0.160
Decile 3	**0.041**	0.374	0.285	0.096	**0.008**	**0.004**	0.441	**0.014**	**0.007**	0.051	0.199	0.118
Decile 4	**0.032**	0.407	0.366	0.098	**0.013**	**0.005**	0.476	**0.020**	**0.010**	**0.039**	0.153	0.131
Decile 5	**0.010**	0.203	0.178	0.064	**0.013**	**0.004**	0.258	**0.008**	**0.007**	**0.015**	0.075	0.109
Decile 6	**0.004**	0.215	0.197	0.082	**0.011**	**0.004**	0.179	**0.006**	**0.009**	**0.012**	0.054	0.139
Decile 7	**0.003**	0.202	0.165	0.097	**0.021**	**0.004**	0.174	**0.009**	**0.012**	**0.013**	**0.036**	0.155
Decile 8	**0.001**	0.127	0.100	0.107	**0.012**	**0.004**	0.099	**0.010**	**0.021**	**0.013**	**0.036**	0.233
Decile 9	**5e-4**	0.083	0.067	0.140	**0.010**	**0.004**	0.067	**0.018**	0.057	**0.025**	**0.029**	0.287
Decile 10	**1e-5**	**0.012**	**0.013**	0.062	**0.003**	**0.008**	0.070	0.094	0.189	**0.030**	0.158	0.234
Index	**2e-5**	**0.017**	**0.015**	0.082	**0.003**	**0.005**	0.078	0.060	0.126	**0.020**	0.146	0.242

The table reports p-values for H_0: linear against the STR model. The linear part consists of $X_t = [1\ DY_t\ Tb_t]$, where DY_t is the dividend yield, and Tb_t is the return on a one-month T-bill. For deciles 1—9 the statistics are calculated according to Newey-West (1987) with 1 lag, and for decile 10 and the index they are calculated according to White (1980). The test procedure is further described in the appendix.

Table 7
LM Tests for Linearity against the TV-STR Model

		Δ_{12}CLI	Δ_6CLI	Infl.	Δ_{12}IP	DP	TP	DY	EP	Δ_{12}M1	Real Vol.	Tb
Decile 1	H_0: No TV-STR	0.281	0.208	0.147	0.056	0.076	0.433	0.075	0.124	0.217	0.162	0.192
	H_{01}: No t	(0.158)	(0.120)	(0.317)	(0.666)	(0.984)	(0.341)	(0.379)	(0.522)	(0.146)	(0.564)	(0.322)
	H_{02}: No st.var	(0.341)	(0.237)	(0.183)	(0.014)	(0.045)	(0.944)	(0.086)	(0.153)	(0.185)	(0.409)	(0.285)
Decile 2	H_0: No TV-STR	0.225	0.162	0.146	0.084	0.100	0.339	0.133	0.101	0.242	0.148	0.263
	H_{01}: No t	(0.129)	(0.090)	(0.408)	(0.578)	(0.931)	(0.241)	(0.588)	(0.580)	(0.221)	(0.366)	(0.341)
	H_{02}: No st.var	(0.287)	(0.173)	(0.159)	(0.033)	(0.078)	(0.889)	(0.187)	(0.129)	(0.204)	(0.385)	(0.424)
Decile 3	H_0: No TV-STR	0.135	0.099	0.138	0.062	0.053	0.134	0.123	0.053	0.164	0.080	0.222
	H_{01}: No t	(0.072)	(0.051)	(0.321)	(0.436)	(0.648)	(0.087)	(0.613)	(0.468)	(0.162)	(0.135)	(0.271)
	H_{02}: No st.var	(0.314)	(0.188)	(0.323)	(0.069)	(0.085)	(0.854)	(0.326)	(0.154)	(0.295)	(0.400)	(0.730)
Decile 4	H_0: No TV-STR	0.131	0.103	0.124	0.068	0.060	0.156	0.114	0.050	0.137	0.066	0.186
	H_{01}: No t	(0.069)	(0.052)	(0.312)	(0.368)	(0.624)	(0.115)	(0.515)	0.476	(0.149)	(0.154)	(0.220)
	H_{02}: No st.var	(0.322)	(0.215)	(0.315)	(0.085)	(0.103)	(0.924)	(0.331)	0.170	(0.244)	(0.319)	(0.608)
Decile 5	H_0: No TV-STR	0.049	0.045	0.057	0.035	0.031	0.057	0.040	0.022	0.055	0.018	0.068
	H_{01}: No t	0.029	0.031	(0.174)	0.147	0.409	(0.066)	0.329	0.321	(0.073)	0.062	(0.066)
	H_{02}: No st.var	0.243	0.204	(0.417)	0.095	0.139	(0.915)	0.268	0.223	(0.232)	0.251	(0.678)
Decile 6	H_0: No TV-STR	0.040	0.037	0.046	0.027	0.025	0.036	0.030	0.017	0.032	0.013	0.055
	H_{01}: No t	0.023	0.023	0.132	0.117	0.319	0.054	0.257	0.225	0.047	0.058	(0.048)
	H_{02}: No st.var	0.218	0.172	0.414	0.085	0.136	0.860	0.221	0.260	0.109	0.243	(0.742)
Decile 7	H_0: No TV-STR	0.036	0.031	0.043	0.034	0.024	0.036	0.033	0.009	0.018	0.010	0.052
	H_{01}: No t	0.021	0.022	0.112	0.090	0.262	0.051	0.199	0.138	0.044	0.046	(0.048)
	H_{02}: No st.var	0.228	0.146	0.467	0.200	0.146	0.904	0.298	0.186	0.053	0.212	(0.805)
Decile 8	H_0: No TV-STR	0.016	0.014	0.021	0.011	0.010	0.009	0.015	0.007	0.009	0.005	0.018
	H_{01}: No t	0.011	0.014	0.051	0.045	0.135	0.023	0.083	0.066	0.024	0.019	0.011
	H_{02}: No st.var	0.237	0.171	0.588	0.155	0.158	0.717	0.312	0.339	0.075	0.288	0.914
Decile 9	H_0: No TV-STR	0.011	0.008	0.016	0.007	0.007	0.006	0.013	0.008	0.006	0.005	0.012
	H_{01}: No t	0.009	0.010	0.035	0.033	0.127	0.024	0.039	0.029	0.024	0.012	0.006
	H_{02}: No st.var	0.193	0.123	0.799	0.132	0.158	0.641	0.377	0.441	0.059	0.444	0.977
Decile 10	H_0: No TV-STR	4e-4	2e-4	0.001	6e-5	2e-4	0.001	0.001	0.001	4e-4	4e-4	0.001
	H_{01}: No t	0.001	0.001	0.004	0.001	0.004	0.003	4e-4	0.001	0.001	0.001	0.001
	H_{02}: No st.var	0.080	0.063	0.839	0.049	0.205	0.773	0.266	0.191	0.172	0.148	0.961
Index	H_0: No TV-STR	0.001	0.001	0.002	2e-4	0.001	0.001	0.002	0.001	0.001	0.001	0.002
	H_{01}: No t	0.002	0.002	0.008	0.003	0.010	0.005	0.001	0.001	0.002	0.002	0.001
	H_{02}: No st.var	0.087	0.060	0.817	0.061	0.178	0.765	0.282	0.204	0.146	0.149	0.957
Summary		1-4: - 5-I: t	1-4: - 5-I: t	1-5: - 6-8: ? 9-I: t	1-4: - 5-7: ? 8-I: t	1-4: - 5-9: ? 10-I: t	1-5: - 6-7: ? 8-I: t	1-4: - 5-8: ? 9-I: t	1-3: - 4-8: ? 9-I: t	1-5: - 6-I: t	1-4: - 5-6: ? 7-I: t	1-7: - 8-I: t

The table reports p-values for three tests concerning the TV-STR model. The linear part consists of $X_t = [1\ DY_t\ Tb_t]$, where DY_t is the dividend yield, and Tb_t is the return on a one-month T-bill.

The first value in each cell tests H_0: linear against the TV-STR model. The second value tests H_{01}: no transition in *t* (i.e., only transition in the state variable) against TV-STR. The third value tests H_{02}: no transition in the state variable (i.e., only transition in *t*) against TV-STR. If the "No TV-STR" is not rejected, the two other tests are not applicable, since they are conditioned on rejecting that hypothesis. Such p-values have therefore been enclosed within parentheses.

The "Summary" row summarizes a rough conclusion for each variable. E.g., for inflation we find no evidence of a structural shift for deciles 1—5, we have unclear results for deciles 6—8, and we find a structural shift explained by *t* for deciles 9—10 and the index.

For deciles 1—9 the statistics are calculated according to Newey-West (1987) with 1 lag, and for decile 10 and the index they are calculated according to White (1980). The test procedure is further described in the appendix.

Table 8
Parameter Estimates for the TV-R Model

	Decile 5		Decile 6		Decile 7		Decile 8	
	Unrestricted	$\beta_{2,Tb} = 0$	Unrestricted	$\beta_{2,Tb} = 0$	Unrestricted	$\beta_{2,Tb} = 0$	Unrestricted	$\beta_{2,Tb} = 0$
$\beta_{1,C}$	-0.0610	-0.0610	-0.0630	-0.0630	-0.0530	-0.053	-0.0587	-0.0585
	(-3.48)	(-3.48)	(-3.79)	(-3.79)	(-3.73)	(-3.73)	(-3.90)	(-3.88)
$\beta_{1,DY}$	3.30	3.30	3.33	3.33	3.03	3.03	3.09	3.09
	(5.22)	(5.22)	(5.48)	(5.48)	(5.41)	(5.41)	(5.56)	(5.54)
$\beta_{1,Tb}$	-10.2	-10.2	-9.92	-9.92	-9.51	-9.51	-8.97	-8.95
	(-5.40)	(-5.39)	(-5.25)	(-5.24)	(-5.21)	(-5.21)	(-5.12)	(-5.12)
$\beta_{2,C}$	-0.00172	-0.00693	-0.00552	-0.0101	-0.0221	-0.0288	-0.0511	-0.0392
	(0.10)	(-0.39)	(-0.31)	(-0.56)	(-0.83)	(-0.97)	(-1.22)	(-1.08)
$\beta_{2,DY}$	1.03	0.997	1.12	1.09	2.36	2.42	3.69	3.18
	(1.22)	(1.22)	(1.32)	(1.32)	(1.47)	(1.44)	(1.61)	(1.41)
$\beta_{2,Tb}$	-1.93	–	-1.74	–	-1.95	–	1.87	–
	(-0.52)	–	(-0.46)	–	(-0.52)	–	(0.32)	–
γ	900	1000	900	1000	905	762	27.4	28.5
	(0.08)	(0.10)	(0.10)	(0.13)	(1.08)	(1.12)	(1.80)	(1.93)
c	0.674 [1/1991]	0.673 [1/1991]	0.672 [11/1990]	0.672 [1/1991]	0.762 [1/1995]	0.762 [2/1995]	0.761 [1/1995]	0.752 [8/1994]
	(120)	(133)	(181)	(281)	(726)	(593)	(22.1)	(23.4)
AIC/BIC	-3150.1/-3115.6	-3151.8/-3121.7	-3170.6/-3136.2	-3172.4/-3142.2	-3203.8/-3169.3	-3205.4/-3175.3	-3273.8/-3239.3	-3275.6/-3245.5

	Decile 9		Decile 10		Index	
	Unrestricted	$\beta_{2,Tb} = 0$	Unrestricted	$\beta_{2,Tb} = 0$	Unrestricted	$\beta_{2,Tb} = 0$
$\beta_{1,C}$	-0.051	-0.0505	-0.0359	-0.0353	-0.0389	-0.0383
	(-3.65)	(-3.64)	(-2.91)	(-2.89)	(-3.14)	(-3.13)
$\beta_{1,DY}$	2.81	2.80	2.06	2.05	2.23	2.22
	(5.32)	(5.32)	(4.28)	(4.28)	(4.67)	(4.66)
$\beta_{1,Tb}$	-8.43	-8.42	-6.50	-6.52	-7.03	-7.04
	(-5.08)	(-5.08)	(-4.75)	(-4.76)	(-5.04)	(-5.04)
$\beta_{2,C}$	-0.0646	-0.0575	-0.102	-0.0663	-0.0936	-0.0643
	(-1.54)	(-1.63)	(-3.13)	(-2.45)	(-2.87)	(-2.34)
$\beta_{2,DY}$	4.64	4.34	6.24	4.69	5.88	4.58
	(2.03)	(1.98)	(3.42)	(2.77)	(3.21)	(2.66)
$\beta_{2,Tb}$	1.12	–	4.65	–	3.82	–
	(0.22)	–	(1.30)	–	(1.05)	–
γ	38.3	41.3	60.1	85.4	59.2	81.3
	(1.82)	(2.00)	(2.37)	(2.01)	(2.30)	(2.03)
c	0.773 [7/1995]	0.770 [6/1995]	0.778 [10/1995]	0.768 [4/1995]	0.777 [9/1995]	0.768 [4/1995]
	(36.3)	(46.3)	(78.9)	(99.9)	(75.1)	(95.3)
AIC/BIC	-3344.5/-3310.0	-3346.4/-3316.2	-3520.8/-3486.3	-3521.2/-3491.0	-3496.2/-3461.7	-3497.1/-3467.0

The full model is: $R^e_{t+1} = (\beta_{1,C} + \beta_{1,DY}DY_t + \beta_{1,Tb}Tb_t)(1 - G_t) + (\beta_{2,C} + \beta_{2,DY}DY_t + \beta_{2,Tb}Tb_t)G_t + \varepsilon_{t+1}$; $\varepsilon_t \sim N(0,\sigma^2)$; $G_t = (1 + \exp\{-\gamma(t - c)\})^{-1}$, where R^e_t is the decile excess return, DY_t is the dividend yield, and Tb_t is the return on a one-month T-bill. G_t is the transition function with argument t, and parameters γ and c.

In the upper part of Figure 2, we see the estimated transition function, i.e., a plot of $G(t; \widehat{\gamma}, \widehat{c})$ in (3.5) against t. The plots for the different deciles are close to identical, so the figure only shows the plot for the index with the restriction. (This corresponds to the estimates for the index in Table 8, column 2.) The transition function splits the data, roughly, into two parts with a smooth transition during 1993–1997. The second graph of Figure 2 contains model predictions together with realizations. It can be instructive to analyze what happened in the data at the point when the transition is estimated to have occurred. In about 1995 the DY starts to trend away from the bounds it has kept itself within during the first part of the sample (see Figure 1b). It would therefore appear that we have a new state in DY, similar to what is argued in Lettau and van Nieuwerburgh (2006).

For deciles 5–8, return is predictable in state 1 but not so in state 2. This result is most likely due, at least in part, to the fact that the deciles of small firms are noisier than the deciles of large firms. One indication of this is that the t-values in state 2 increase with firm size: Going from decile 5 to decile 10, the t-values for $\widehat{\beta}_{2,DY}$ are 1.08, 1.19, 1.32, 1.41, 1.98, and 2.79. Furthermore, for deciles 5 and 6, I find the shift to have occurred in about January 1991, instead of in 1995 as for the others. It is likely that the model here is picking up something related to the end of the recession at that time.

The estimated betas are all of similar magnitudes across deciles, and all significant estimates have the expected sign (i.e., positive parameters for DY and negative for Tb), but there is systematic variation. In state 1, deciles of large firms are less sensitive to the variables than deciles of small firms (i.e., the parameter estimates are lower in absolute value for deciles of large firms). But in state 2, deciles of large firms are *more* sensitive to the variables. The latter is in contrast to what is found, for instance, in Perez-Quiros and Timmermann (2000), i.e., that small firms are more sensitive to risk factors than large firms.

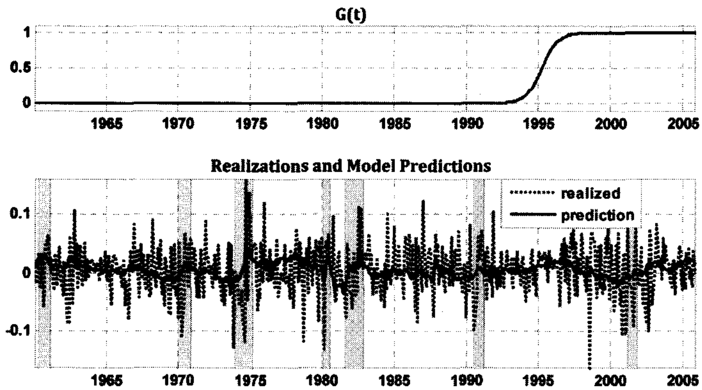
Table 9 contains parameter estimates for the linear model (3.1) for the remaining deciles. They show a similar pattern of increasing t-values, which, again, is probably due to systematic differences in noise. There is also evidence that returns are predictable for these deciles.[23] Note also that there are no indications of more predictability in deciles of small firms than in deciles of large firms.

What differences are there between the state before and the state after the shift? For deciles 7–8, the parameter estimates of $\beta_{1,DY}$ are similar in size to the estimates for $\beta_{2,DY}$, whereas for deciles 9–10 and the index the latter are substantially higher than the former.[24] Comparisons of the average predictions before and after the shift are presented in Table 10. Some weak evidence is found, that the average predicted return has increased after the shift. Note, however, that the increase is only significant for deciles 5–7. Furthermore, the increase is mainly due to a much larger constant in state 2 than in state 1.

[23] However, when testing for subsamples that are similar to state 2, i.e. starting in the 1990's and ending at the sampe end, no variable turns out to be significant (not shown). It seems wise not to put too much emphasis on the predictability of returns in latter years for these deciles either.

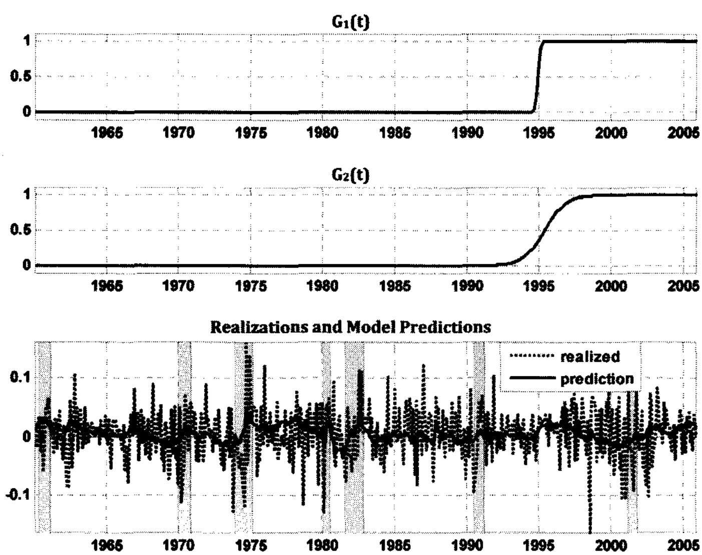
[24] Note that this does not immediately imply that the predicted returns are higher after the shift than before, since DY is lower in the period after the shift. (See Figure 1b.)

Figure 2
TV-R Transition Function and Model Predictions for the Index



The transition function is the function $G(t)$ in the TV-R model. The estimated model is:
$R^e_{t+1} = \{\beta_{1,C} + \beta_{1,DY}DY_t + \beta_{1,Tb}Tb_t\}(1 - G_t) + \{\beta_{2,C} + \beta_{2,}\mathrm{DYDYt}\}G_t + \varepsilon_{t+1}$; $\varepsilon_t \sim N(0,\sigma^2)$; $G_t = (1 + \exp\{-\gamma(t - c)\})^{-1}$,
where R^e_t is the index excess return, DY_t is the dividend yield, and Tb_t is the return on a one-month T-bill.

Figure 3
TV(2)-R Transition Functions and Model Predictions for the Index



The transition functions are the functions $G_1(t)$ and $G_2(t)$ in the TV(2)-R model. The estimated model is:
$R^e_{t+1} = [\beta_{1,C} + \beta_{1,DY}DY_t + \beta_{1,Tb}Tb_t](1 - G_{1t})(1 - G_{2t}) + [\beta_{4,C} + \beta_{4,DY}DY_t]G_{1t}G_{2t} + \varepsilon_{t+1}$; $\varepsilon_t \sim N(0,\sigma^2)$;
$G_{1t} = (1 + \exp\{-\gamma_1(t - c_1)\})^{-1}$, $G_{2t} = (1 + \exp\{-\gamma_2(t - c_2)\})^{-1}$,
where R^e_t is the index excess return, DY_t is the dividend yield, and Tb_t is the return on a one-month T-bill.

Table 9
Linear Parameter Estimates for Deciles 1—4

	Decile 1	Decile 2	Decile 3	Decile 4
β_C	0.0143	0.0102	0.00606	0.00399
	(1.14)	(0.97)	(0.62)	(0.42)
β_{DY}	1.01	**1.08**	**1.08**	**1.18**
	(1.71)	**(2.28)**	**(2.44)**	**(2.75)**
β_{Tb}	**-7.08**	**-7.45**	**-6.94**	**-7.19**
	(-3.15)	**(-3.90)**	**(-3.79)**	**(-4.03)**
AIC/BIC	-2820.5/-2807.6	-2990/-2977.1	-3056.7/-3043.7	-3101.5/-3088.6

The model is: $R^e_{t+1} = \beta_C + \beta_{DY}DY_t + \beta_{Tb}Tb_t) + \varepsilon_{t+1}$; $\varepsilon_t \sim N(0,\sigma^2)$, where R^e_t is the decile excess return, DY_t is the dividend yield, and Tb_t is the return on a one-month T-bill.

Table 10
Comparing the Pre- and Post-Shift Means

	$\bar{R}^e_{t<\hat{c}}$	$\bar{R}^e_{t\geq\hat{c}}$	t-value
Decile 5	**0.00563**	**0.01182**	**5.88**
Decile 6	**0.00601**	**0.01036**	**4.08**
Decile 7	**0.00634**	**0.00994**	**3.38**
Decile 8	0.00738	0.00677	-0.58
Decile 9	0.00700	0.00647	-0.44
Decile 10	0.00487	0.00563	0.62
Index	0.00513	0.00652	1.18

Mean Predicted R^e before and after the estimated structural shift. The t-value is calculated as $t = \frac{\bar{R}^e_{t\geq\hat{c}} - \bar{R}^e_{t<\hat{c}}}{\sqrt{s^2_{t\geq\hat{c}}/T_{t\geq\hat{c}} + s^2_{t<\hat{c}}/T_{t<\hat{c}}}}$.

For this part of the analysis, we can conclude that we have found evidence of a structural shift, that is, in time, similar to that of Paye and Timmermann (2006), Pastor and Stambaugh (2001), and Lettau and van Nieuwerburgh (2006). In contrast to those papers, the shift is a smooth transition lasting a few years in the mid-1990s, and it is most easily observed for deciles of large firms. After the shift, returns are more sensitive to the dividend yield, and the short rate no longer predicts returns. This result is in some contrast to the statement by Ang and Bekaert (2007) that the short rate is the most robust predictor of returns. In addition, there are still indications that the predictability of returns has decreased after the shift, in the sense that pre-shift t-values are larger in absolute value than post-shift ones.

4.5. Evaluation and Extension. Table 11 presents diagnostics for the estimated TV-R models, with the estimated linear models also included for comparison. The information criteria, AIC and BIC, select different models for deciles 5–9: AIC selects the TV-R model, whereas BIC selects the linear alternative.[25] For decile 10 and the index, both criteria choose the TV-R model. Quite unsurprisingly, for the deciles of large firms there seems to be some heteroscedasticity in the residuals.

Interestingly, there are indications of remaining nonlinearity explained by t for the deciles of large firms and the index (see the RNL(t) column). This indicates that the structural shift we have found may not be enough to model the time-varying behavior of these portfolios. The structure of the shift appears more complicated than what can be described by a simple TV-R model. To study this feature, I estimate a variant of the TV-STR model (3.8)-(3.10). We have already rejected the notion that any of the variables examined explains a structural shift in the predictive regression. But, since there are signs of a further structural shift in t, it makes sense to simply use t as the transition variable in both transition functions, (3.9) and (3.10). I call the resulting model the TV(2)-R model.[26] The extra transition function adds flexibility to the model.

Table 12 presents parameter estimates. The first column contains estimates for all 16 parameters. Most of them turn out insignificant, and some estimates are of a magnitude (such as $\widehat{\beta}_{3,Tb} = -242$) that makes economic interpretation difficult. Note, however, that states 2 and 3 in this model do not correspond to any real time period: State 1 corresponds, roughly, to 1960–2001 and state 4, roughly, to 2002–2005, but states 2 and 3 correspond to a period that is simultaneously between 1960–2001 and 2002–2005, and that never happens. I eliminate variables using the general-to-specific approach, deleting the beta that has the lowest t-value until all are higher than 1 in absolute value. This procedure eliminates all betas in states 2 and 3, as well as $\beta_{4,Tb}$, and then all remaining betas are significant. The estimates of the restricted model are presented in the second column of Table 12, and are quite similar in magnitude to those of the TV-R model (see

[25] It is not completely clear how to interpret the information criteria in the present case. When calculating them, each parameter in the estimated model carries a weight of 1. But since the TV-R model is linear under the restriction $\gamma = 0$, but not identified, one might argue that this punishes that model too heavily.

[26] It is, of course, a special case of the TV-STR model (3.8)-(3.10). The name TV(2)-R indicates that it has two time-dependent shifts. In that respect, it is similar to the TV-R model (3.4)-(3.5) which has one time-dependent shift.

Table 8). Similarly to the TV-R model, Tb does not predict returns after the estimated break in about 1995. Most estimates and t-values are not very affected by the restrictions. The main exceptions are that $\beta_{4,DY}$ is decreased to a value similar to that from the TV-R model, and that the estimated transition functions, G_1 and G_2, have their mid-points moved from about 2002 to about 1995.

Figure 3 contains plots of the two transition functions for the restricted model, as well as a plot of the model predictions. G_2 is similar to the transition function in the TV-R model, although the transition is a bit longer in duration. G_1, in contrast, displays a very rapid transition in January 1995. Macroeconomically, this is the time of the Mexican peso crisis that followed the December 1994 devaluation of the peso. The patterns are consistent with the result that the significance level for a linear regression of returns on lagged DY has decreased over most of the 1990s (Lettau and van Nieuwerburgh (2006)).

As mentioned in Section 2, Goyal and Welch (2003 and 2007) argue that most of the predictability in linear regressions comes from unique events during 1973–1975 (the oil crises). Paye and Timmermann (2006) similarly find an important break somewhere during the second half of 1974. Since it is conceivable that the results in this paper might have been affected by such a break, the third column of Table 12 presents parameter estimates after deleting the 36 observations from 1973–1975. Neither the levels nor the predictability conclusion (significant t-values) are affected much by this, so I conclude that the results are not due to unique events during this period.

Figure 4 contains a plot of model predictions from both the TV-R and the TV(2)-R models. They overlap almost completely, except for a period from about 1993 to 1997. During that period, the TV-R model favors a gradual shift, whereas the TV(2) model favors a more sudden one. The predicted return can be understood as the equity risk premium.[27] By and large, the predictions from both models adhere to some intuitive criteria of what the risk premium should look like: It is very persistent, seldom substantially negative, and seldom very high.[28] Furthermore, these predictions also resemble how many people, with the added benefit of hindsight, perceive that the risk premium on stocks might have varied over the last two decades: they mimic the bull market of the mid-1990s and the crash at around 2000. Interestingly, the predictions are negative from about 1999, about a year before the IT bubble burst. Furthermore, they mimic the new bull market from about 2003 onwards.

The results from both the TV-R model and the extended TV(2)-R model indicate that the sources of instability in the predictability of returns are mainly to be found in the mid-1990s. Although the purpose of this paper is not to model the risk premium per se, the results suggest that future research should look more closely at smooth transitions

[27] But this is not necessary. There might be additional variables that predict returns. In that case, the model here only predicts part of what is possible, and as a model of the risk premium it is then misspecified. Furthermore, investors might not be rational, and then predictability could also reflect that irrationality.

[28] The risk premium should not undergo large changes in the short run, and should therefore be persistent. Since one has to compensate investors for taking on risk, the risk premium must be nonnegative. And, according to the Equity Premium Puzzle literature, it should be low.

Table 11
Diagnostic Tests for the Estimated Models

		$T\hat{\sigma}_{\varepsilon}^{2}$	AIC	BIC	Skew.	Kurt.	JB, p-val	ARCH (1)	ARCH (4)	RNL (t)
Decile 1	Linear	3.2247	-2820.5	-2807.6	1.4476	10.13	**0**	0.135	0.528	
Decile 2	Linear	2.3693	-2990.0	-2977.1	0.7033	6.325	**0**	0.108	0.230	
Decile 3	Linear	2.0989	-3056.7	-3043.7	0.3958	5.691	**0**	0.070	0.220	
Decile 4	Linear	1.9346	-3101.5	-3088.6	0.3995	5.777	**0**	0.290	0.168	
Decile 5	Linear	1.8131	-3137.2	**-3124.3**	0.0940	5.147	**0**	0.298	0.346	
	TV-R	1.7418	**-3151.3**	-3121.1	-0.0218	4.963	**0**	0.125	0.114	0.365
Decile 6	Linear	1.7468	-3157.7	**-3144.7**	0.0346	4.793	**1e-16**	0.185	0.074	
	TV-R	1.6777	**-3171.9**	-3141.7	-0.0874	4.662	**1e-14**	0.074	**0.012**	0.295
Decile 7	Linear	1.6442	-3191.0	**-3178.0**	-0.0863	4.616	**7e-14**	0.122	0.058	
	TV-R	1.5784	**-3205.4**	-3175.2	-0.2248	4.590	**3e-14**	0.058	**0.020**	0.185
Decile 8	Linear	1.4477	-3261.0	**-3248.0**	-0.1896	4.227	**6e-9**	0.163	0.146	
	TV-R	1.3892	**-3275.6**	-3245.5	-0.3036	4.244	**3e-10**	0.108	0.063	0.062
Decile 9	Linear	1.2710	-3332.6	**-3319.6**	-0.1558	4.020	**2e-6**	0.156	0.066	
	TV-R	1.2215	**-3346.4**	-3316.2	-0.2694	4.041	**1e-7**	0.105	**0.046**	**0.040**
Decile 10	Linear	0.9318	-3503.3	-3490.4	-0.1249	3.641	**0.0044**	**1e-4**	**3e-6**	
	TV-R	0.8890	**-3521.2**	**-3491.0**	-0.2300	3.696	**3e-4**	**1e-4**	**1e-5**	**0.009**
Index	Linear	0.9728	-3479.6	-3466.7	-0.2100	3.568	**0.0033**	**0.004**	**2e-4**	
	TV-R	0.9287	**-3497.1**	**-3467.0**	-0.3160	3.658	**7e-5**	**0.003**	**9e-4**	**0.018**

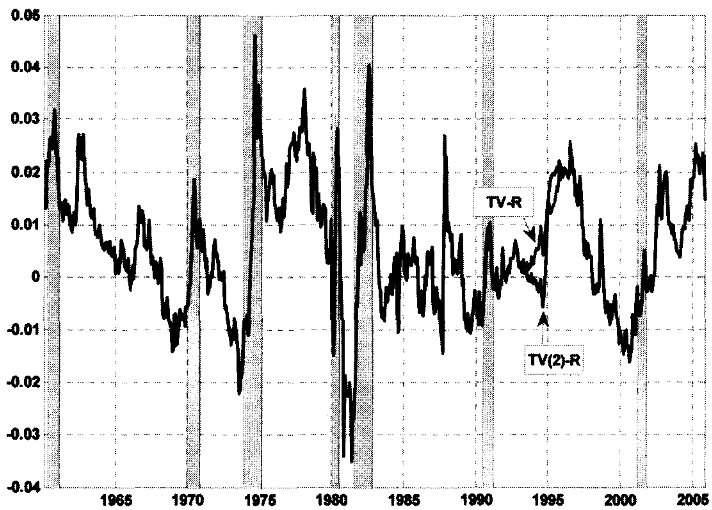
The TV-R model is estimated with the restriction $\beta_{2,Tb} = 0$, corresponding to the estimates in the second columns in Table 8.
"AIC" and "BIC" are the Akaike and Schwarz information criteria.
"JB, p-val" is the p-value for the Jarque-Bera statistic for normality.
"ARCH(1)" and "ARCH(4)" are p-values for tests for ARCH-type heteroscedasticity with lag lengths of 1 and 4, respectively (see Engle (1982)).
"RNL(t)" is a test for remaining nonlinearity with respect to t, and reported numbers are p-values (see Lundbergh, Teräsvirta and van Dijk (2003)).

Table 12
Parameter Estimates for the TV(2)-R Model for the Index

	Unrestricted	Restricted	Deleting 73-75
$\beta_{1,C}$	**-0.0465**	**-0.0369**	**-0.0322**
	(-3.73)	**(-3.11)**	**(-2.93)**
$\beta_{1,DY}$	**2.44**	**2.18**	**1.97**
	(5.21)	**(4.67)**	**(4.42)**
$\beta_{1,Tb}$	**-7.20**	**-7.06**	**-6.14**
	(-5.32)	**(-5.05)**	**(-4.35)**
$\beta_{2,C}$	-2.98	–	–
	(-0.41)	–	–
$\beta_{2,DY}$	287	–	–
	(0.40)	–	–
$\beta_{2,Tb}$	15.7	–	–
	(0.14)	–	–
$\beta_{3,C}$	0.0777	–	–
	(0.17)	–	–
$\beta_{3,DY}$	7.81	–	–
	(-0.46)	–	–
$\beta_{3,Tb}$	-242	–	–
	(-0.75)	–	–
$\beta_{4,C}$	-0.319	**-0.0641**	**-0.0642**
	(-0.79)	**(-2.22)**	**(-2.23)**
$\beta_{4,DY}$	18.9	**4.57**	**4.57**
	(0.85)	**(2.49)**	**(2.50)**
$\beta_{4,Tb}$	40.2	–	–
	(0.47)	–	–
γ_1	**143**	500	500
	(4.24)	(0.43)	(0.38)
c_1	**0.908 [9/2001]**	**0.760 [12/1994]**	**0.760 [12/1994]**
	(115)	**(141)**	**(134)**
γ_2	**26.6**	55.1	55.7
	(7.79)	(0.55)	(0.55)
c_2	**0.950 [8/2003]**	**0.771 [6/1995]**	**0.771 [6/1995]**
	(8.75)	**(44.6)**	**(44.6)**
AIC/BIC			

The full model is:
$R^e_{t+1} = \{[\beta_{1,C} + \beta_{1,DY}DY_t + \beta_{1,Tb}Tb_t](1 - G_{1t}) + [\beta_{2,C} + \beta_{2,DY}DY_t + \beta_{2,Tb}Tb_t]G_{1t}\}(1 - G_{2t}) + \{[\beta_{3,C} + \beta_{3,DY}DY_t + \beta_{3,Tb}Tb_t](1 - G_{1t}) + [\beta_{4,C} + \beta_{4,DY}DY_t + \beta_{4,Tb}Tb_t]G_{1t}\}G_{2t} + \varepsilon_{t+1}$; $\varepsilon_t \sim N(0,\sigma^2)$;
$G_{1t} = (1 + \exp\{-\gamma_1(t - c_1)\})^{-1}$, $G_{2t} = (1 + \exp\{-\gamma_2(t - c_2)\})^{-1}$,
where R^e_t is the decile excess return, DY_t is the divident yield, and Tb_t is the return on a one-month T-bill. G_{1t} and G_{2t} are the transition functions with argument t, and parameters γ_1 and c_1, and γ_2 and c_2.

Figure 4
Comparing Model Predictions from the TV-R and TV(2)-R Models



The predictions from the TV-R model are the same as in Figure 2, and the predictions from the TV(2)-R model are the same as in Figure 3.

to increase our understanding of the predictability of asset returns, and the equity risk premium.

5. Summary and Conclusions

Theoretical arguments and a rich empirical literature suggest that the dividend yield and the short interest rate predict the return on equity. A recent literature also argues that this relation has undergone one or several structural shifts. In this paper, I study whether such shifts can be explained by underlying economic variables, and whether models that allow for smooth gradual shifts, from one economic state to another, can explain the shifts better than an ordinary structural break model can.

In the first part of the study, I show that none of the macro or financial variables investigated can explain structural shifts. The conclusion is that economic phenomena, such as the business cycle, inflation, or the level of market risk, do not cause structural shifts in the predictive relation. Instead, I argue that these shifts are better explained by unique occurrences that extend over time. Evidence of at least one shift is mostly found in portfolios of large firms. The portfolios of small firms show almost no signs of shifts, possibly because the presence of more noise in these portfolios makes detecting shifts difficult.

In the second part, I estimate a structural shift to have occurred over a few years, from about 1993 to 1997. After the shift, returns are more sensitive to the dividend yield, but the short rate no longer predicts returns at all. It is noteworthy that if one allows for at least one structural shift, returns are predictable in-sample over the whole 1960–2005 period for the deciles of large firms. The timing of the shift is similar to that of Paye and Timmermann (2006), who find a structural break around the turn of 1994/1995, and that of Lettau and van Nieuwerburgh (2006), who find one between 1991 and 1992. To further study the dynamics of the shift, I estimate an alternative model that allows for an even more complex structural shift. The results turn out to be similar to the results from the previous structural shift model, in terms of the levels of the estimated parameters and the timing of the shift. However, an instantaneous break, such as those used in previous research, seems to be a rather poor approximation of the dynamics.

The results indicate that the 1990s had an unusual impact on return predictability, and they also suggest that future research concerning predictability and the equity risk premium should focus more closely on smooth transitions and not just structural breaks.

Appendix

A.1. Data.

A.1.1. *Macro Variables.* The macro variables are collected from the St. Louis Fed's public database FRED, and rates of change are calculated as yearly log differences, except where indicated.

- Inflation (the rate of change in the seasonally adjusted consumer price index, CPI; code: CPIAUCSL).

- The rate of change in seasonally adjusted industrial production (IP; code: INDPRO).
- The rate of change in seasonally adjusted monetary supply (M1; code: M1SL).
- The rate of change in the Composite Leading Indicator (CLI; code: USCYLEAD), published by The Conference Board.

CLI is a special case in this collection. It is a composite index that is believed to predict the business cycle, but the composition and the weights of the included variables have changed several times during its existence. This makes it problematic to use as a predictive variable. Since it is constructed to predict the business cycle as well as possible with an additional benefit of hindsight, I include it only to see if it may contain any interesting information.[29] To give CLI even more advantages, I include a different version of the rate of change for this variable. The version that The Conference Board itself uses is a 12-month moving average that is centered at 6 months before t:

$$\Delta_6 CLI_t = \left(\left[\frac{CLI_t \times 12}{\sum_{i=1}^{12} CLI_{t-i}} \right]^{12/6.5} - 1 \right) \times 100. \tag{A.1}$$

It is noteworthy that, even with all the advantages, CLI turns out to be one of the variables with least predictive ability.

A.1.2. *Financial Variables.* The financial variables are collected from CRSP, except where noted.

- The one-month realized return on the CRSP value-weighted index for NYSE/AMEX/NASDAQ (code: 1000080) and the ten decile portfolios ordered by size (codes: 1000082, ..., 1000091). Excess returns are then calculated as the difference between returns on the portfolios and the return on a one-month T-bill (from the CRSP, Fama risk-free rates).
- Realized volatility is the sum of squared daily returns on the index over one-month. The variable used is the log of volatility. This is to make the distribution of the variable better from a statistical perspective, as it comes closer to a normal distribution.
- The term premium (TP) is calculated as the difference in yields between a five-year bond (from the Fama-Bliss file) and the one-month T-bill.
- T-bill (Tb) is the yield on a one-month T-bill.
- The dividend yield (DY) is calculated as the log difference between dividends and the level of the index. Dividends are calculated as the difference between the index including and excluding dividends, and the value used for a certain month is the sum of all dividends during the previous twelve months.
- The default premium (DP) is calculated as the difference in yields between AAA (code: AAA) and BAA (code: BAA) rated bonds, collected from the St. Louis Fed.

[29] Perez-Quiros and Timmermann (2000) use a "real-time" version of CLI, i.e., the value that was actually accessible at time t, so their results do not have the benefit of hindsight.

- Earnings-price (EP) is calculated as the log difference between earnings and the level of the index. Earnings data are collected from Robert Shiller's website

  `www.econ.yale.edu/~shiller/data.htm`.

A.1.3. *Timing Conventions.* All variables are sampled at the monthly frequency, and the extreme observation of October 1987 has been deleted. For the portfolio returns this means, for instance, that the return from January 1 to February 1 is considered to be known on February 1. Prices (for calculation of yields and financial ratios) are considered known when they are listed, i.e., on the same day. The values of the macro variables often refer to a certain month. For instance, January inflation refers to the rate of inflation in January. This variable is considered to be known on February 1. This is the earliest date an agent could possibly know this value by inferring it from prices in the real economy. The statistic is not published on that date. On the contrary, many macro statistics are published with a lag and are also sometimes updated one or several times. The use of the earliest date could mean, for instance, that actual inflation in January (finally completely realized on January 31) presumptively affects excess returns. I do not theorize that agents typically have to obtain the statistic from a publication.

A.2. Procedure for Finding p-Values for Structural Shifts. Luukkonen, Saikkonen and Teräsvirta (1988) develop a test for $H_0 : \gamma = 0$ against $H_1 : \gamma > 0$ in (3.4)-(3.5), or in (3.6)-(3.7). Under H_0 there is no nonlinearity, no structural shift, explained by the suggested variable. In (3.5), t is the suggested variable, whereas in (3.7) it is z_t.

To circumvent the problem that (3.4)-(3.5) and (3.6)-(3.7) are not identified under H_0, let $\overline{\mathbf{X}}_t$ be a vector of explanatory variables, not containing a constant, and y_{t+1} the independent variable. Define

$$\widetilde{X}_t = \begin{bmatrix} 1 & \overline{\mathbf{X}}_t & z_t \times \overline{\mathbf{X}}_t \end{bmatrix}',$$

and project y_{t+1} on $\widetilde{X}_t$ to get the parameter estimates $\begin{bmatrix} \widehat{\beta}_0 & \widehat{\boldsymbol{\beta}}_1 & \widehat{\boldsymbol{\beta}}_2 \end{bmatrix}$. Then test

$$H_0' : \boldsymbol{\beta}_2 = \mathbf{0} \tag{A.2}$$

instead of H_0. I use an LM test with p-values calculated according to White (1980), or Newey and West (1987).

Lundbergh, Teräsvirta and van Dijk (2003) develop a similar test for $H_0 : \gamma_1 = \gamma_2 = 0$ in (3.8)-(3.10). Under H_0, there is no nonlinearity in either t or in z_t. Again, the model is not identified under the null. Let $\overline{\mathbf{X}}_t$, z_t, and y_{t+1} be as before, and let t be time. Define

$$\widetilde{W}_t = \begin{bmatrix} 1 & \overline{\mathbf{X}}_t & z_t \times \overline{\mathbf{X}}_t & t & t \times \overline{\mathbf{X}}_t & z_t \times t \times \overline{\mathbf{X}}_t \end{bmatrix}',$$

and regress y_{t+1} on $\widetilde{W}_t$ to get the estimates $\begin{bmatrix} \widehat{\beta}_0 & \widehat{\boldsymbol{\beta}}_1 & \widehat{\boldsymbol{\beta}}_2 & \widehat{\beta}_3 & \widehat{\boldsymbol{\beta}}_4 & \widehat{\boldsymbol{\beta}}_5 \end{bmatrix}$. Then test (with a slight abuse of notation)

$$H_0' : \boldsymbol{\beta}_2 = \beta_3 = \boldsymbol{\beta}_4 = \boldsymbol{\beta}_5 = \mathbf{0}. \tag{A.3}$$

If we reject H_0', and consequently H_0, this might be because of nonlinearity w.r.t. t, or z_t, or both. We then want to test if we can do without the t dimension (given, then, that there is nonlinearity in z_t) with the sub-test

$$H_{01}' : \beta_3 = \boldsymbol{\beta}_4 = \boldsymbol{\beta}_5 = 0, \tag{A.4}$$

and if we can do without the z_t dimension (given that there is nonlinearity w.r.t. t) with

$$H'_{02} : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_5 = 0. \tag{A.5}$$

For details, discussions, and other versions of some of the test procedures, see the cited literature.

References

Ang, A., and G. Bekaert, 2007, "Stock Return Predictability: Is it There?," *Review of Financial Studies*, 20, 651–707.

Bossaerts, P., and P. Hillion, 1999, "Implementing statistical criteria to select return forecasting models: what do we learn?," *Review of Financial Studies*, 12, 405–428.

Campbell, J. Y., and John H. Cochrane, 1999, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205–251.

Campbell, J. Y., 1987, "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18, 373–399.

Campbell, J. Y., and R. J. Shiller, 1988, "The dividend-price ratio and expectations of future dividends and discount factors," *Review of Financial Studies*, 1, 195–228.

Campbell, J. Y., and S. Thompson, 2005, "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?," NBER Working Paper No. 11468.

Cochrane, J. H., 1992, "Explaining the variance of price-dividend ratios," *Review of Financial Studies*, 5, 243–280.

Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, Princeton, NJ.

Cochrane, J. H., 2006, "The Dog That Did Not Bark: A Defense of Return Predictability," NBER Working Paper No. 12026.

Cornell, B., 1999, *The Equity Risk Premium: The Long-Run Future of the Stock Market.* John Wiley & Sons, New York, NY.

Engle, R., 1982, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Fama, E. F., and G. W. Schwert, 1977, "Asset Returns and Inflation," *Journal of Financial Economics*, 5, 115–146.

Fama, E. F., and K. R. French, 1988, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–27.

French, K., W. Schwert, and R. Stambaugh, 1987, "Expected Stock Returns and Variance," *Journal of Financial Economics*, 17, 5–26.

Ghysels, E., P. Santa-Clara, and R. Valkanov, 2005, "There is a Risk-Return Tradeoff After All," *Journal of Financial Economics*, 76, 509–548.

Glosten, L., R. Jagannathan, and D. Runkle, 1993, "On the Relation between the Expected Value and the Variance of the Nominal Excess Return on Stocks," *Journal of Finance*, 48(5), 1779–1801.

Gordon, M., 1962, *The Investment, Financing, and Valuation of the Corporation*, Homewood, Ill.: Richard D. Irwin.

Goyal, A., and I. Welch, 2003, "Predicting the Equity Premium with Dividend Ratios," *Management Science*, 49(5), 639–654.

Goyal, A., and I. Welch, 2007, "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," forthcoming in *Review of Financial Studies.*

Hamilton, J. D., 1989, "A New Approach to the Analysis of Nonstationary Returns and the Business Cycle," *Econometrica*, 57, 357–384.

Hodrick, R. J., 1992, "Dividend yields and expected stock returns: alternative procedures for inference and measurement," *Review of Financial Studies*, 5, 357–386.

Lamont, O., 1998, "Earnings and Expected Returns," *Journal of Finance*, 53(5), 1563–1587.

Lettau, M., and S. C. Ludvigson, 2001, "Consumption, Aggregate Wealth and Expected Stock Returns," *Journal of Finance*, 56(3), 815–849.

Lettau, M., and S. van Nieuwerburgh, 2006, "Reconciling the Return Predictability Evidence," NBER Working Paper No. W12109.

Lundbergh, S., T. Teräsvirta, and D. van Dijk, 2003, "Time-varying smooth transition autoregressive models," *Journal of Business & Economic Statistics*, 21, 104–121.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta, 1988, "Testing Linearity Against Smooth Transition Autoregressive Models," *Biometrika*, 75, 491–499.

Nelson, D., 1991, "Conditional Heteroscedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Newey, W. K., and K. D. West, 1987, "A Simple, Positive Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Pastor, L., and R. F. Stambaugh, 2001, "The Equity Premium and Structural Breaks," *Journal of Finance*, 56(4), 1207–1239.

Paye, B. S., and A. Timmermann, 2006, "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13, 274–315.

Perez-Quiros, G., and A. Timmermann, 2000, "Firm Size and Cyclical Variations in Stock Returns," *Journal of Finance*, 55(3), 1229–1262.

Pesaran, M. H., and A. Timmermann, 1995, "Predictability of stock returns: Robustness and economic significance," *Journal of Finance*, 50(4), 1201–1228

Poterba, J., and L. Summers, 1988, "The Persistence of Volatility and Stock Market Fluctuations," *American Economic Review*, 75, 1142–1151.

Rozeff, M. S., 1984, "Dividend Yields are Equity Risk Premiums," *Journal of Portfolio Management*, 49, 141–160.

Shiller, R. J., 1989, *Market Volatility*, MIT Press.

Shiller, R. J., and A. E. Beltratti, 1992, "Stock prices and bond yields: Can their comovements be explained in terms of present value models?," *Journal of Monetary Economics*, 30, 25–46.

Stock, J. H., and M. W. Watson, 1996, "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, 14(1), 11–30.

Teräsvirta, T., 1994, "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89(425, 208–218.

Viceira, L., 1996, "Testing For Structural Change in Predictability of Asset Returns," manuscript, Harvard University.

White, H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(4), 817–838.

Restoring the Predictability of Equity Returns

ABSTRACT. Several recent papers have questioned the predictability of equity returns. With return data including the 1990s, variables that previously predicted returns, such as the dividend yield, are no longer significant and results of out-of-sample tests are often weak. A possible reason is that the underlying structure of the economy has changed.

I use an econometric model (the TV-STR model) that allows for regime shifts over time as well as due to changes in a state variable, in this case the price-earnings ratio. This makes it possible to separate influences from these two sources and to determine whether one or both sources have affected return predictability. The results indicate that, first, a structural change occurred during the 1990s, and, second, that the unusually high level of price earnings in the late 1990s and early 2000s temporarily affected predictability. Accounting for these factors, equity return is still predictable.

1. Introduction

The early financial literature often assumed that equity returns are inherently unpredictable, and the related empirical literature also seemed to support this assumption. Starting in the mid-1980s, however, an enormous flood of papers began to find that several variables actually do seem to predict returns, and the conventional wisdom is now that returns are predictable.[1] This flood of papers has recently met two types of barriers. The first one is that there are statistical problems surrounding the models used to predict returns, and that the claims of predictability have therefore been strongly exaggerated.[2] These statistical problems include small-sample biases, data-mining problems, and unusual distributions for near-nonstationary variables. The other barrier consists of the finding that, while there might be evidence of in-sample predictability, there is little evidence of out-of-sample predictability (Bossaerts and Hillion (1999) and Goyal and Welch (2003 and 2007)). Perhaps even more troublesome for the predictability literature is the fact that even ordinary linear regressions, such as the ones used by many of the early researchers, fail to predict returns in-sample when data from the 1990s are included. This could indicate that the predictability of equity returns was just a statistical fluke or a misleading data-mining result.

[1] Some of the most important papers are those by Rozeff (1984), Fama and French (1988), Campbell & Shiller (1988), Shiller (1989), Cochrane (1992), Hodrick (1992), Lamont (1998), Lettau and Ludvigson (2001), and Lewellen (2004). See also Campbell, Lo, and MacKinlay (1997) for an overview.

[2] Among others, Goetzman and Jorion (1993), Nelson and Kim (1993), Stambaugh (1999), Ang and Bekaert (2006), Ferson, Sarkissian & Simin (2003), and Valkanov (2003).

This criticism poses a serious challenge to both academic researchers and to investors. The risk premium is a cornerstone of most asset pricing models, and knowledge of its size and possible time-variation is crucial to investors. Several researchers have also introduced asset pricing models that build on predictability of returns as a stylized fact (Campbell and Viceira (1999), Barberis (2000), and Ang and Liu (2004) among others). If there is no predictability, this literature is built on a vacuous claim.

A few researchers have recently attempted to save the predictability literature by introducing structural breaks into the empirical models (Lettau and van Nieuwerburgh (2006), and Paye and Timmermann (2006)). Their statistical tests show that the equity return process seems to have undergone a break somewhere during the 1990s.[3] When Lettau and van Nieuwerburgh "correct" for the break, then returns are again predictable in-sample. While this type of approach marks a significant breakthrough in the predictability literature, it still leaves a few major questions unanswered. Most seriously, the detected breakpoint coincides more or less with the IT bubble, i.e., the period around the year 2000. During this period, the equity market behaved in a very unusual way, even irrationally according to some (most notably Shiller (2000)). It is not clear, then, if a test for a structural break discovers something about the underlying economy, or if it just rediscovers the IT bubble period. However, depending on which of these explanations we accept, our conclusions for investing might be radically different. If the economy has entered into a new state, as a structural break would suggest, then investors should expect future returns to be different from past returns. But, on the other hand, if the economy has just passed through an unusual period, then investors should expect future returns to be little or no different from past returns.

This paper presents a way to disentangle two sources of influence from each other, and to test if they, alone or together, have any significant impact on returns. One source of influence is assumed to depend on time. We can think of this as a change in the underlying economy that has permanently changed the equity return process, such as a change in regulations or the development of new technology. The model also allows for the transition over time to be smooth, which is a more realistic kind of change than a break is. The second source of influence is assumed to depend on the price-earnings ratio. Shiller (2000) argues that the high price-earnings ratio at the end of the 1990s was a sign that the market was irrationally overvalued. Others have argued that it might have been a sign of unusually high optimism about future growth. In either case, it is a natural candidate to proxy for the unusualness of this period. The model can then distinguish between a change that depends on the IT bubble period and a change that depends on the fundamental economy.

I show that the conclusions regarding the IT bubble period depend largely on which investment horizon–12, 36, or 60 months–we are looking at. For the shortest horizon, it appears that predictability disappeared during that period. This is consistent with the

[3] This is only one of the breaks these authors find. Lettau and van Nieuwerburgh, for instance, also find evidence of a break in the mid 1950s, and Paye and Timmermann test several predictors of returns and find different break points in many of them.

intuition that returns would be less correlated with fundamentals if prices were temporarily irrational. However, we can also find evidence of dividend growth turning predictable during this period, possibly at the expense of return predictability. For the longer horizons, the IT bubble period appears to have had no particular impact on predictability at all. The results regarding changes over time are similar for all horizons. I reject the occurrence of a structural break; however a smooth change appears to have taken place over several years from the mid- to late 1990s.

The rest of this paper is organized as follows. Section 2 gives an overview of the literature and a background to the problem of return predictability. Section 3 briefly describes the data used in the empirical parts. Two different underlying arguments for the change in return predictability during the 1990s are explored in the two subsequent sections: in Section 4, that there has been a structural change in the economy, and, in Section 5, that the IT bubble period has affected predictability. Section 6 presents the econometric framework, and Section 7 discusses the statistical estimates. The main discussion of interpretations and evaluations of the underlying arguments is given in Section 8, and Section 9 sums up and concludes.

2. Background and Literature

Although practitioners had used linear regressions to predict equity returns for a long time, the academic predictability literature took off in the mid-1980s with papers by, among others, Rozeff (1984), Fama and French (1988), and Campbell and Shiller (1988). The new papers showed that a simple linear regression of returns on one or a few explanatory factors produced significant estimates. Campbell and Shiller (1988) elaborate on Gordon's (1962) growth model, and thereby show that the dividend yield is a theoretically sound factor to use for predicting equity returns. Fama and French (1988) also use the dividend yield, and show that predictability increases with horizon length. The R^2 is, for instance, higher for the five-year horizon than for the one-year horizon. Building on these and other results, many papers report that returns are predictable by variables such as interest rates, financial ratios, and macro variables (Cochrane (1992), Hodrick (1992), Lamont (1998), and Lettau and Ludvigson (2001), among others). While the set of proposed variables is now large, the main variable in this paper is the dividend yield, as this remains one of the most popular choices in the literature.

Over the years, however, a counter literature has evolved. Several papers in this literature focus on statistical problems. Goetzman and Jorion (1993), Nelson and Kim (1993), and Stambaugh (1999) argue that the dividend yield (and several other financial ratios) is "near nonstationary." This refers to the fact that a unit root is not rejected in a test for nonstationarity, but neither is a root very close to unity. And a unit root can, arguably, be rejected on theoretical grounds. This feature, they argue, makes the use of a linear regression complicated, and t-statistics should be adjusted downwards. After such an adjustment, the dividend yield no longer significantly predicts returns. On the other hand, Lewellen (2004) argues for another correction of t-statistics, and in his paper the dividend yield does, again, significantly predict returns.

The counter literature also includes papers by Bossaerts and Hillion (1999), and Goyal and Welch (2003 and 2007). Their claim is very straightforward: Out-of-sample there is no predictability. Suppose, for example, that an investor wants to take advantage of return predictability, and that he uses a linear model to predict returns for the next period. Goyal and Welch (2003), in particular, argue that such an investor would do no better (and often worse) than someone who has only used the historical average return to predict returns. Since out-of-sample predictability could be viewed as the acid test for predictability, this is bad news for the predictability literature.

However, there is a defense of predictability that tries to circumvent the above-mentioned problems altogether. Remember that, according to Gordon's growth model, the price of an asset, P_t, is a function of the required return, r, next period's dividend, D_{t+1}, and dividend growth, g, such that

$$P_t = \frac{D_{t+1}}{r-g}. \tag{2.1}$$

Solving for the dividend-price ratio, we get:

$$\frac{D_{t+1}}{P_t} = r - g. \tag{2.2}$$

In this model, r and g are assumed to be constants, and, consequently, the dividend-price ratio is also constant. Campbell and Shiller (1988) derive a "dynamic version" of this expression. Their result can be expressed as (see Equation (6) of their paper)

$$\log\left(\frac{D_t}{P_t}\right) \approx E_t\left[\sum_{j=0}^{\infty} \rho^j \left(\widetilde{r}_{t+j} - \Delta d_{t+j}\right)\right] + const, \tag{2.3}$$

where ρ is a constant, $\widetilde{r}$ is log return, and Δd $\left(= \log\left(\frac{D_{t+j}}{D_{t+j-1}}\right)\right)$ is dividend growth, i.e., it corresponds to the constant g in (2.2). Here, a high dividend-price ratio must be followed by either higher future returns or lower future growth, possibly both. Consequently, the dividend-price ratio must predict at least one of them. Campbell and Shiller also find no indication of growth being predictable. And if growth is not predictable, the dividend yield *must* predict returns, regardless of any statistical problems with near-nonstationary variables or out-of-sample predictability. Cochrane (1998 and, particularly, 2006) argues that this is a strong argument in favor of the dividend yield being able to predict returns. Unfortunately for this argument, Ang and Bekaert (2007) come to the opposite conclusion regarding the predictability of dividend growth: In their analysis, it is predictable in some settings.[4] It would appear, then, that we might be back to showing that returns are predictable by actually getting significant regression parameter estimates.

One further fact that has been difficult for the predictability literature to handle is that the degree of significance of coefficient estimates in predictive regressions has been decreasing over time since the works of Campbell and Shiller, and Fama and French, even

[4] Although the predictability of growth is not robust to the choice of sample period. In a univariate regression of growth on dividend yields, Ang and Bekaert obtain significant parameter estimates for the sample period 1952–2001 when the horizon is 1 or 4 quarters, but not when the horizon is 20 quarters or the sample period is 1935–2001.

using the same econometric setups. Starting somewhere in the 1990s, the dividend yield no longer significantly predict returns in a linear regression (Lettau and van Nieuwerburgh (2006)). There could be several reasons for this. Those who do not believe in return predictability could argue that this shows the predictability literature to be based on a statistical accident. Using pre-1990 data produced significant results, but that was a type I error. With more data, the error was undone, and we now, correctly, no longer reject the null of no predictability. The decreasing significance levels during the 1990s are, of course, also related to the out-of-sample results of Bossaerts and Hillion (1999), and Goyal and Welch (2003 and 2007). Their out-of-sample periods occur largely during the 1990s and as this period changes the in-sample results it also changes the out-of-sample results.

However, the change in predictability may have a different explanation. It could be that the equity return process itself has undergone some sort of change during this period. If this is the case, then a simple linear regression might produce nonsignificant results even if returns are predictable. This type of explanation is put forth by Lettau and van Nieuwerburgh (2006) and Paye and Timmermann (2006). The new feature in these papers is that they employ models that allow for structural breaks. For instance, Lettau and van Nieuwerburgh allow the dividend-price ratio to have different means before and after a breakpoint, and then estimate the break to have occurred in 1991/1992. Using a dividend-price ratio that has been corrected for the break, returns are in-sample predictable again. Similarly, Paye and Timmermann find a break at the end of 1994.[5] In a slightly different setting, Pastor and Stambaugh (2001) use a Bayesian model, together with several different assumptions about risk preferences, transition periods, etc., to study the equity risk premium. Their estimates indicate that the risk premium has gone through several different regimes (i.e., levels), but the timings of the breakpoints are quite dependent on which underlying assumptions they use.

The approach taken in these papers constitutes a major development in the predictability literature but, as argued in the introduction, it does suffer from a few important drawbacks. I will briefly present the data used in this paper and then discuss the problems with the structural break approach in the following sections.

3. Data Description

The data I use in this paper are U.S. stock index prices, dividends, and earnings, together with secondary market three month T-bill rates. All input data are on a monthly base and log excess returns are calculated from prices, dividends and T-bill rates for horizons of 12, 36, and 60 months. The log dividend-price ratio is denoted DP_t and the log price-earnings ratio PE_t. I also use two alternatives to PE_t, where earnings are averaged over 12 and 60 months, respectively; these variables are denoted PE_t^{12} and PE_t^{60}. The variables are further described in the appendix.

The sample period starts in January 1955 and ends in June 2004. Lettau and van Nieuwerburgh (2006) identify a structural break in the relation between the dividend-price

[5] Both Lettau & van Nieuwerburgh and Paye & Timmermann test for several different explanatory variables. The results reported here are for the dividend yield.

ratio and stock returns in both 1955 and in 1994 (when testing for two breaks). Since the model I use here is not well suited for more than one break in the time dimension, I restrict the period and start after the first break.

DP and the three versions of PE are plotted in Figure 1. As can readily be seen from the graphs, DP and PE are very (negatively) correlated. There is also quite a noticeable trend in DP; it is not obvious if it started as early as in the mid-1980s, but between 1995 and 2001 the decline in the variable is steep. After 2001 the trend appears to either have reversed or, at least, evened out. The trend in DP over the 1990s is probably responsible for many of the predictability literature's statistical problems over the last decade.

Summary statistics for the variables are presented in Tables 1a and 1b. Excess returns for the period are between 4.49% and 4.82%, depending on the horizon. As expected, the standard deviation for the 12-month horizon is the highest at 15%, falling to about 6% for the 60-month horizon. The two shorter horizons are negatively skewed, and the Jarque-Bera test (JB) rejects normality for them. The longest horizon has a slight positive skewness and normality is not rejected. All the PE variables are highly correlated, and DP is highly negatively correlated with them all. It has been common since Lintner (1956) to assume that firms have a target dividend that is a constant fraction of earnings, so the high empirical correlation between PE and DP makes a great deal of sense.

4. Has There Been a Structural Break in Time?

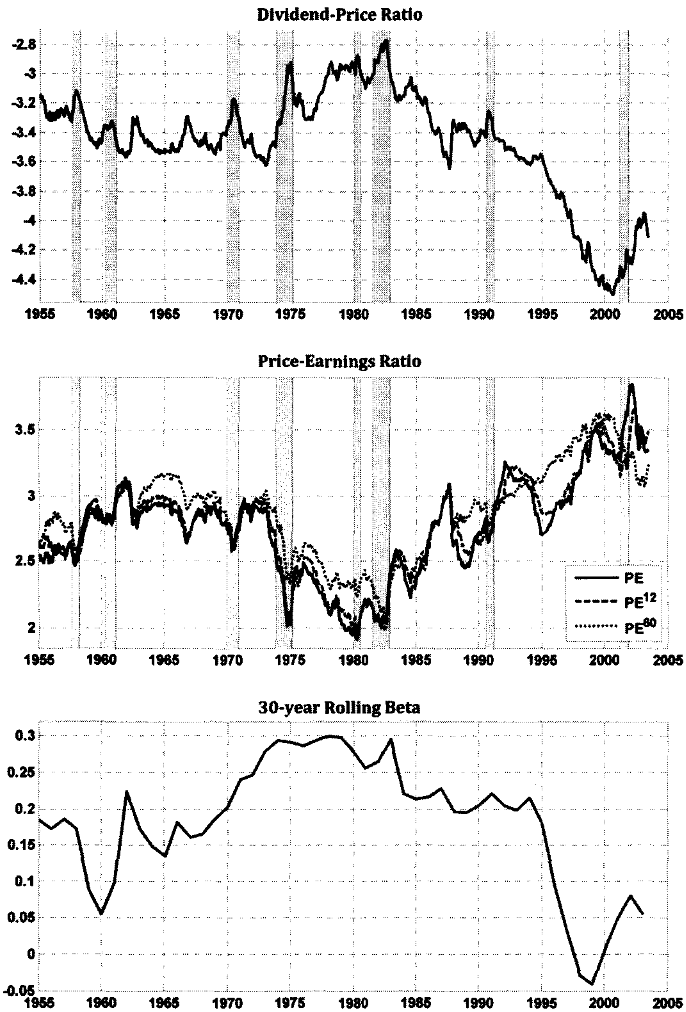
It is quite easy to find explanations for how the underlying economy has changed in a way that might have influenced the relation between the dividend yield and returns. For instance the 1980s and 1990s saw a gradual change in tax regulations (McGrattan and Prescott 2005), and in the 1990s many companies started to reward shareholders through share buy-backs at the expense of smaller dividend payouts. Other underlying changes include increased stock market participation, particularly during the 1990s, increased risk sharing through the development of new financial instruments, and lower and more stable inflation. Another trend that might be important here is that over the last decade more firms have entered the stock market when they are young and have very low (or even negative) earnings and pay very low (if any) dividends. As a consequence, the dividend-price ratio would have trended downwards. Furthermore, the development of better computers, the growth of the internet, and a generally more internationalized world might have helped to make transaction costs lower.

Let us briefly examine some empirical evidence. The most straightforward way to assess if DP_t is able to predict stock returns is to use a linear regression such as

$$r^e_{t+k} = \beta_0 + \beta_1 DP_t + \varepsilon_{t+k}, \tag{4.1}$$

where, r^e_{t+k} is the excess return between time t and $t+k$, and k is the horizon. If $\widehat{\beta}_1$ turns out to be significantly different from zero, we would interpret that as evidence that DP_t predicts stock returns. This is what Fama and French (1988) do. The first panel of Table 2 reports the results of these estimates for the whole sample period, 1955–2004. The t-values on DP_t lie between 1.45 and 1.57, which is insignificant at any standard level. However, the value of the t-statistic, as already mentioned, is very dependent on

Figure 1
Dividend-Price Ratio, Price-Earnings Ratio, and Rolling Beta Estimates, 1955—2004



The upper graph shows $\log(D_t/P_t)$, and the middle graph shows three versions of the price-earnings ratio. PE is $\log(P_t/E_t)$, PE^{12} uses the average of E_t over 12 months in the denominator, and PE^{60} uses the average over 60 months.
The lower graph shows estimates of β_1 from $r^e_{t+1} = \beta_0 + \beta_1 DP_t + \varepsilon_{t+1}$, using a 30-year rolling window with yearly data.
The shaded areas correspond to NBER recession periods.

Table 1a
Summary Statistics

	r^e_{t+12}	r^e_{t+36}	r^e_{t+60}	DP	PE	PE^{12}	PE^{60}
mean	0.0461	0.0449	0.0482	-3.46	2.76	2.78	2.88
min	-0.486	-0.198	-0.0876	-4.50	1.92	1.95	2.08
max	0.373	0.228	0.195	-2.77	3.84	3.66	3.63
st.dev	0.150	0.0792	0.0591	0.374	0.388	0.375	0.341
skewness	-0.571	-0.404	0.0678	-1.00	0.0516	-0.0948	-0.0279
kurtosis	3.026	3.52	2.673	3.77	2.87	2.61	2.60
JB	**31.6**	**21.4**	2.78	**112**	0.68	4.53	4.00
	(2.7e-4)	**(9.3e-4)**	(0.22)	**(0)**	(0.70)	(0.10)	(0.13)

Summary statistics for the basic variables. JB is the Jarque-Bera statistic for testing normality with p-values in parentheses. All variables are in logs (see the data appendix).

Table 1b
Correlations of Explanatory Variables

	PE	PE^{12}	PE^{60}
DP	-0.895	-0.903	-0.935
PE		0.979	0.898
PE^{12}			0.927

Table 2
Parameter Estimates for the Linear Model

	1955—2004					
horizon	12 months		36 months		60 months	
#obs	582		558		534	
	beta	t-value	Beta	t-value	beta	t-value
C	0.333	1.79	0.287	1.80	**0.225**	**2.10**
DP_t	0.083	1.51	0.071	1.45	0.052	1.57
	1955—1991					
horizon	12 months		36 months		60 months	
#obs	582		558		534	
	beta	t-value	Beta	t-value	beta	t-value
C	**0.848**	**2.95**	**0.421**	**2.06**	**0.392**	**2.64**
DP_t	**0.244**	**2.82**	0.115	1.85	**0.106**	**2.32**
	1955—2004					
horizon	12 months		36 months		60 months	
#obs	582		558		534	
	beta	t-value	Beta	t-value	beta	t-value
C	**0.413**	**2.23**	**0.340**	**2.14**	**0.268**	**2.41**
DP_t	**0.119**	**2.12**	**0.0958**	**2.00**	**0.0751**	**2.20**
t/T	0.0886	1.45	0.0673	1.19	0.0693	1.31

This table presents parameter estimates for the linear model (4.1) for horizons of 12, 36, and 60 months. The first panel uses the full sample 1955—2004, and the second panel uses the shorter period 1955—1991. The third panel adds a time trend (t/T) as an additional variable, and uses the full sample 1955—2004.
T-values are calculated according to Newey and West (1987) and use a lag length equal to the horizon length. All variables are in logs (see the data appendix).

the sample period we use. For example, restricting the period to be between 1955 and 1991, the t-values become 2.82, 1.85, and 2.32 (see the second panel of Table 2). Lettau and Ludvigson (2001), Goyal and Welch (2003), and Ang and Bekaert (2007) also show that including of excluding the 1990's in the estimation gives very different significance levels.

Lettau and van Nieuwerburgh (2006) also estimate time-varying beta values, corresponding to β_1 in (4.1), using a 30-year rolling window. The last panel of Figure 1 shows the results from doing the same with the data used here. The estimated beta is reasonably stable until 1994, except for a few years around 1960, varying between 0.15 and 0.30. After that, the value drops dramatically and even becomes negative for a short period. Lettau and van Nieuwerburgh use a test developed by Perron (1989) to find breakpoints in the relation between the dividend-price ratio and stock returns, and find a break between 1991 and 1992. Similarly, Paye and Timmermann (2006) find a breakpoint in a regression of returns on dividend yields in December 1994.

Another simple way to find indications of a structural change over time is to include a time trend variable (t/T) in the linear regression. Controlling for a time trend makes $\widehat{\beta}_1$ significant over the entire sample period, as can be seen in the last panel of Table 2.

Consequently, the evidence suggests that the regression in (4.1) is not stable over time, and that particularly the 1990s seem to cause the problems. However, this is not enough to show that any of the changes to the underlying economy mentioned above has actually caused the change in predictability. One of the major phenomena during the 1990s and early 2000s was the "IT bubble." When we find evidence of a structural break that singles out this period as unusual in some way, we cannot simply conclude that the underlying economy has changed. The results might be driven solely by the IT bubble. I will discuss this further in the next section.

It is also somewhat counter-intuitive to model any of the events listed above as a break. They are all processes that have taken time to establish themselves, and a structural break seems to be quite a brutal approximation of reality. I will explain how to avoid this problem in Section 6, and model it as a trend instead.

5. The IT Bubble

Alan Greenspan held his famous "irrational exuberance" speech, in which he expressed his view that the stock market was dangerously overvalued, in December 1996. One indicator of overvaluation was the unusually high value of the price-earnings ratio (the PE ratio). With hindsight, the PE level at the end of 1996 does not seem alarmingly high, but it continued to rise and reached record levels toward the end of the 1990s. The PE ratio can be interpreted as the price investors are prepared to pay for a certain amount of earnings, and Shiller (2000) notes that historically this ratio has kept itself within rather tight bounds (between, say, 2.0 and 3.2 over the last fifty years; see the second panel of Figure 1, particularly PE^{60}). The major exceptions are the late 1920s (not shown), which were followed by the infamous crash of 1929, and the late 1990s, which, in turn, was later followed by the crash of the early 2000s. The "IT bubble" refers to this latter period, which started somewhere between 1996 and 1998, and ended a few years into the new millennium. In recent years, this has been one of the hottest topics in the popular financial literature.

We can get a crude picture of how the PE ratio relates to returns by, again, looking at Gordon's growth model. If we assume that dividends are a constant fraction, b, of earnings in (2.1), we can express the PE ratio as

$$\frac{P_t}{E_{t+1}} = \frac{b}{r-g}. \tag{5.1}$$

With prices set according to (2.1), a high PE ratio signals a higher constant payout ratio, b, a lower required return, r, or higher growth, g.

A popular interpretation of the PE ratio is that it is primarily a signal related to the business cycle (see, for instance, Lamont (1998)). If, for example, the economy heads into a recession, we would expect growth, g, over the immediate future to be lower than usual. In panel two of Figure 1, it is also quite evident that the PE ratio often drops at the beginning of a recession and rises toward the end of it. In line with this, the high PE ratio at the end of the 1990s might have been caused by shocks to g. The rise and subsequent fall of the market is then interpreted as unusually high optimism about future growth (that did not materialize), followed by a return to a more ordinary level.

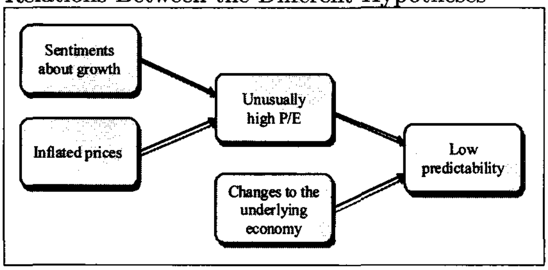
Another popular explanation for the high PE ratio during this period is that it was evidence of irrationally overvalued markets. Shiller (2000) is one of the principal proponents of this interpretation. According to this view, the high PE ratio might partly have signaled changes in r and/or in g, as above, but in addition there was a further increase in prices that was not due to changes in fundamentals. In other words, Shiller argues that the IT bubble was a financial bubble.

Regardless of whether prices were rational or not, such an unusual period as the IT bubble would likely make statistical inference problematic. More specifically, during the 1990s the economy might have undergone a structural change as argued in Section 4. But even if this was not so, a standard econometric test for break points is nonetheless likely to identify a change because of the unusual, but temporary, period at the end of the 1990s.

We thus have two different potential explanations for why return predictability has been low since sometime in the 1990s, two explanations that could possibly be true simultaneously. First, it might be because of a structural change in the underlying economy, and, second, because of the IT bubble, interpreted as an unusually high PE ratio. Since these two events occurred during almost the same period, it is difficult to see which one, if either, could be the underlying reason for the change in return predictability.

In the following, I will try to separate the two by modeling them as different regimes. A structural change over time, t, as in Section 4 (i.e., corresponding to a change in the underlying economy), is modeled as two different regimes with a transition period between them. For some period of time, the economy is in the first state. Then it starts to shift, and finally reaches a second state, where it remains permanently. The timing of the shift and the speed of the transition (where a break is a special case) are measured from the data. Furthermore, a change related to the PE ratio, as in Section 5, is also modeled as two different regimes with a transition period. When the PE ratio is low the economy is in one state, and when it is high it is in another state. The speed and location of the transition are, as with the t dimension, measured from the data, and the states of the PE dimension are independent from the states of the t dimension.

Figure 2
Relations Between the Different Hypotheses



Note that regarding the IT bubble period, we also have two underlying explanations that could both be true simultaneously. First, sentiments about growth could have caused the IT bubble, and, second, irrationally inflated prices could have caused it. I will return to the question of if and how we can distinguish these two from each other in the discussion in Section 8. Figure 2 gives a rough picture of how the different hypotheses relate to each other.

6. The TV-STR Regime Shift Model

In this section I describe the Time-Varying Smooth Transition Regression (TV-STR) model of Lundbergh, Teräsvirta and van Dijk (2003). This is primarily to give an intuition for how the model works and how it can provide us with crucial information about the issues addressed in this paper. I do this by showing how one can augment the linear model (4.1) step-by-step to arrive at the TV-STR model. This also makes it obvious that the linear model is nested in the TV-STR model. For a detailed exposition, see Lundbergh, Teräsvirta and van Dijk, and also Teräsvirta (1994), which serves as the basic building block for the TV-STR model.

The linear model (4.1) can be written in vector form as

$$r^e_{t+k} = \beta_0 + \beta_1 DP_t + \varepsilon_{t+k} = \boldsymbol{\beta}'\mathbf{X}_t + \varepsilon_{t+k}, \tag{6.1}$$

$$\boldsymbol{\beta} = [\beta_0 \; \beta_1]', \tag{6.2}$$

$$\mathbf{X}_t = [1 \; DP_t]'. \tag{6.3}$$

Let us introduce two regimes, regime 0 and regime 1, in which the beta values are different. This can be written as

$$r^e_{t+k} = \begin{cases} \boldsymbol{\beta}_0'\mathbf{X}_t + \varepsilon_{t+k}, \text{ in regime 0} \\ \boldsymbol{\beta}_1'\mathbf{X}_t + \varepsilon_{t+k}, \text{ in regime 1} \end{cases}, \tag{6.4}$$

$$\boldsymbol{\beta}_0 = [\beta_{00} \; \beta_{01}]', \tag{6.5}$$

$$\boldsymbol{\beta}_1 = [\beta_{10} \; \beta_{11}]'. \tag{6.6}$$

A more compact way of writing the same thing is to introduce an indicator function, I_t, that takes on the value 0 in regime 0 and 1 in regime 1:

$$r^e_{t+k} = \boldsymbol{\beta}_0'\mathbf{X}_t (1 - I_t) + \boldsymbol{\beta}_1'\mathbf{X}_t I_t + \varepsilon_{t+k}, \tag{6.7}$$

Say, for example, that the regimes are "summer" and "winter"; when it is summer, $I_t = 0$, and when it is winter, $I_t = 1$.

However, we might not be satisfied with just a clean break between the summer and the winter, but would like to allow for a smooth transition between them. For instance, we might like it to be possible to be 10% in regime 0 and 90% in regime 1, corresponding to some point in early spring or late fall. One way to do that is to change the indicator function to a smooth continuous "transition function" that is bounded on $[0, 1]$. Here, I will consider the logistic function as the only candidate, as this function is one of the most popular choices in the literature. Substituting $G(t)$ for I_t, where the former is the logistic function with time as argument, the model can be written as

$$\begin{aligned} r^e_{t+k} &= \boldsymbol{\beta}_0'\mathbf{X}_t (1 - G(t)) + \boldsymbol{\beta}_1'\mathbf{X}_t G(t) + \varepsilon_{t+k}, && (6.8)\\ G(t) &= \left[1 + \exp\{-\gamma (t - c)\}\right]^{-1}, \quad \gamma > 0. && (6.9) \end{aligned}$$

G has an S-shape, where the parameter γ controls the steepness of the regime shift (a higher value corresponds to a steeper shift), and c controls the position along the X-axis where the midpoint of the shift takes place (i.e., $G = 0.5$ at c). For examples of what G might look like, see the two middle plots in Figures 3a-c. This model is similar to that of Lindgren (1978) in that it allows for regimes with different betas. Another similar model, that has recently become popular among financial economists, is that of Hamilton (1989). Hamilton's model, however, assumes that the economy is either in regime 0 or in regime 1 (and not in between), but that the true state is not observable. It then estimates from the data the probability of being in a certain regime. In (6.8), in contrast, we can observe the variable that determines which regime the economy is in (i.e., we can observe t). Another difference is that Hamilton's model only allows for regime shifts in the constant, whereas (6.8) allows for shifts in the sensitivity to all factors.

The model (6.8)-(6.9) only allows for changes over time, and we want to extend it even further so that we can also add a regime shift over the PE ratio. The suggestion given by Lundbergh, Teräsvirta and van Dijk (2003) is to add a second transition function, so that instead of just two regimes with a smooth transition between them, we have two-by-two regimes with smooth transitions between them all. The final model, the TV-STR model, is

$$\begin{aligned} r^e_{t+k} &= \left[\boldsymbol{\beta}_1'\mathbf{X}_t (1 - G(t)) + \boldsymbol{\beta}_2'\mathbf{X}_t G(t)\right]\left[1 - G(PE_t)\right] && (6.10)\\ &\quad + \left[\boldsymbol{\beta}_3'\mathbf{X}_t (1 - G(t)) + \boldsymbol{\beta}_4'\mathbf{X}_t G(t)\right] G(PE_t) + \varepsilon_{t+k}, \\ G(PE_t) &= \left[1 + \exp\{-\gamma_1 (PE_t - c_1)\}\right]^{-1}, \quad \gamma_1 > 0, && (6.11)\\ G(t) &= \left[1 + \exp\{-\gamma_2 (t - c_2)\}\right]^{-1}, \quad \gamma_2 > 0. && (6.12) \end{aligned}$$

6.1. Testing Different Transition Variables. Before applying the TV-STR model, we want to test if it is appropriate for the data. In other words, we want to test if there really is any regime-switching, as in (6.11) and (6.12) in the data. Note that if $\gamma_1 = 0$, then $G(PE)$ becomes a constant. Hence, if $\gamma_1 = 0$, there is no regime shift in the PE dimension. Similarly, if $\gamma_2 = 0$, there is no regime shift in the t dimension. Lundbergh, Teräsvirta and van Dijk (2003) develop a simple test for the appropriateness of using a certain variable as regime indicator that utilizes this property. The test statistic is described in the appendix.

We want to test for several situations. We might have regime-switching behavior in both dimensions, we might have it in one dimension but not the other, or we might have it in neither dimension. This makes it necessary to use a series of tests. Following the suggestions of Lundbergh, Teräsvirta and van Dijk[6], I start by testing

$$H_0: \ \gamma_1 = 0 \text{ and } \gamma_2 = 0 \text{ against } H_1: \ \gamma_1 > 0 \text{ and/or } \gamma_2 > 0. \tag{6.13}$$

This corresponds to testing (4.1) against (6.10)-(6.12). Rejecting H_0 means that we reject the linear model in favor of the TV-STR model.

However, this test would reject the linear model even if there is regime-switching behavior in only one of the two dimensions. If we reject H_0, we therefore move further and test if we really need both dimensions. This amounts to two conditional tests: Given that we have rejected H_0, can we reject that there is no change in the PE dimension? Or can we reject that there is no change in the t dimension? That is, we want to test

$$H_{01}: \ \gamma_1 = 0 | \gamma_2 > 0 \text{ against } H_{11}: \ \gamma_1 > 0 | \gamma_2 > 0, \tag{6.14}$$

and

$$H_{02}: \ \gamma_2 = 0 | \gamma_1 > 0 \text{ against } H_{12}: \ \gamma_2 > 0 | \gamma_1 > 0. \tag{6.15}$$

If we do not reject H_{01}, the remaining model has regime-switching behavior only in the t dimension. The TV-STR model then collapses to the Time-Varying Regression (TV-R) model. Note that, as a special case in this event, if γ_2 turned out to be a very large number, we would arrive at an almost instantaneous transition similar to a structural break. The structural break model is, consequently, nested in the TV-STR model. On the other hand, if we do not reject H_{02}, the remaining model has regime-switching behavior only in the PE dimension. The TV-STR model in that case collapses to the Smooth Transition Regression (STR) model. We could think of this as a test for a structural change (possibly a break) while controlling for the PE ratio. Finally, if we reject both H_{01} and H_{02}, the TV-STR model is accepted.

7. Empirical Results

In this section, I discuss the statistical results from the estimations and some related problems. Different interpretations of the results are discussed in the next section.

[6] See their paper for a discussion of strategies for finding appropriate transition variables.

7.1. Testing for Model Specification. Shiller (2000) discusses different versions of the PE ratio as an indicator of possibly inflated prices and, citing the famous work of Graham and Dodd (1934), suggests that earnings should be measured as an average over a long interval. I therefore consider three versions of the PE ratio: $PE_t = \frac{P_t}{E_t}$ incorporates no historical data previous to time t. PE_t^{12} uses the price from time t but the average of earnings over 12 months up to t, and, similarly, PE_t^{60} uses an average of earnings over 60 months (see also the appendix). Table 3 reports p-values for tests (6.13)-(6.15) for each choice of horizon and for all three versions of the PE ratio.

For both the 12-month and 36-month horizons, the lowest p-value in the TV-STR column is attained using PE_t. For the 60-month horizon, PE_t^{12} receives the lowest p-value, but the magnitude of the p-values in this case seems to make the choice between PE_t and PE_t^{12} more or less arbitrary. To make comparisons easier, I choose to use PE_t for all horizons. Furthermore, when using PE_t as transition variable, we reject H_{01} and H_{02}, and consequently choose the TV-STR model for both the 12-month and 60-month horizons. For the 36-month horizon, when testing H_{02}, we obtain a p-value as high as 0.092, which indicates that the time variation in this case may not be as strong as in the other cases. We will see in Section 7.2 that the data on the 36-month horizon have a different fit than the others; I will discuss this further in Section 8.

Consequently, the choice between PE_t and PE_t^{12} is close to irrelevant in the present setting. The same cannot be said about using the 60-month average, PE_t^{60}. Here, with earnings averaged over a long period, the model does not perform as well.

7.2. Model Estimates. The TV-STR model is estimated using non-linear least squares for each of the horizons: 12, 36, and 60 months. T-statistics are calculated according to Newey and West (1987), as the use of overlapping observations creates a high level of serial correlation in the errors. The number of lags in the Newey-West estimates is chosen so as to equal the horizon length.

I begin by estimating the TV-STR model with $\mathbf{X}_t' = [1\ DP_t]$ in all instances in (6.10). As long as the lowest absolute value of the t-statistic corresponding to a DP_t variable is less than 1, I restrict the corresponding β to zero and reestimate the model. Table 4 reports the estimated parameter values of the final model, with "–" referring to a β restricted to zero. The DP component in β_2 is omitted for all horizons due to low t-values. For the 12-month horizon, the DP component in β_4 is also omitted for the same reason.

Comparing the results from the linear model (4.1) with those from the TV-STR model (6.10)-(6.12), we notice that almost all beta estimates are (much) higher in the latter (see Tables 2 and 4). Estimates of coefficients of DP_t in the linear model range between 0.05 and 0.10 for the entire sample period, while for the TV-STR model most non-restricted estimates range between 0.40 and 1.00. The situation is similar for the constants: Almost all are (much) higher in the TV-STR model. These results are consistent with those obtained by Lettau and van Nieuwerburgh (2006) find when comparing their structural break model with a linear one. Furthermore, betas for longer horizons are typically lower than for shorter horizons, which is similar to the findings of Fama and French (1988).[7]

[7] Fama & French do not annualize returns. Since the size of the betas in their study grows slower than the horizon, the results are similar.

Table 3
P-values for Different Choices of Transition Variable

	H_0: Linear vs. TV-STR	H_{01}: No shift in PE dimension	H_{02}: No shift in t dimension
		12 Months	
PE_t	**0.00130**	**0.000716**	**0.00172**
PE_t^{12}	**0.00292**	**0.00330**	**0.00692**
PE_t^{60}	**0.00387**	**0.00564**	**0.00462**
		36 Months	
PE_t	**0.00853**	**0.00129**	0.0924
PE_t^{12}	**0.0100**	**0.00161**	0.0710
PE_t^{60}	0.107	**0.0455**	0.195
		60 Months	
PE_t	**6.69e-5**	**0.000264**	**0.00252**
PE_t^{12}	**1.53e-8**	**2.27e-9**	**0.00384**
PE_t^{60}	**0.0155**	**0.0176**	**0.00724**

This table presents p-values to assess the appropriateness of different transition variables in the TV-STR framework of Lundbergh, Teräsvirta and van Dijk (2003). P-values are presented for horizons of 12, 36, and 60 months. See also the appendix for a brief description of the methodology.
I use a Wald test, and p-values are calculated according to Newey and West (1987) with a lag length equal to the horizon length.

Table 4
Parameter Estimates for the TV-STR Model

		12 Months		36 Months		60 Months	
#obs		582		558		534	
		beta	t-value	beta	t-value	beta	t-value
"early/low"	$\beta_{1,C}$	**2.303**	**3.27**	1.572	1.70	**1.145**	**7.04**
	$\beta_{1,DP}$	**0.805**	**3.21**	0.604	1.65	**0.370**	**7.00**
"late/low"	$\beta_{2,C}$	0.509	1.81	5.042	0.86	16.54	0.71
	$\beta_{2,DP}$	—		—	—	—	—
"early/high"	$\beta_{3,C}$	**8.695**	**2.07**	**1.457**	**4.06**	**0.697**	**6.04**
	$\beta_{3,DP}$	**2.361**	**2.12**	**0.409**	**4.10**	**0.197**	**5.45**
"late/high"	$\beta_{4,C}$	-0.132	-0.83	**4.086**	**4.25**	**1.575**	**6.13**
	$\beta_{4,DP}$	—	—	**0.948**	**4.60**	**0.382**	**6.22**
G(PE)	γ_1	**1.147/0.3880**	**1.98**	**1.295/0.3557**	**1.96**	**12.44/0.3364**	**4.46**
	c_1	**2.881**	**8.86**	**2.012**	**4.92**	**2.355**	**227**
G(t)	γ_2	28.045	1.44	**22.330**	**6.00**	**29.13**	**2.88**
	c_2	**0.720 (1989.9)**	**21.27**	**0.905 (1997.0)**	**36.99**	**0.839 (1992.3)**	**49.8**

This table presents (non-linear least squares) parameter estimates for the TV-STR model (6.10)–(6.12) for horizons of 12, 36, and 60 months. T-values are calculated according to Newey and West (1987) with a lag length equal to the horizon length. All variables are in logs (see the data appendix). The betas are labeled "early/low" etc., to make interpretation easier. An observation that occurs far before the c_2 point (i.e., the point where the transition over time is halfway between regimes) is considered to be "early," and its behavior is mainly governed by the "early" betas. Similarly, an observation that occurs at a time when PE_t is far lower than c_1 is considered to be "low." In the estimation, γ_1 was normalized by its standard deviation and I present both values here. The full value of γ_1 is the value of the ratio (e.g. 1.147/0.388 = 2.956 for the 12-month horizon). The value beneath the estimates of c_2 is the year that the estimated c_2 corresponds to.

With the exception of an insignificant constant in the TV-STR model, all estimates are positive. Since the log dividend yield is negative (see Table 1a), the constants are positive to compensate.

The betas in Table 4 are labeled "early/low," "late/low," etc., to make interpretations easier. "Early" refers to the predominant regime before the midpoint, c_2, and "late" to the predominant regime after the midpoint. Note that the midpoints have different locations for different horizons, so that what is "early" in one horizon might be "late" in another. Also, note that the transitions are smooth, so that for several years, the economy is between the "early" and the "late" states. Similarly, "low" refers to values of $PE < c_1$, and vice versa for "high."

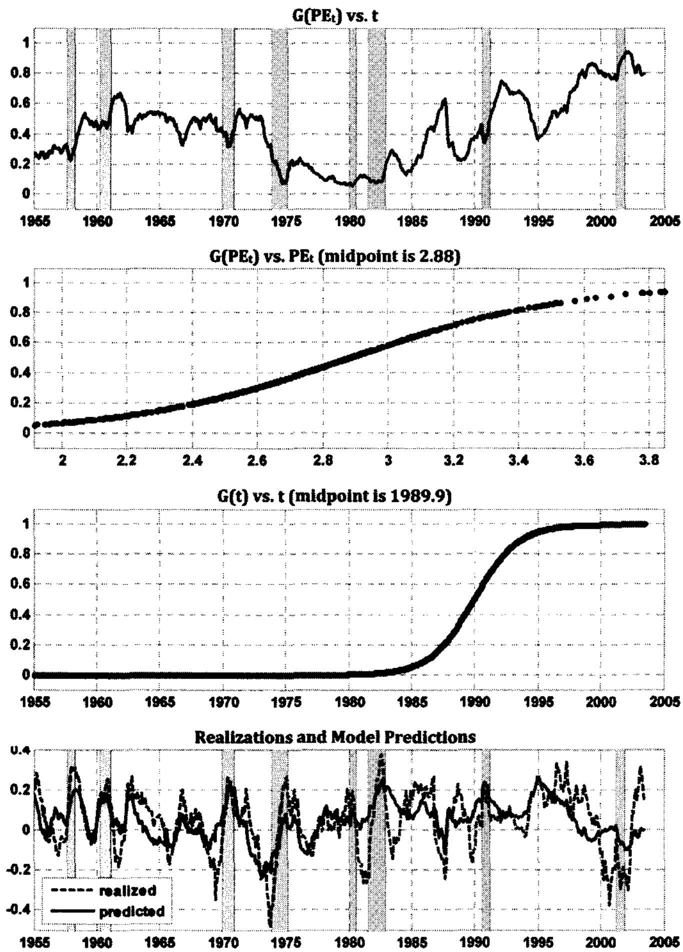
Figures 3a-c show plots of the transition functions, $G(PE)$ and $G(t)$. "Early," "low," etc. are easy to interpret in the plots. For instance, "early" is the part in the third plot where $G(t)$ is close to 0, while "late" is the part where it is close to 1. "Low" and "high" can be given similar interpretations in the first and second plots. β_1, for example, is then labeled "early/low" since it is the major determinant of the model behavior when $G(t)$ is close to 0 ("early") and $G(PE_t)$ is also close to 0 ("low").

The transitions with respect to time (the third plot in Figures 3a-c) look similar for all horizons. They all show a shift that starts at some point in the second half of the 1980s and ends at some point between 1995 and 2005. These results are consistent with those of Lettau and van Nieuwerburgh (2006), and Paye and Timmermann (2006), who find evidence of a structural break during this period. They also agree with those of McGrattan and Prescott (2005), who argue that changes in tax regulations during this period impacted stock returns. For the 12-month and 60-month horizons, the midpoint in the transition occurs around 1990 to 1992. For the 36-month horizon it occurs several years later, around 1997. This latter fact is probably one reason why we do not reject H_{02} in (6.15) for this horizon (see Table 3). There are simply too few observations in the "late" part to render any significance.

The transition with respect to PE (the second plot in Figures 3a-c) looks quite different for each horizon. For the 12-month and 36-month horizons, the shift is very smooth, while for the 60-month horizon it is quite abrupt. The midpoint of the estimates also changes with the horizon. For the shortest horizon, it is rather high at around 2.9 (38% of the PE observations are above this figure), while for the other horizons it is much lower, at around 2–2.35 (98% and 82%, respectively, are above these figures). This difference between the 12-month horizon on the one hand and the 36-month and 60-month horizons on the other becomes even more apparent if we concentrate on observations that are very close to the "high" regime. For instance, count an observation of PE_t as "high" if $G(PE_t) \geq 0.8$. Then 7%, 82%, and 81% of the observations are "high" for the 12-month, 36-month, and 60-month horizons, respectively. For the 12-month horizon, "high" occurs quite rarely, while for the 36-month and 60-month horizons it is more of a normal state. I will discuss this further in Section 8.

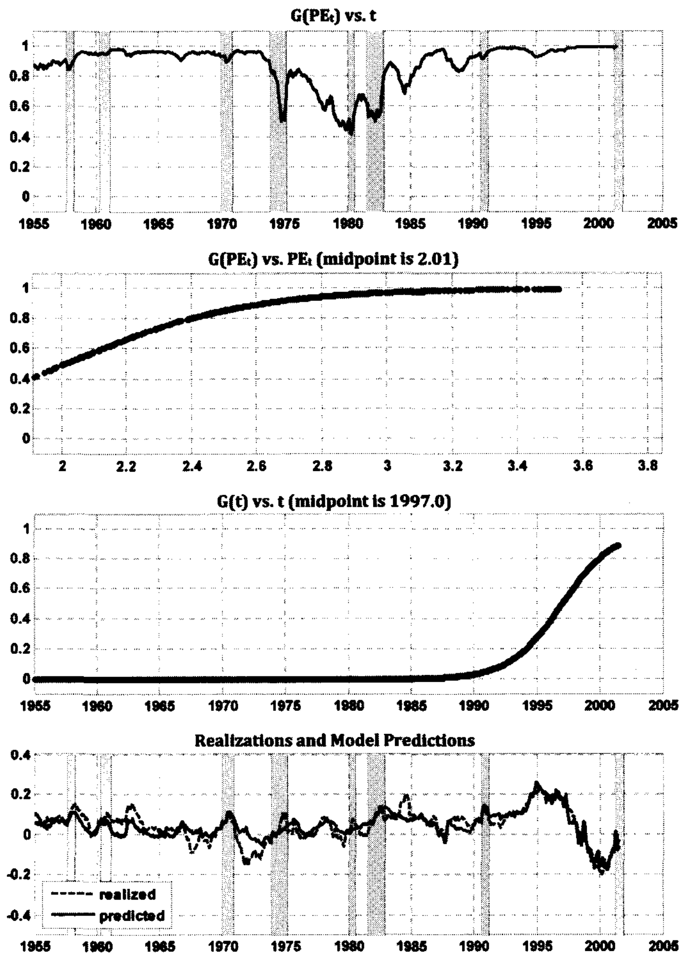
When it comes to the 36-month horizon, the shape of $G(PE)$ is quite different from that of the others. It does not display the full S-shape, but only the upper part of it. The curvature that is needed to explain the data in this case is consequently also different

Figure 3a
Transition Functions and Predicted Excess Returns at the 12-Month Horizon, 1955—2004



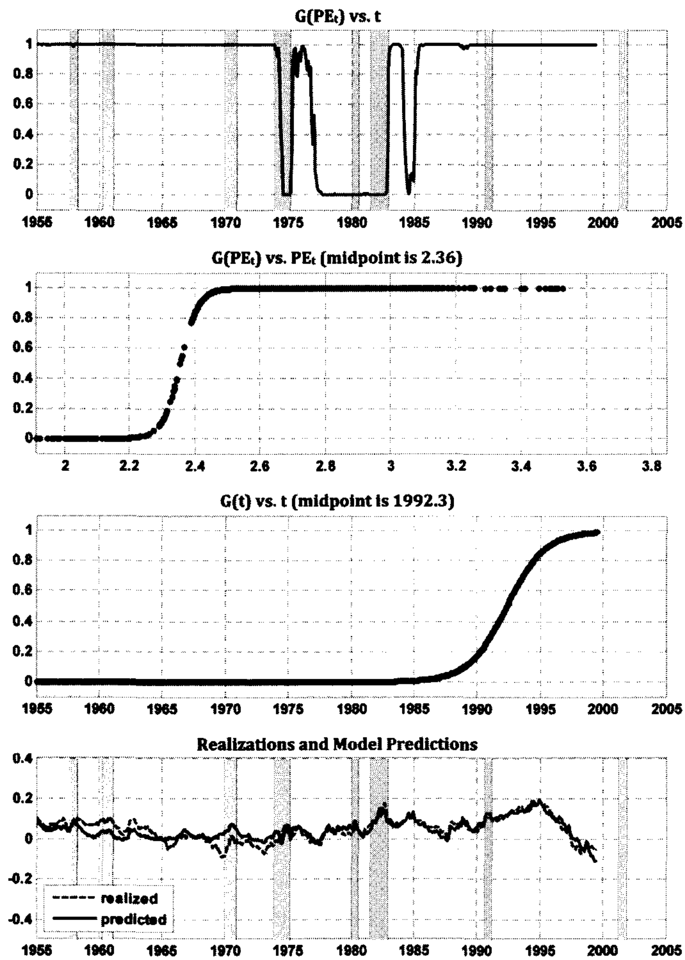
The first graph shows the estimated values of $G(PE_t)$ versus time, and the second graph shows $G(PE_t)$ versus PE_t, i.e., where and how rapid the transition is with respect to the price earnings ratio.
The third graph shows the estimated values of $G(t)$ versus time, i.e., where and how rapid the transition is over time.
The fourth graph shows the 12-month ahead predicted excess return together with the 12-month ahead realized excess return (i.e., the realized return from t to $t + 12$).
The shaded areas correspond to NBER recession periods.

Figure 3b
Transition Functions and Predicted Excess Returns at the 36-Month Horizon, 1955—2004



The first graph shows the estimated values of $G(PE_t)$ versus time, and the second graph shows $G(PE_t)$ versus PE_t, i.e., where and how rapid the transition is with respect to the price earnings ratio.
The third graph shows the estimated values of $G(t)$ versus time, i.e., where and how rapid the transition is over time.
The fourth graph shows the 36-month ahead predicted excess return together with the 36-month ahead realized excess return (i.e., the realized return from t to $t + 36$).
The shaded areas correspond to NBER recession periods.

Figure 3c
Transition Functions and Predicted Excess Returns at the 60-Month Horizon, 1955-2004



The first graph shows the estimated values of $G(PE_t)$ versus time, and the second graph shows $G(PE_t)$ versus PE_t, i.e., where and how rapid the transition is with respect to the price earnings ratio.
The third graph shows the estimated values of $G(t)$ versus time, i.e., where and how rapid the transition is over time.
The fourth graph shows the 60-month ahead predicted excess return together with the 60-month ahead realized excess return (i.e., the realized return from t to $t + 60$).
The shaded areas correspond to NBER recession periods.

from that of the others. Primarily, the function singles out low values of PE as differing more and more (i.e., the second derivative of $G(PE)$ is negative). For most of the time, the value of G is very close to 1; the main exception being a period during about 1974 to 1985 when it occasionally drops to about 0.5. Consequently, the 36-month horizon spends most of its time in the "high" regime, none of its time in the "low" regime, and only a small amount of its time in an in-between regime.

Looking at Figure 3c, we see a somewhat similar pattern for the 60-month horizon. However, the period from 1974 to 1985 here demonstrates a sharp break from the pattern of the rest of the sample period. For this horizon we have a very pronounced either/or pattern with very few observations in between.

The behavior of $G(PE)$, to some extent, explains why $\widehat{\beta}_2$ in (6.10) has such low t-values for all horizons. The parameter corresponds to the "late/low" regime but, particularly for the 36-month and 60-month horizons, no, or very few, observations are simultaneously "late" and "low." We thus have very little information with which to estimate the parameter, and it comes out insignificant for all horizons.

Table 5 presents some residual diagnostics for both the TV-STR model and the linear alternative for each horizon. The TV-STR model has markedly lower standard deviation, as well as lower values for both the Akaike Information Criterion (AIC) and the Schwartz Information Criterion (BIC). I conclude that it represents a significant improvement over the linear model.

8. Discussion

The k-month-ahead predicted values of r^e_{t+k} from the TV-STR model (6) are plotted in the last graphs in Figure 3a-c. These values may be interpreted as the equity risk premium. For the 12-month horizon, we have many predictions that are below zero. It is, of course, difficult to believe that investors would hold stocks if they predicted that the reward would be negative, but this is a problem that mars just about any regression model in the literature.[8] For the two longer horizons, except at the end of the sample period, very few predictions are below zero. Note also that since $G(PE)$ does not single out the years around 2000 for the longer horizons, they attribute most of the low realized returns during these years to a new regime in t.

8.1. The IT Bubble and the Transition with Respect to the PE Ratio. The transition function $G(PE)$ shows some interesting differences for different horizons. The estimate of c_1, corresponding to the point where $G(PE) = 0.5$ in the second graphs in Figures 3a-c, is much higher for the 12-month horizon than for the other two horizons. As shown in Section 7, very "high" observations (meaning that $G(PE)$ is very close to 1) are also rare for this horizon. The behavior of $G(PE)$ for the 12-month horizon is

[8] Campbell & Thompson (2005) argue in a similar context that if we want to test predictions of returns out-of-sample, then negative predictions should be set to zero. In the present case, and with the TV-STR model, it makes little sense to test the results out-of-sample. Most of the regime-switching behavior, particularly in the t-dimension, occurs in the last period of the sample. Deleting that part amounts to deleting exactly those properties that the model is meant to capture.

Table 5
Residual Diagnostics

	12 Months		36 Months		60 Months	
	Linear	TV-STR	Linear	TV-STR	Linear	TV-STR
min	-0.529	-0.383	-0.195	-0.163	-0.135	-0.119
max	0.332	0.320	0.200	0.130	0.155	0.0657
st.dev	0.147	0.118	0.0751	0.0419	0.057	0.0326
skewness	-0.475	-0.50	0.265	-0.418	0.404	-0.369
kurtosis	2.99	2.92	3.38	4.18	3.06	3.38
JB	**21.9**	**24.5**	**9.85**	**48.4**	**14.6**	**15.4**
(p-val)	**(1.77e-5)**	**(4.86e-6)**	**(0.0073)**	**(3.2e-11)**	**(6.7e-4)**	**(6.1e-4)**
AIC	-2230.86	-2468.44	-2885.44	-3517.42	-3054.80	-3635.42
BIC	-2222.13	-2424.78	-2876.79	-3469.85	-3046.24	-3588.34

This table presents residual diagnostics from both the linear model (4.1) and the TV-STR model (6.10)–(6.12) for horizons of 12, 36, and 60 months. JB is the Jarque-Bera statistic for testing normality with p-values in parentheses. AIC is the Akaike Information Criterion and BIC is the Schwartz Information Criterion.

Table 6
Did the DP Ratio forecast Returns or Growth during the IT bubble?

	$r^e_{t+12} = \beta_0 + \beta_1 DP_t + \varepsilon_{t+12}$			$g_{t+12} = \gamma_0 + \gamma_1 DP_t + \eta_{t+12}$	
#obs	65			65	
	Estimate	t-value		Estimate	t-value
β_0	3.25	1.61	γ_0	**0.0672**	**2.66**
β_1	0.767	1.68	γ_1	**0.0155**	**2.62**

This table presents parameter estimates of two linear regressions performed at the 12-month horizon and over a sample period between 1998 and 2004. The first panel shows estimates for predicting excess returns with the DP ratio, and the second panel shows estimates for predicting dividend growth with the DP ratio.
All variables are in logs (see the data appendix) and t-values are according to Newey and West (1987) with a lag length equal to the horizon length.

quite consistent with the IT bubble arguments in Section 5. If there ever was a bubble or a period of extreme optimism, it would have occurred around 2000. Furthermore, the bubble should be the exception, not the rule. Looking at the first graph in Figure 3a, it is easy to interpret it along those lines. Between, roughly, 1998 and 2003, $G(PE)$ is closer to "high" than during any other period in the sample, and for most of the remaining time it is much lower at between 0 and 0.5.

The same is not true for the 36-month and 60-month horizons. The estimates of c_1 being so low, most observations fall in the "high" regime. The estimate of γ_1, measuring the speed of the transition, differs considerably between the two horizons. The transition is very slow for the 36-month horizon, but quite fast for the 60-month horizon. For the latter, this puts most observations either in the "low" regime or in the "high" regime, whereas for the former it puts most in the "high" regime and some in a "in between" state. As the years around 2000 are not singled out, and "high" is the regime in which most observations fall, the IT bubble is not a likely underlying cause for why the PE ratio explains some of the predictability of returns.

We may recall, though, that we rejected that the model could do without the PE ratio (see Section 6.1), so this variable definitely picks up something. What is that? Looking at Figure 3c, we see that $G(PE)$ picks out a few years around, roughly, 1974–1985, as differing from the rest. It is easy to tie a macroeconomic interpretation to this period: It corresponds to the high inflation period (about 1973–1982), and to the Fed's experiment in controlling money supply and the associated high and volatile interest rates of 1980–1982. Tied to both of these phenomena are the recession periods of November 1973–March 1975, January–July 1980, and July 1981–November 1982. As is quite evident from the first graph in Figure 3c, the first of these recession periods coincides with the first change into the "low" regime–they both begin and end at the same time. The end of the third recession period similarly coincides with a change from "low" to "high." It is reasonable to believe that the risk factor the PE ratio picks up for the 60-month horizon is high inflation, with associated recession risk and high interest rates. It is not reasonable to believe that it is the IT bubble.

The two different explanations for the 12-month and 60-month horizons, the IT bubble period and inflation risk, respectively, might also explain the results for the 36-month horizon. For the 12-month horizon, the impact of the IT bubble appears to dominate over the inflation risk. For the 60-month horizon, it seems to be the other way around. As the 36-month horizon lies between them, it appears that the inflation risk and the IT bubble counter-balance each other. And while we have seen that the PE ratio could proxy for either one of them, in this case the picture becomes blurred.

8.2. The Transition with Respect to Time. The evidence suggesting the economy has undergone a regime shift that can be translated into a time trend is strong. The estimated transition for the 36-month horizon is quite similar to those for the other two horizons, lending credibility to the interpretation that the time trend has occurred even though we rejected it for this horizon in Table 3. Lettau and van Nieuwerburgh (2006), and Paye and Timmermann (2006) also find a structural break somewhere during the 1990s. The TV-STR model allows us to use a smooth transition, with a break nested in the model, and I find that a smooth transition, somewhere between 1985 and 2000, fits the data better than a break does.

One concern has been that the empirical evidence researchers have found of a new regime, starting somewhere in the 1990s could have been an artefact of the IT bubble period. Although that period might have influenced previous empirical work, as it seems to have done at the 12-month horizon in this paper, I conclude that the data point to a regime shift actually having taken place.

Looking at the hypotheses schema in Figure 2, we would then conclude that the "Changes to the underlying economy" part is a very likely explanation for changes in predictability for all horizons. The "Unusually high PE ratio" part is a likely explanation only for the shortest horizon, as the PE ratio is never *unusually* high for the other horizons. The PE ratio remains an explanation for the longer horizons, though, but as a signal of inflation/recession.

8.3. Predictability at the 12-month Horizon. So, does the dividend yield predict equity return? For the 12-month horizon and "early" in the sample period, DP is significant regardless of whether PE is "high" or "low." During this part of the sample period, it appears reasonable that the PE ratio fluctuates jointly with the business cycle. During a recession, the required rate of return tends to be high, which consequently pushes down prices and, hence, the PE ratio. A low PE ratio would consequently predict higher future returns. Looking at the first graph in Figure 3a, we see some evidence that the PE ratio is lower during recession periods. During the "early" part of the sample period, both the constant and the beta on DP are higher in the "high" PE regime than in "low" regime (see Table 4). This could lead us to believe that the model would forecast low returns rather than high returns after such a period. In fact, precisely the opposite is true. The reason is that PE and DP are, as noted in Section 3, very negatively correlated. This means that when PE is "high," the corresponding constant and beta on DP are higher than when PE is "low." But because of the correlation, DP is low. The result can be read off the bottom graph in Figure 3a: During the recession periods, the model predicts high future returns.

I argued briefly in the introduction that during a period of irrational prices returns might be less predictable by a rational pricing model such as the one underlying Campbell and Shiller (1988). We could get a simple picture of this argument by augmenting (2.2) with a bubble term, B_t, so that we obtain

$$\frac{D_{t+1}}{P_t} = r - g + B_t. \tag{8.1}$$

Obviously, with an extra term in the formula, the dividend yield no longer has to predict either returns, r, or growth, g. As B_t may vary freely, the predictability of returns could vanish completely during a bubble period.

Quite consistent with this argument, the beta estimate on DP in the "late/high" regime in Table 4 has a t-value below 1 (and has therefore been omitted). The dividend-price ratio appears to be useless as a predictor of returns in the bubble period. In contrast to the "late/low" regime, where the estimate on DP is also below 1, this cannot be due to a lack of information, since most observations in the "late" part also fall close to the "high" part. The insignificance of the DP component in the "late" period is also in contrast to the "early" period, where all t-values are significant and, consequently, the dividend-price ratio does explain returns. Note also that for the other two horizons, for which we found the bubble-interpretation invalid, the beta estimates on DP in the "late/high" regime are significant. These results would then support the "inflated prices" argument of Figure 2.

However, we can also find results that support the "growth" argument of the same figure. In Figure 3a we see that from 1998 on, $G(t)$ is firmly in the "late" regime, and $G(PE)$ is as close as it ever gets to the "high" regime. During this period, the economy is then predominantly in "late/high." Table 6 presents subsample estimates, over 1998–2004, of two regressions. In the first, excess returns are regressed on the dividend-price ratio; in the second, growth, g, is regressed on the dividend-price ratio. As we see in Table 6, growth is strongly predictable during this period, whereas returns are only weekly significant with a t-value of 1.68. Consequently, we have evidence to support both the underlying explanations for the high PE ratio in Figure 2.

8.4. Predictability at the 36-month Horizon. The 36-month horizon has proven interesting mainly as an in-between horizon. It shares properties with both the 12-month and 60-month horizons, but the PE ratio seems to fail to provide a clear signal of either irrational prices or inflation risk for this horizon. I have argued that this might be because it signals a little of both.

As all t-values corresponding to "high" are significant, and, as this horizon spends almost all of its time close to the "high" regime, I conclude that the dividend-price ratio predicts returns across the entire sample period. The beta on DP is higher "late" in the sample period, meaning that the predicted return has become more sensitive to this variable over time (roughly from the early 1990s to the early 2000s). Note that as DP is very low during this time period, this does not mean that the predicted return is higher during the "late" part. Rather, it means that it is more volatile: It reached both its highest value (around 1995) and its lowest value (around 2000) during this period.

8.5. Predictability at the 60-month Horizon. For the longest horizon, no observations whatsoever fall in the "late/low" regime, so the corresponding parameter estimates are insignificant. All others are strongly significant. As noted earlier, it is hard to give a bubble interpretation to the transition along the PE ratio in this case. There are indications, though, that the recessions, the high inflation, and the Fed's experiment between about 1974 and 1985 have influenced the regime shift.

The dividend-price ratio is useful for predicting returns over the whole sample period, but with different marginal contributions. During the "early/high" period, the beta on DP is about 0.20, while both in the "early/low" period and in the "late/high" period it is about 0.37–0.38. There appear to be no important changes in predictability at all: The lowest t-value on a DP variable is 5.4, which is still strongly significant.

9. Conclusion

In this paper, I investigate whether the dividend yield is useful for predicting excess returns. To this end, I estimate the TV-STR model, a very general econometric model that allows for simultaneous smooth regime-shifts in two different dimensions: time and state (represented by the PE ratio). Within this model the dividend yield is indeed able to predict future returns at different horizons. The results indicate consistently that a smooth transition in the relation between dividend yield and returns occurred during the 1990s.

The results concerning the dependence on the PE ratio, however, are different for different horizons. For the 12-month horizon, the transition function primarily singles out the IT bubble period, during which predictability seems to disappear. There is, instead, some evidence that dividend growth is predictable by the dividend yield during this period. For the 60-month horizon, the PE ratio seems to control for inflation/recession risk and there is no evidence that the IT bubble period has had any impact on predictability. For this horizon, the dividend yield is useful for predicting returns over the whole sample period. The behavior of returns over the 36-month horizon is more difficult to interpret; I argue that this is because the PE ratio at this horizon shares the properties of both the 12-month horizon and the 60-month horizon, thus blurring the picture. Still, the dividend yield is a significant predictor of returns over the whole sample period for this horizon as well.

Appendix

A.1. Data. The data used in this study are stock prices (P), dividends (D), and earnings (E) from Robert Shiller's website, and secondary market three-month T-bill rates (R^{3m}) from the St. Louis Fed's database FRED. Shiller computes monthly dividends and earnings from Standard & Poor's four-quarter total for the quarter, with linear interpolation to monthly figures. For the stock price data, Shiller uses monthly averages of daily closing prices of Standard & Poor's Composite Stock Price Index. For further information on how Shiller's data were constructed, see Shiller (1989 or 2000) or

`www.econ.yale.edu/~shiller/data.htm`.

The sample period I use starts in January 1955 and ends in June 2004.

The variables used in the empirical part are calculated as follows. $r_{t,k}$ is the annualized stock return realized at time t for an investment made k months earlier:

$$r_{t,k} = \frac{12}{k}\left[\log\left(\frac{P_{t-k+12}+D_{t-k+12}}{P_{t-k}}\right) + \ldots + \log\left(\frac{P_t + D_t}{P_{t-12}}\right)\right], \quad k = 12,\ 36,\ 60. \tag{A.1}$$

The annualized risk free return, $r^f_{t,k}$, over the same interval is calculated as

$$r^f_{t,k} = \frac{12}{k}\left[\log\left(1+\frac{R^{3m}_{t-k}}{4}\right) + \log\left(1+\frac{R^{3m}_{t-k+3}}{4}\right) + \ldots + \log\left(1+\frac{R^{3m}_{t-3}}{4}\right)\right], \tag{A.2}$$

and then the annualized excess return over the interval is

$$r^e_{t,k} = r_{t,k} - r^f_{t,k}. \tag{A.3}$$

Other variables used are

$$
\begin{aligned}
DP_t &= \log\left(\frac{D_t}{P_t}\right), && \text{(A.4)}\\
PE_t &= \log\left(\frac{P_t}{E_t}\right), && \text{(A.5)}\\
PE_t^{12} &= \log\left(\frac{P_t}{\frac{1}{12}\sum_{j=t-11}^{t} E_j}\right), && \text{(A.6)}\\
PE_t^{60} &= \log\left(\frac{P_t}{\frac{1}{60}\sum_{j=t-59}^{t} E_j}\right), && \text{(A.7)}\\
g_{t+k} &= \log\left(\frac{D_t}{D_{t-11}}\right). && \text{(A.8)}
\end{aligned}
$$

A.2. Finding the Appropriate Transition Variable. The methodology is the following. Let X_t be a matrix of explanatory variables, not including a constant; let t be a time index, and tr a suggested transition variable. y_t is the dependent variable. Define

$$\widetilde{X}_t = [\ 1 \quad \mathbf{X}_t \quad tr \times \mathbf{X}_t \quad t \quad t \times \mathbf{X}_t \quad tr \times t \times \mathbf{X}_t\ ], \tag{A.9}$$

and regress y_t on $\widetilde{X}_t$ to get the estimates $[\ \widehat{\beta}_0 \quad \widehat{\boldsymbol{\beta}}_1 \quad \widehat{\boldsymbol{\beta}}_2 \quad \widehat{\beta}_3 \quad \widehat{\boldsymbol{\beta}}_4 \quad \widehat{\boldsymbol{\beta}}_5\ ]$. This is the unrestricted model.

Testing a linear model against the TV-STR model (i.e., H_0 in (6.13)) amounts to testing the restriction $\boldsymbol{\beta}_2 = \beta_3 = \boldsymbol{\beta}_4 = \boldsymbol{\beta}_5 = \mathbf{0}$. Testing if the transition along tr is necessary, given that we have rejected the linear model (i.e., H_{01} in (6.14)), amounts to testing the restriction $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_5 = \mathbf{0}$. And, finally, testing if the transition along time, t, is necessary, given that we have rejected the linear model (i.e., H_{02} in (6.15)), amounts to testing the restriction $\beta_3 = \boldsymbol{\beta}_4 = \boldsymbol{\beta}_5 = \mathbf{0}$. For details, see Lundbergh, Teräsvirta and van Dijk (2003). I use a Wald test, and p-values are calculated according to Newey and West (1987) with a lag length equal to the horizon length.

References

Ang, A. and G. Bekaert, 2007, "Stock Return Predictability: Is it There?," *Review of Financial Studies*, 20, 651–707.

Ang, A., and J. Liu, 2004, "How to Discount Cashflows with Time-Varying Expected Returns," *Journal of Finance*, 59(6), 2745-2783.

Barberis, N., 2000, "Investing for the Long Run when Returns are Predictable," *Journal of Finance*, 55(1), 225–264.

Bossaerts, P., and P. Hillion, 1999, "Implementing statistical criteria to select return forecasting models: what do we learn?," *Review of Financial Studies*, 12, 405–428.

Campbell, J. Y., A. W. Lo, and C. MacKinlay, 1997, *The Econometrics of Financial Markets.* Princeton University Press, Princeton, NJ.

Campbell, J. Y., and L. M. Viceira, 1999, "Consumption and Portfolio Decisions when Expected Returns are Time Varying," *Quarterly Journal of Economics*, 114(2), 433–495.

Campbell, J. Y., and R. J. Shiller, 1988, "The dividend-price ratio and expectations of future dividends and discount factors," *Review of Financial Studies*, 1, 195–228.

Campbell, J. Y., and S. Thompson, 2005, "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?," NBER Working Paper No. 11468.

Cochrane, J. H., 1992, "Explaining the variance of price-dividend ratios," *Review of Financial Studies*, 5, 243–280.

Cochrane, J. H., 1998, "Where is the Market Going? Uncertain Facts and Novel Theories," NBER Working Paper No. 6207.

Cochrane, J. H., 2006, "The Dog That Did Not Bark: A Defense of Return Predictability," NBER Working Paper No. 12026.

Fama, E. F., and K. R. French, 1988, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–27.

Ferson, W. E., S. Sarkissian, and T. T. Simin, 2003, "Spurious regressions in financial economics?," *Journal of Finance*, 58(4), 1393–1413.

Goetzman, W. N., and P. Jorion, 1993, "Testing the Predictive Power of Dividend Yields," *Journal of Finance*, 48(2), 663–679.

Gordon, M., 1962, *The Investment, Financing, and Valuation of the Corporation*, Homewood, Ill.: Richard D. Irwin.

Goyal, A., and I. Welch, 2003, "Predicting the Equity Premium with Dividend Ratios," *Management Science*, 49(5), 639–654.

Goyal, A., and I. Welch, 2007, "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," forthcoming in *Review of Financial Studies.*

Graham, B., and D. Dodd, 1934, *Securities Analysis.* New York: McGraw–Hill.

Hamilton, J. D., 1989, "A New Approach to the Analysis of Nonstationary Returns and the Business Cycle," *Econometrica*, 57, 357–384.

Hodrick, R. J., 1992, "Dividend yields and expected stock returns: alternative procedures for inference and measurement," *Review of Financial Studies*, 5, 357–386.

Lamont, O., 1998, "Earnings and Expected Returns," *Journal of Finance*, 53(5), 1563–1587.

Lettau, M., and S. C. Ludvigson, 2001, "Consumption, Aggregate Wealth and Expected Stock Returns," *Journal of Finance*, 56(3), 815–849.

Lettau, M., and S. van Nieuwerburgh, 2006, "Reconciling the Return Predictability Evidence," NBER Working Paper No. W12109.

Lewellen, J. W., 2004, "Predicting Returns With Financial Ratios," *Journal of Financial Economics*, 74(2), 209–235.

Lindgren, G., 1978, "Markov regime models for mixed distributions and switching regressions," *Scandinavian Journal of Statistics*, 5, 81–91.

Lintner, J.,, 1956, "Distribution of Incomes of Corporations Among Dividends, Retained Earnings, and Taxes," *American Economic Review*, 46(2), 97–113.

Lundbergh, S., T. Teräsvirta, and D. van Dijk, 2003, "Time-varying smooth transition autoregressive models," *Journal of Business & Economic Statistics*, 21, 104–121.

McGrattan, E. R., and E. C. Prescott, 2005, "Taxes, Regulations, and the Value of U.S. and U.K. Corporations," *Review of Economic Studies*, 72, 767–96.

Nelson, C. C., and M. J. Kim, 1993, "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance*, 48(2), 641–661.

Newey, W. K., and K. D. West, 1987, "A Simple, Positive Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Pastor, L., and R. F. Stambaugh, 2001, "The Equity Premium and Structural Breaks," *Journal of Finance*, 56(4), 1207–1239.

Paye, B. S., and A. Timmermann, 2006, "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13, 274–315.

Perron, P., 1989, "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57(6), 1361–1401.

Rozeff, M. S., 1984, "Dividend Yields are Equity Risk Premiums," *Journal of Portfolio Management*, 49, 141–160.

Shiller, R. J., 1989, *Market Volatility*, MIT Press.

Shiller, R. J., 2000, *Irrational Exuberance*, Princeton University Press.

Stambaugh, R. F., 1999, "Predictive Regressions," *Journal of Financial Economics*, 54, 375–421.

Teräsvirta, T., 1994, "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89(425), 208–218.

Valkanov, R., 2003, "Long-Horizon Regressions: Theoretical Results and Applications," *Journal of Financial Economics*, 68, 201–232.

Coupling and Decoupling: Changing Relations between Stock and Bond Market Returns

ABSTRACT. Comovement between stocks and bonds varies over time, making portfolio management and risk management difficult. I use GARCH estimates of stock volatility, simple regressions, and regime-switching econometric models to assess whether level of volatility, or changes in volatility, can be used to explain some of the changes in comovement in seven different countries. The main objective is to explain comovement changes from positive to negative and vice versa.

As regards volatility level, strong support is found in almost all countries to suggest that high volatility predicts lower, or negative, comovement. I argue that this can be evidence of a market-timing type of behavior. As for changes in volatility, the results are more mixed. Only for the U.S. market do we find strong support to conclude that large changes tend to coincide with lower, or negative, comovement. This can be evidence of a flight-to-quality (or cross-market hedging) type of behavior.

1. Introduction

The correlation between returns on different assets is of central importance for financial decision-making in many situations, such as portfolio management, risk management, and the pricing of some derivative instruments, to name just a few. In this paper, I present stylized facts on the comovement between broad national stock indices and government bonds from seven large economies. Usually, stock-bond correlation is positive, but it can be negative for long periods (Fleming, Kirby and Ostdiek (2003), Gulko (2002), and Li (2002)). I follow Connolly, Stivers and Sun (2005), and argue that stock market volatility explains some of the comovement between stocks and bonds. The main objective here is to show that the correlation is positive or negative depending, to some extent, on the level of volatility (high volatility predicts low correlation and vice versa). For several countries, it also depends on daily changes in volatility (large changes tend to coincide with low correlation). Connolly, Stivers and Sun show that this is the case in the U.S. economy. I extend their analysis to six other economies and also introduce a richer econometric framework.

Furthermore, I study stock-bond comovement in an international setting. Several researchers have studied comovement between stock markets in different countries, especially after the Asian crisis of 1997. A common conclusion is that comovement between stock markets is much higher, with a correlation close to one, during times of high volatility (Longin and Solnik (2001), and Campbell, Koedijk and Kofman (2002)). A practical implication of this phenomenon is that diversification between stock markets is of limited value during such times. This is a serious problem, as the main reason for diversifying is to protect oneself against high correlation. And this is particularly important during times of financial turbulence, which, in turn, often coincide with times of high volatility. Comovement between stocks and bonds from different domestic markets has attracted somewhat less attention, but is equally important. This prompts me to also study comovement between different stock markets and U.S. bonds, since U.S. bonds are often considered to be a "safe haven."

It is instructive to begin by looking at a graph. In Figures 1a and 1b, we see time-series of estimated time-varying correlations between daily returns on stocks and government bonds for seven countries, together with standard deviations for the stocks. It is easy to see that the high standard deviations, which occurred in particular in Germany, France, the U.K., and the U.S. between 1998 and 2002, coincide with low stock-bond correlation. The standard deviations do not vary as much prior to 1998, except for a brief period at the end of 1987, that is also accompanied by lower correlation.

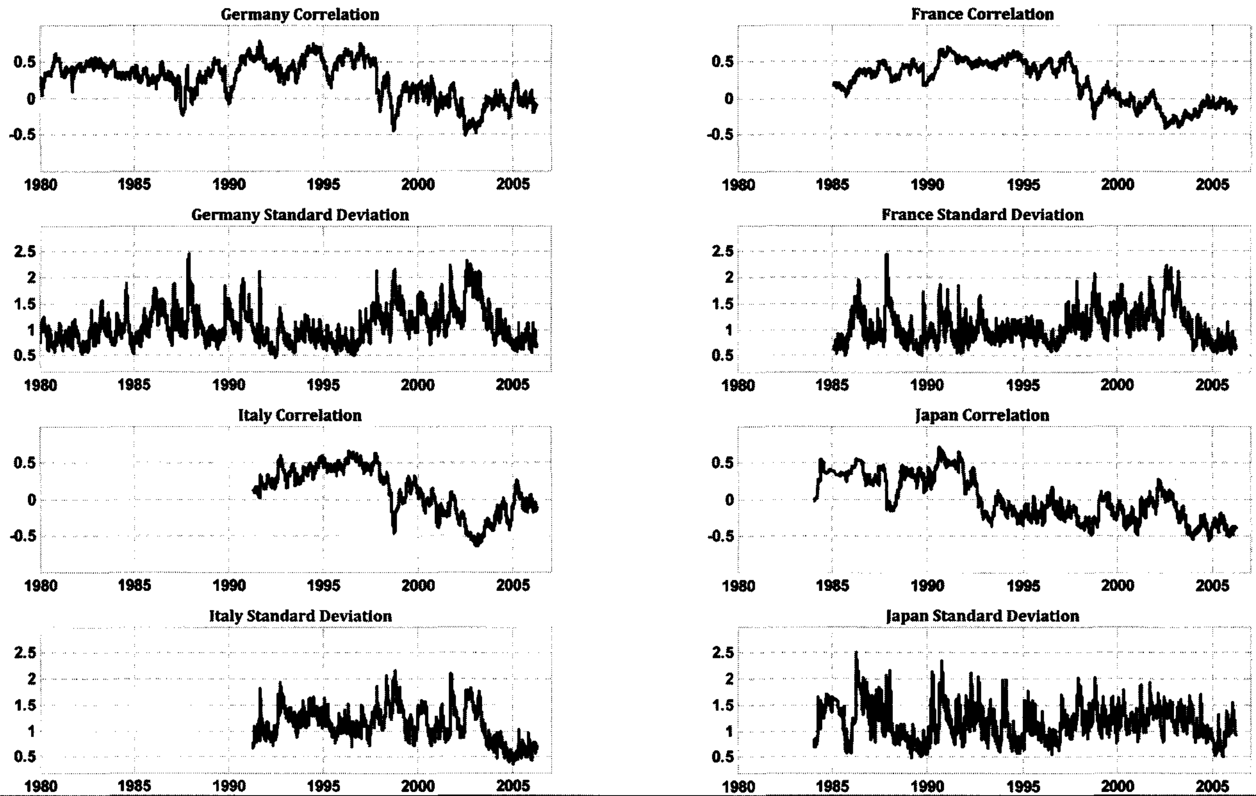
Regarding changes in standard deviations, the picture is more difficult to interpret. Many of the high volatility periods start with a large change (increase) in volatility. And as high volatility periods are associated with low correlation, many large changes in volatility will also be associated with low correlation. However, one can also find large jumps in standard deviations that are not accompanied by lower correlation. One example of this is the jump in the standard deviation in the U.S. market in 1982.

I use the simple method of regressing bond returns on control variables and the product of stock returns and the level of standard deviations ($S_t \times \log(\widehat{\sigma}_t)$, where S_t is the daily return on stocks and $\widehat{\sigma}_t$ is a GARCH estimate of the conditional standard deviation). When this latter term turns out significant, I reject that the level of volatility does not influence the comovement between stocks and bonds. This is similar to what Connolly, Stivers and Sun do on U.S. data, and I show that the same result holds in several other economies as well.

The idea that changes in volatility might influence stock-bond comovement is frequently put forward, particularly in the popular press. However, this is difficult to show empirically. Using a linear regression produces inconclusive results. I therefore apply a regime-switching econometric model, which allows us to test if changes in volatility carry information about stock-bond comovement, and assess how sensitive the comovements are to different sizes of change. It seems that comovement is primarily affected by large changes in volatility; small changes appear to have almost no effect at all.

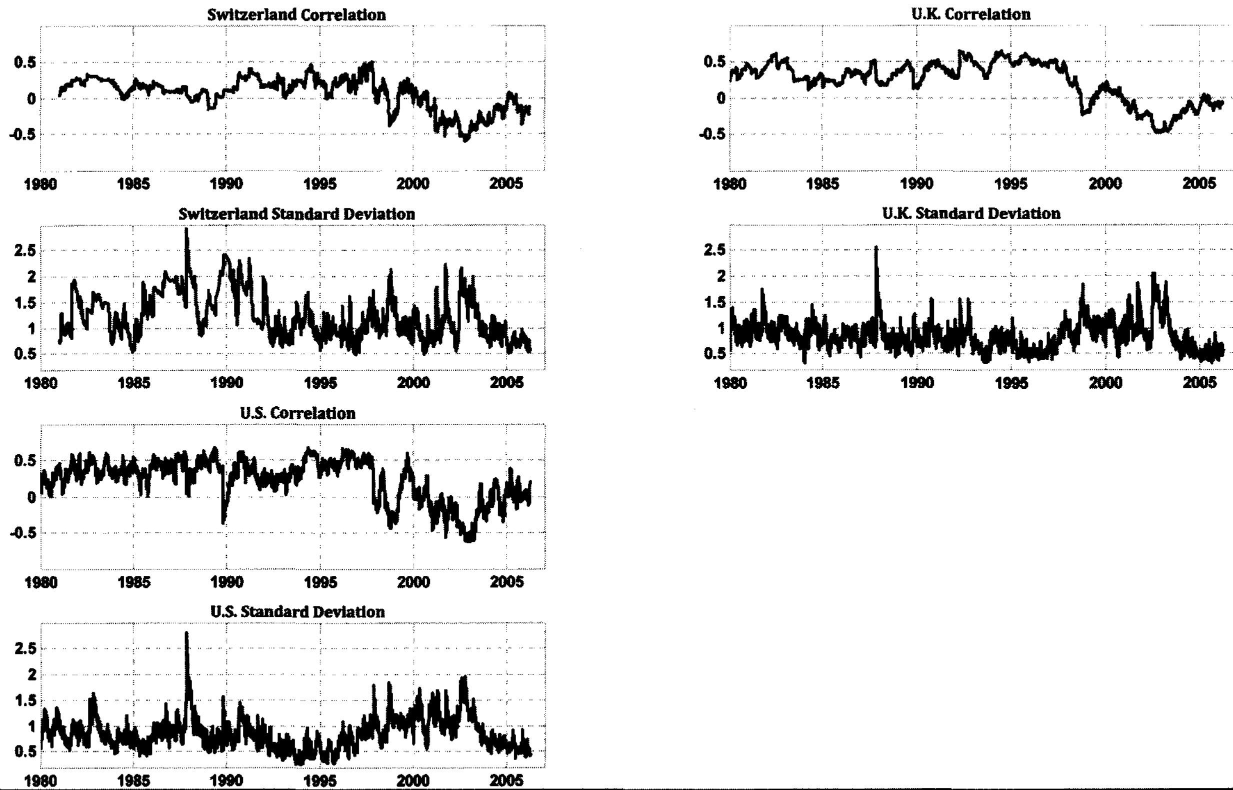
For the international case, a simple regression shows that comovement between several domestic stock markets and U.S. bonds does indeed change. When there is a change, it appears to benefit investors who wish to keep themselves well hedged; in other words, comovement tends to be lower, even negative, during times of high volatility.

Figure 1a
Stock Market Standard Deviations and Stock-Bond Correlations



The standard deviations are in logs. The estimates are from a DCC-GARCH(1,1) model; see Engle (2002) for details.

Figure 1b
Stock Market Standard Deviations and Stock-Bond Correlations



The standard deviations are in logs. The estimates are from a DCC-GARCH(1,1) model; see Engle (2002) for details.

1.1. Literature and Background. What do we know about stock-bond comovement? Several researchers have shown that comovement is, in fact, time-varying. For instance, Scruggs and Glabadanidis (2003) use the Asymmetric Dynamic Covariance (ADC) model of Kroner and Ng (1998) to model the second moments of stock and bond returns, and reject that the correlation is constant. Cappiello, Engle and Sheppard (2003) use the Dynamic Conditional Correlation GARCH (DCC-GARCH) model, developed by Engle (2002), to study the correlation between a number of stock indices (FTSE All-World) and five-year government bond indices. They find, among other things, that high stock market volatility seems to coincide with low stock-bond correlation, and suggest that this is evidence of flight-to-quality. Gulko (2002) studies the joint behavior of U.S. stocks and bonds. He shows that after a stock market crash (defined as a day when the S&P500 loses more than 5% of its value), the correlation is usually negative, and argues that U.S. bonds therefore provide a good hedge against stock market crises. His paper, just like that by Cappiello, Engle and Sheppard, points to the possibility that high volatility, which often coincides with crisis periods, could cause lower, or negative, correlation.

The paper most akin to the present one is the paper by Connolly, Stivers and Sun (2005). They examine whether time-variation in the comovement of daily U.S. stock and government bond returns can be linked to stock market uncertainty, as measured by the Chicago Board Option Exchange's Volatility Index[1] (VIX) and stock turnover. They find that high uncertainty predicts lower, possibly negative, future correlation of stock and bond returns. Furthermore, they find that on days when the implied volatility of equity index options increases, bond returns tend to be high relative to stock returns. I describe their methodology in some detail in Sections 3.1 and 3.2.2. Briefly, they run a regression of the type

$$\text{“bond returns”} = \alpha \times \text{“stock returns”} \times \text{"volatility"} + \beta \times \text{“controls,”}$$

to support the first claim (α comes out significantly negative), and sort data into percentiles w.r.t. daily changes in VIX to support the second (the in-sample correlation is lower in the percentile with the largest changes). The former approach is similar to what I do in this paper, and my main contribution here is to show that the results are largely the same in other countries. In Sections 3.2.1 and 3.2.2, I argue that the latter approach is inappropriate. My main criticism is of the way Connolly, Stivers and Sun analyze changes. The underlying argument why changes in volatility would affect comovement is that it is due to flight-to-quality: Large increases in volatility lower comovement and could possibly make it negative. But large decreases in volatility should, conversely, induce flight-*from*-quality, which should also lower the comovement. Hence, positive and negative changes of similar sizes should have similar impact on comovement. However, in their analysis Connolly, Stivers and Sun use raw changes in volatility as explanatory variable when analyzing comovement, and this approach implicitly groups negative changes with small positive changes. I, instead, first analyze the impact from the absolute value of changes and subsequently extend the analysis by introducing a much more flexible econometric

[1] The Volatility Index is constructed so as to represents the implied volatility of an at-the-money option on the S&P 100 index with one month to expiration.

framework. This framework allows me to directly test which of the two approaches that is most supported by the data.

The previous papers rely little on financial theory, but many papers do use theory as the starting point of their analysis. Campbell and Ammer (1993) use a present value function to break down excess stock returns into news about future dividends, the real rate of return, inflation expectations, and future excess returns. They then go on to study what drives stock-bond correlation. However, as Connolly, Stivers and Sun point out, this method cannot easily explain negative stock-bond comovement if inflation expectations do not change. As inflation has been rather steady since the mid-1980s, and periods of negative comovement long, the approach of Campbell and Ammer seems unable to tell the whole story. I argue, as do Connolly, Stivers and Sun, that volatility is part of that untold story. Shiller and Beltratti (1992), in a similar study to that of Campbell and Ammer, forecast the fundamental variables, and then use a dynamic present value formula to calculate a theoretically motivated stock-bond correlation. This theoretical correlation turns out to be very small at 0.064, whereas the estimated unconditional correlation is 0.366. They conclude that the correlation is "too high" to be explained by fundamentals. Again, they do not consider any impact from volatility.

The literature on financial contagion is also related to the stock-bond comovement literature. Financial contagion is, briefly, the phenomenon that if a financial shock occurs in one country's financial market this shock will cause changes in financial markets in other countries. In other words, the comovement between stock markets changes (they become more correlated) because of an increase in risk in the first market. This is similar to what we study here: the comovement between a stock market and a bond market changes (they become *less*, and possibly negatively, correlated) because of an increase in risk in the stock market. A theoretical paper in the financial contagion tradition has been conducted by Kodres and Pritsker (2002). They develop a model where cross-market hedging can cause financial contagion, even if two markets share no fundamentals at all. The underlying idea is that a change in one stock market can influence a portfolio manager's needs in another stock market. And this is true, even if there are no changes at all in the fundamentals of the second market. It is enough that his hedging needs have changed. We may extend this, and notice that the portfolio manager's hedging needs in the bond market might also have changed; hence, the stock-bond comovement might also be affected. An empirical paper, which starts with a similar cross-market hedging argument, is that by Fleming, Kirby and Ostdiek (1998). The authors find volatility linkages between the stock, bond, and T-bill markets. We can regard the present paper as an extension: not only are there volatility linkages, there are also volatility-comovement linkages. A recent study has also been performed by Li (2002), who derives a theoretical stock-bond correlation model for a CRRA investor.

2. Data

The data used in this paper consist of Morgan Stanley Capital International (MSCI) country indices and Datastream constructed 10-year average maturity bond indices (BMXX10Y, where XX is a country code) for seven countries. MSCI country indices are equity total return indices, i.e., they consist of prices plus dividends. All the data come from Datastream and have been converted to daily log returns and annualized by 252 trading days/year.

The seven countries selected are major developed economies: five European countries, the U.S., and Japan. When I use data from only one country at a time, returns are in the domestic currency. When I use data for the U.S. together with data from another country, returns are converted into U.S. dollars. One feature that makes Italy and Japan different from the other economies is their lower credit ratings. Standard and Poor's rate them AA-, while all the others are AAA.

MSCI country indices are constructed in the same way for each country, which makes them easy to compare. Choosing MSCI, however, makes comparisons with the findings of Connolly, Stivers and Sun more difficult, because these authors use stock data from the Center for Research in Securities Prices (CRSP). As we will see, the results are very similar for the U.S. market regardless of whether we use CRSP or MSCI, so it is reasonable to assume that the choice of data has no significant effect on the results.

Indices are calculated from closing prices on domestic markets, and those prices are not collected at the same time (GMT) for different markets. This might cause problems in the international analysis in Section 4. Marten and Poon (2001) have shown that using closing prices leads to a downward bias on correlations between different domestic stocks markets, as compared to sampling prices at the same time (GMT). The main objective here, however, is to study if and when comovement changes sign, i.e., to study coupling and decoupling. Thus the problem is not serious as long as the bias does not affect the sign, although we might lose some power when testing.

The time period varies from case to case depending on the length of the bond data series. I have chosen 1984 as the starting date for most of the series, since this gives enough observations for the estimation of standard deviations described below. The shortest series is for Italy, where the daily bond observations start in 1991. In the regime-switch model, used for the case of changes in volatility (Section 3.2.2), I use 1988 as the starting year. That model is sensitive to outliers, and I want to avoid having the results affected by the October 1987 crash. All time periods are indicated in the result tables.

In addition to stock and bond returns, we also need some measure of market risk. Connolly, Stivers and Sun use the VIX index, which is constructed using implied index option volatility. However, this kind of index is not available for all countries investigated here, and where it is available it is so only for a short period of time. As an alternative, I model stock returns as a GARCH$(1,1)$ process. For this to be useful, we first need to estimate the parameters of that process. To avoid using information from the future, I do the following: For some countries the stock index series is longer than the bond series. If this is the case, I take the 1000 observations immediately before the start of the bond series. If this is not possible, I erase enough bond observations so that we have exactly 1000 stock index observations before the start of the bond series. I then partition the data as:

$$\begin{bmatrix} S_{-999} \\ \vdots \\ S_0 \end{bmatrix}$$
$$\begin{bmatrix} S_1 \\ \vdots \\ S_T \end{bmatrix} \quad \begin{bmatrix} B_1 \\ \vdots \\ B_T \end{bmatrix}$$

where S_t is the stock return and B_t is the bond return at time t. I assume that stock returns evolve according to a GARCH(1, 1):

$$\begin{aligned} S_t &= \beta_0 + \eta_t & (2.1) \\ \eta_t &= y_t \sigma_t & (2.2) \\ y_t &\sim \mathrm{N}(0,1) & (2.3) \\ \sigma_t^2 &= \phi_0 + \phi_1 \eta_{t-1}^2 + \phi_2 \sigma_{t-1}^2. & (2.4) \end{aligned}$$

I estimate the parameters, $\left(\widehat{\beta}_0, \widehat{\phi}_0, \widehat{\phi}_1, \widehat{\phi}_2\right)$, of this process from the first 1000 observations, $S_{-999} \ldots S_0$, and then go on to estimate the conditional standard deviations for the rest of the series, using these parameter estimates. Thus we obtain the series $\{\widehat{\sigma}_t\}_{t=1}^T$, which I treat as an observed series of conditional standard deviations. We can think of this as market participants using a GARCH(1, 1) model to estimate the unobserved volatility.

All bond index return series are graphed in Figure 2, and all stock index return series in Figure 3. Since one of the markets we are studying is the U.S. market, we may also compare results using the VIX and GARCH estimates. Figure 4 shows the VIX[2] and GARCH series together with changes in both series. The correlation between the two series of levels is 0.890, and between the two series of changes it is 0.584. One notable difference between the two series of changes is that large negative changes almost never occur in the GARCH series. This is partly due to the functional form of the GARCH model, which almost forces this property on the estimates.

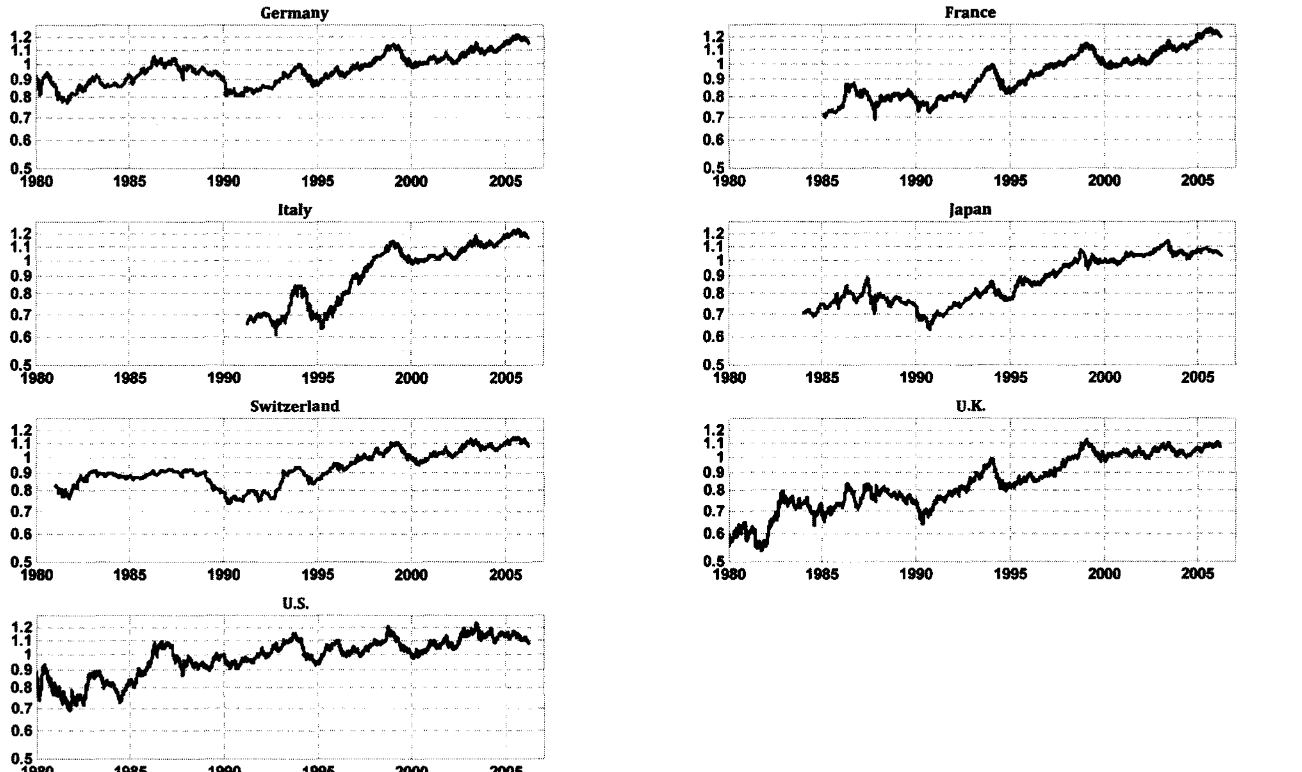
On the market, the prices of stocks and bonds are determined simultaneously, but when modeling I do not specify a joint distribution of stock and bond returns. Instead, I follow Connolly, Stivers and Sun and study the distribution of bonds, conditional on stocks. Note, also, that I do not claim to study causality–stock returns do not cause bond returns. Instead, I characterize the conditional distribution. What could be causing the results is discussed at the end of the paper in Section 5.

Summary statistics for the stock and bond return series are reported in Table 1. They have typical features of financial returns: The mean return on stocks is higher than for bonds, but the standard deviation is also higher. Almost all the series have negative skewness and excess kurtosis; Japan is the only exception, with a skewness of 0.175.

In Table 2 we find unconditional correlations between all stock and bond returns. The correlations between different bond returns range roughly between 0.3 and 0.6 except for Japan, whose correlations are lower at 0.06 to 0.19. Correlations between stock returns range between 0.5 to 0.8, but, again, the correlations between Japanese stocks and other

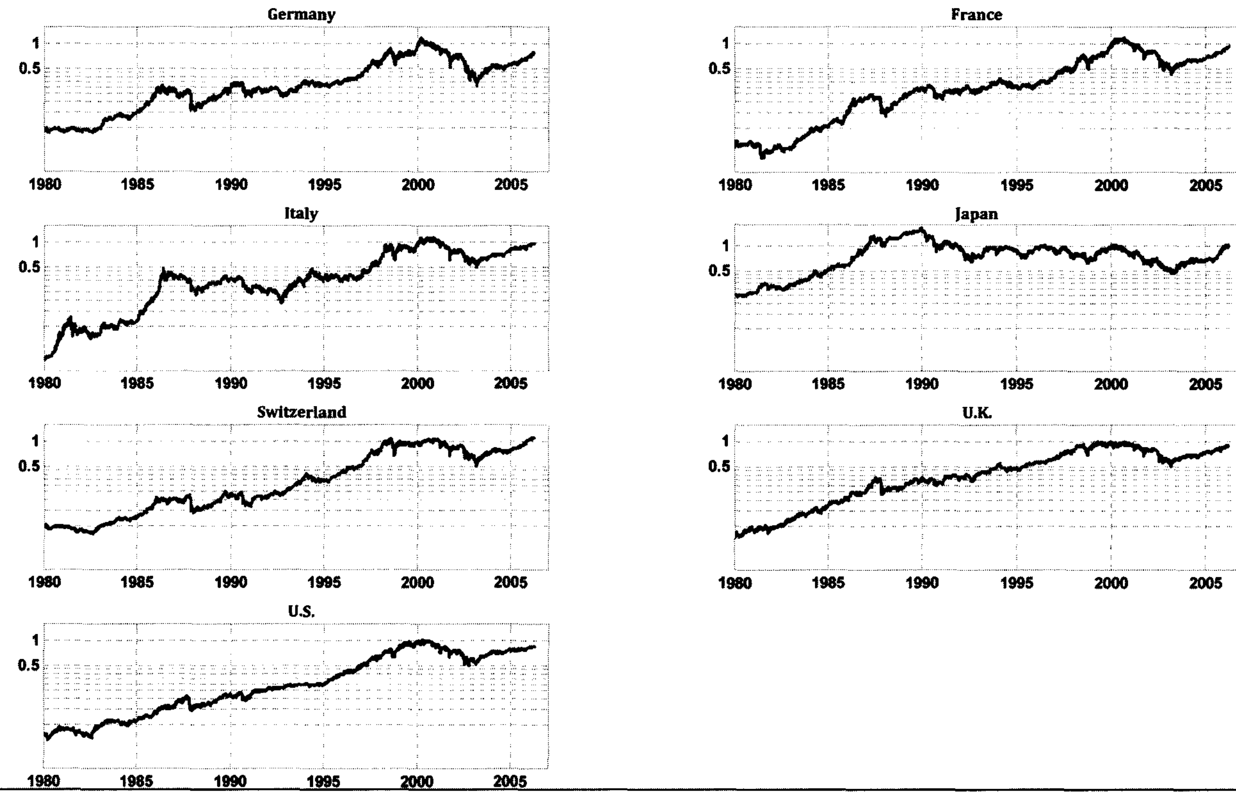
[2] A revised methodology for the VIX was introduced in 2003. The series used here is the old one.

Figure 2
Bond Prices



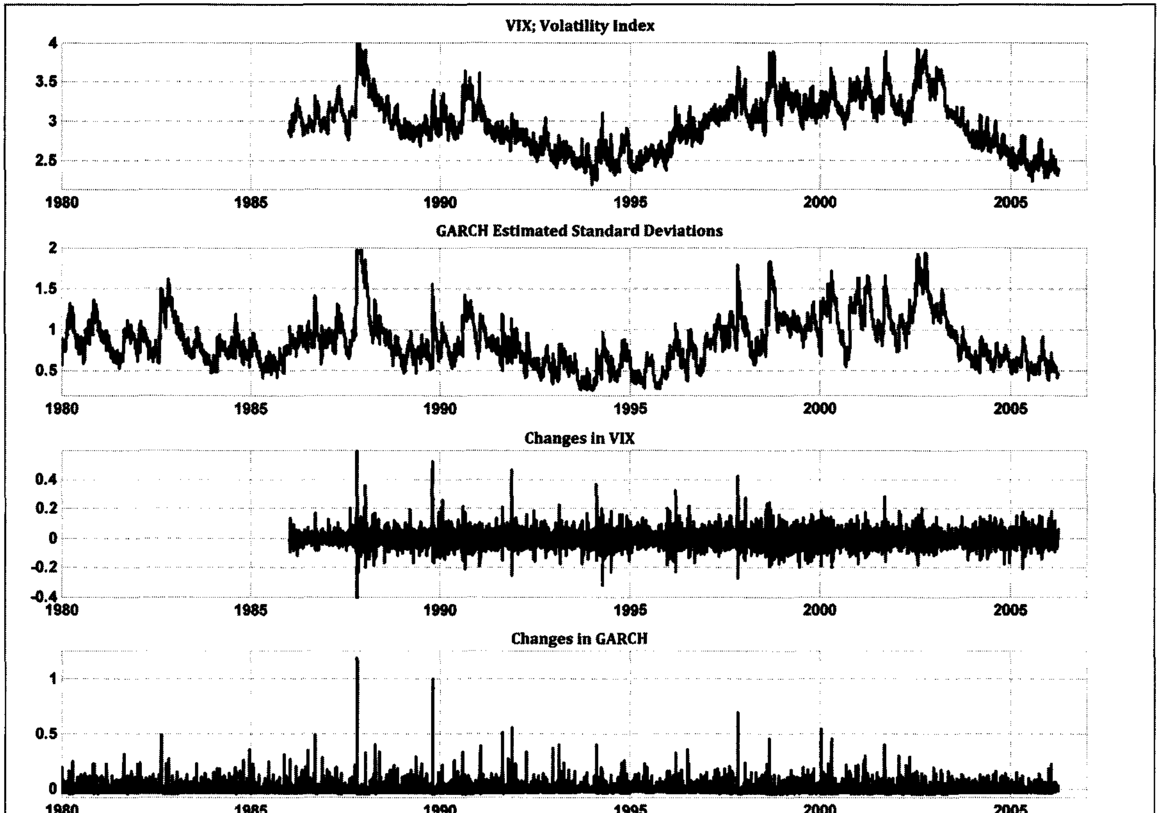
The prices are normalized to equal 1 on Jan 1, 2000. Note the logarithmic scale.

Figure 3
Stock Prices



The prices are normalized to equal 1 on Jan 1, 2000. Note the logarithmic scale.

Figure 4
Comparing the VIX and the GARCH Series



The first two diagrams show log VIX and log GARCH estimated standard deviations for the U.S. stock market. The correlation between these series is 0.880. The second two diagrams show daily changes in log VIX and daily changes in GARCH estimates ($\hat{\sigma}_t$ - $\hat{\sigma}_{t-1}$).

Table 1
Descriptive Statistics

	Germany B	France B	Italy B	Japan B	Switz. B	U.K. B	U.S. B	Germany S	France S	Italy S	Japan S	Switz. S	U.K. S	U.S. S
start	Jan/1984	Feb/1985	Apr/1991	Jan/1984	Jan/1984	Jan/1984	Jan/1984	Jan/1984	Feb/1985	Apr/1991	Jan/1984	Jan/1984	Jan/1984	Jan/1984
mean	0.0155	0.0239	0.0437	0.0247	0.0175	0.0261	0.0123	0.0867	0.106	0.0820	0.0504	0.160	0.0901	0.0949
median	0.0723	0.0504	0.0711	0.0783	0.0863	0.0769	0.0764	0.211	0.153	0.135	-0.0258	0.254	0.159	0.146
min	-4.88	-5.96	-5.92	-5.61	-5.47	-5.95	-6.26	-26.5	-24.8	-20.9	-18.2	-45.1	-26.9	-57.5
max	4.28	5.28	6.35	5.73	5.27	6.72	6.20	18.7	20.3	17.7	26.9	28.7	17.9	21.7
st.d.	0.835	1.07	1.04	0.919	0.897	1.09	1.16	3.54	3.28	3.40	3.34	3.62	2.57	2.71
skewness	-0.297	-0.0739	-0.284	-0.309	-0.371	-0.0598	-0.172	-0.355	-0.326	-0.168	0.175	-1.18	-0.443	-1.96
kurtosis	5.37	5.32	6.96	7.86	6.76	6.00	4.69	7.38	7.52	5.49	7.04	20.8	8.49	44.8

"B" indicates bond return and "S" indicates stock return. Returns are daily log changes, annualized by 252 trading days/year.

Table 2
Correlation Matrix Apr/1991—Mar/2006

	Ge. B	Fr. B	It. B	Jp. B	Sw. B	U.K. B	U.S. B	Ge. S	Fr. S	It. S	Jp. S	Sw. S	U.K. S
Fr. B	0.621												
It. B	0.565	0.525											
Jp. B	0.190	0.117	0.060										
Sw. B	0.577	0.442	0.383	0.112									
U.K. B	0.482	0.530	0.452	0.142	0.410								
U.S. B	0.371	0.372	0.298	0.151	0.322	0.493							
Ge. S	0.014	-0.009	0.060	-0.071	-0.086	-0.028	-0.155						
Fr. S	-0.013	0.051	0.073	-0.064	-0.077	0.040	-0.106	0.795					
It. S	0.018	0.037	0.227	-0.066	-0.031	0.013	-0.095	0.651	0.687				
Jp. S	-0.077	-0.044	0.009	-0.197	-0.074	-0.011	-0.048	0.279	0.321	0.223			
Sw. S	-0.014	0.004	0.041	-0.058	-0.057	0.001	-0.084	0.732	0.733	0.590	0.268		
U.K. S	-0.036	0.016	0.038	-0.042	-0.082	0.088	-0.087	0.703	0.771	0.595	0.284	0.706	
U.S. S	0.007	0.003	0.020	-0.007	-0.044	0.027	-0.025	0.564	0.533	0.387	0.209	0.508	0.519

"B" refer to bond returns and "S" to stock returns.

countries' stocks are much lower at about 0.2 to 0.3. Correlations between stocks and bonds are generally quite low, ranging for the most part between −0.1 and 0.1. Notably, the domestic correlation between Japanese stocks and bonds is −0.197, while the correlation between Italian stocks and bonds is 0.227. These values, and also U.S. bond-German stock correlation at −0.155, are distinctly different from those of the other countries.

3. Domestic Comovement

In this section, I present statistical models and empirical results. First, I consider the relation between stock-bond comovement and level of volatility in seven domestic markets. Second, I do the same using change in volatility instead of level of volatility, and also extend the econometric framework. In Section 4, I consider comovement between domestic stocks and U.S. bonds and their relation to domestic stock volatility.

3.1. The Level of Volatility and Stock-Bond Comovement. The question I want to answer here is whether stock-bond comovement varies with the level of volatility. One of the most straightforward ways to test this is to regress bond returns on the product of stock returns and volatility and some control variables. Should the parameter on $\log(\widehat{\sigma}_t) \times S_t$, in equation (3.1) below, come out significant, I reject both that the comovement is constant and that it is independent of the level of volatility. The model is

$$B_t = \alpha_0 + \alpha_1 B_{t-1} + \alpha_2 S_t + \alpha_3 \log(\widehat{\sigma}_t) \times S_t + \varepsilon_t, \tag{3.1}$$

$$\varepsilon_t \sim \text{GARCH}(1,1), \tag{3.2}$$

where B_t is the bond return, S_t is the stock return, and $\widehat{\sigma}_t$ is the estimated standard deviation from (2.4). B_{t-1} is included as a control for possibly omitted other variables, and S_t alone is included to allow for a constant contribution from S_t to B_t. I take the log of $\widehat{\sigma}_t$ to make the distribution of this variable less skewed. In (3.1) we can think of the contribution from S_t to B_t as time-varying, with the time-variation picked up by a time-varying coefficient on S_t equal to $\alpha_2 + \alpha_3 \log(\widehat{\sigma}_t)$. The error term is modeled as GARCH$(1,1)$ and estimates are maximum likelihood.

The model is close, but not identical, to the one used by Connolly, Stivers and Sun. They use

$$B_t = a_0 + (a_1 + a_2 \log(VIX_{t-1}) + a_3 CV_{t-1})S_t + \varepsilon_t, \tag{3.3}$$

where VIX_t is the level of the VIX at time t. They estimate three different versions of this model. In the first, CV_t is omitted; in the second, it equals estimated stock-bond correlation; and in the third, it equals a dummy for crisis periods. Since we do not have any useful equivalent to the VIX for the economies in this study, and we do not address the impact of crisis periods, I omit this variable completely. The results in Connolly, Stivers and Sun are not in any essential way due to the inclusion or omission of CV_t. Results from estimating (3.1)-(3.2) are presented in Table 3. The parameter of most interest here is α_3; we can note that the estimate is negative and significant for all countries except

Japan, whose estimate is very low in comparison. For almost all countries $\widehat{\alpha}_2$ is positive, but for Japan, again, this estimate has the opposite sign and is insignificant.[3]

To illustrate how the total marginal contribution from S_t to B_t varies with the estimated risk, I plot $(\widehat{\alpha}_2 + \widehat{\alpha}_3 \log(\widehat{\sigma}_t))$ against different realized values of $\widehat{\sigma}_t$ in Figure 5. For Germany, France, Italy, the U.K., and the U.S., between 5% and 10% of the values of $\widehat{\sigma}_t$ have a corresponding marginal contribution below zero. Switzerland has 51% and Japan has 100%. In fact, for Japan both the values of $\widehat{\alpha}_2$ and $\widehat{\alpha}_3$ are so low that it is hard to see any impact at all. The U.S. is the market that is most sensitive to changes in stock market volatility (i.e., the slope is steeper), closely followed by the U.K. Consequently, the changes in hedging opportunities are most dramatic for the U.K. and the U.S., and, particularly in those two countries, the benefits of hedging in bonds increase as volatility increases.

3.1.1. *Robustness and Diagnostics.* For robustness, I perform a series of checks. First, I split the samples into two parts: The first contains data from the start of each series through 1994, and the second contains data from January 1995–April 2006. I then rerun the estimations on the sub-samples. The results are reported in Table 4. The most notable difference is that the t-statistics for the first sub-sample are lower than for the second. Looking at Figures 1a and 1b again, it is likely that this is partly because the comovement does not vary as much during the first sub-sample. In the case of Italy, the estimate of α_3 is significant in the first sub-sample, but has the opposite sign.

Second, I test the predictive power of the model (conditional on S_t) in an out-of-sample test, and compare it to a model with no nonlinear term:

$$B_t = \alpha_0 + \alpha_1 B_{t-1} + \alpha_2 S_t + \varepsilon_t, \tag{3.4}$$

$$\varepsilon_t \sim \text{GARCH}(1,1). \tag{3.5}$$

The out-of-sample period is January 2002–April 2006, and the parameter estimates are updated for each prediction. I report the Clark and McCracken (2001) statistic for the improvement.[4] A 95% significance level of testing H_0 : No improvement with this statistic, is 0.520. All countries, except Japan, exceed this (see Table 5). Consequently, we reject that there is no improvement for all countries but Japan. Third, I test if GARCH$(1,1)$ is an adequate model for the data by testing H_0 : No remaining GARCH. (See Lundbergh and Teräsvirta (2002) for details.) Reported numbers are p-values. Only for Germany and Italy do we reject that there is no remaining GARCH in the errors.

3.1.2. *An Alternative Model.* There is always a risk that econometric results are model driven. For this reason I also test if the results are robust to the choice of econometric model, and estimate an alternative model that also allows for nonlinearities that depend on volatility. One such model is the simple Threshold Regression (TR; see Franses and

[3] For comparison, I also estimated the model of Connolly, Stivers and Sun (omitting the CV-term), using MSCI instead of CRSP stock data. This corresponds to estimating (3.1) with the restriction $\alpha_1 = 0$. I did this for the same period as they do (1986-2000), first with VIX and then with GARCH standard deviations as a volatility measure. The results were very similar. As an example: With VIX $\widehat{\alpha}_3 = -0.19$, and with GARCH $\widehat{\alpha}_3 = -0.20$. I conclude that the change from VIX to GARCH estimated standard deviations is inessential for the phenomenon I study.

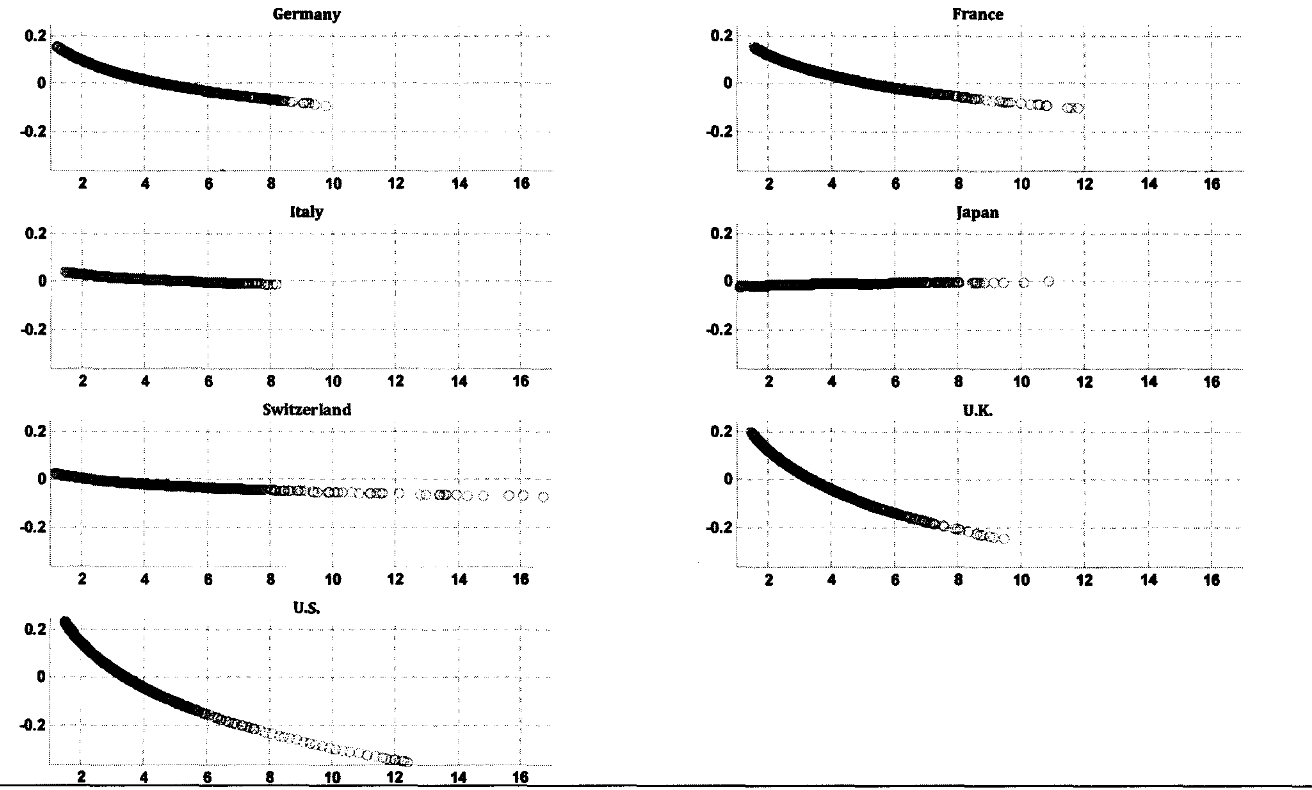
[4] The asymptotic distribution is non-standard. For details, see Clark and McCracken (2001).

Table 3
Parameter Estimates for Model (3.1)-(3.2)

Country	Period	#obs	α_0	α_1	α_2	α_3	γ_0	γ_1	γ_2	LLF
Germany	Jan/1984—Mar/2006	5233	**0.0299**	**0.0354**	**0.177**	**-0.117**	**0.00948**	**0.0691**	**0.917**	-5943.5
			(3.24)	**(2.57)**	**(18.3)**	**(-16.3)**	**(4.42)**	**(9.04)**	**(106)**	
France	Feb/1985—Mar/2006	4932	**0.0250**	-0.0185	**0.207**	**-0.125**	**0.0298**	**0.104**	**0.870**	-6797.3
			(1.96)	(-1.20)	**(13.0)**	**(-8.87)**	**(5.30)**	**(8.56)**	**(63.5)**	
Italy	Apr/1991—Mar/2006	3731	**0.0393**	**0.119**	**0.0499**	**-0.0298**	**0.00571**	**0.0652**	**0.931**	-4894.2
			(3.18)	**(6.47)**	**(2.83)**	**(-2.17)**	**(2.62)**	**(6.95)**	**(103)**	
Japan	Jan/1984—Mar/2006	4731	**0.0404**	**0.0423**	-0.0216	0.00890	**0.0167**	**0.111**	**0.873**	-5668.8
			(3.93)	**(2.63)**	(-1.71)	(0.86)	**(3.51)**	**(7.92)**	**(54.2)**	
Switzerland	Jan/1984—Mar/2006	3425	**0.0301**	**0.0679**	**0.0299**	**-0.0358**	**0.0204**	**0.0814**	**0.89**	-4143.9
			(2.44)	**(3.70)**	**(2.69)**	**(-4.78)**	**(3.91)**	**(5.28)**	**(47.8)**	
U.K.	Jan/1984—Mar/2006	5348	0.024	**0.0520**	**0.288**	**-0.23**	**0.011**	**0.0442**	**0.945**	-7511.4
			(1.94)	**(3.77)**	**(18.5)**	**(-17.8)**	**(3.07)**	**(6.09)**	**(98.4)**	
U.S.	Jan/1984—Mar/2006	5366	0.00587	**0.0713**	**0.334**	**-0.272**	**0.0257**	**0.0465**	**0.933**	-8059.2
			(0.41)	**(5.14)**	**(8.34)**	**(-7.14)**	**(3.88)**	**(5.99)**	**(90.7)**	

A significant α_3 indicates time-varying comovement w.r.t. σ_t. Numbers in parentheses are Bollerslev and Wooldridge t-values.
The model is $B_t = \alpha_0 + \alpha_1 B_{t-1} + \alpha_2 S_t + \alpha_3 \ln(\sigma_t) S_t + \varepsilon_t$; $\varepsilon_t = z_t h_t$, $z_t \sim N(0,1)$, $h^2_t = \gamma_0 + \gamma_1 \varepsilon^2_{t-1} + \gamma_2 h^2_{t-1}$.

Figure 5
Domestic Stock-Bond Comovement w.r.t. $\hat{\sigma}_t$



A plot of the total marginal contribution from S_t to B_t as a function of $\hat{\sigma}_t$ (i.e., $\alpha_2 + \alpha_3\ln(\hat{\sigma}_t)$ as a function of $\hat{\sigma}_t$); see model (3.1)–(3.2). The model is $B_t = \alpha_0 + \alpha_1 B_{t-1} + \alpha_2 S_t + \alpha_3\ln(\hat{\sigma}_t)S_t + \varepsilon_t$, $\varepsilon_t = z_t h_t$, $z_t \sim N(0,1)$, $h^2_t = \gamma_0 + \gamma_1\varepsilon^2_{t-1} + \gamma_2 h^2_{t-1}$.

Table 4
Subsample Tests for Model (3.1)-(3.2)

Country	Period	#obs	α_0	α_1	α_2	α_3	γ_0	γ_1	γ_2	LLF
Germany	Jan/1984—Dec/1994	2439	0.0210	**0.0573**	**0.175**	**-0.0861**	**0.0100**	**0.097**	**0.890**	-2657
			(1.71)	**(2.89)**	**(10.9)**	**(-6.63)**	**(3.17)**	**(6.48)**	**(58.9)**	
	Jan/1995—Mar/2006	2793	**0.0411**	0.0019	**0.137**	**-0.141**	**0.0093**	**0.0383**	**0.947**	-3272.6
			(2.89)	(0.10)	**(6.30)**	**(-6.67)**	**(2.22)**	**(4.44)**	**(71.7)**	
France	Feb/1985—Dec/1994	2330	0.0074	**0.0630**	**0.145**	0.0088	**0.0238**	**0.141**	**0.839**	-2983.4
			(0.46)	**(3.15)**	**(6.38)**	(0.41)	**(4.20)**	**(7.62)**	**(48.5)**	
	Jan/1995—Mar/2006	2602	**0.0468**	**-0.130**	**0.165**	**-0.160**	**0.0420**	**0.0713**	**0.889**	-3667.9
			(2.49)	**(-5.92)**	**(4.71)**	**(-4.79)**	**(3.55)**	**(4.88)**	**(43.1)**	
Italy	Apr/1991—Dec/1994	916	0.0162	**0.160**	-0.085	**0.173**	0.0140	**0.110**	**0.889**	-1414.6
			(0.58)	**(4.08)**	(-1.93)	**(4.45)**	(1.87)	**(4.58)**	**(37.0)**	
	Jan/1995—Mar/2006	2816	**0.0437**	**0.107**	**0.0913**	**-0.0671**	2e-7	**0.0905**	**0.909**	-3435.5
			(3.23)	**(5.25)**	**(3.64)**	**(-3.75)**	(9e-5)	**(7.09)**	**(68.3)**	
Japan	Jan/1984—Dec/1994	2117	**0.0423**	**0.0617**	**0.0656**	-0.0060	**0.0185**	**0.108**	**0.877**	-2720.6
			(2.85)	**(2.48)**	**(3.41)**	(-0.42)	**(2.46)**	**(6.29)**	**(44.5)**	
	Jan/1995—Mar/2006	2613	**0.0372**	0.0095	**-0.0842**	0.0313	**0.0153**	**0.105**	**0.872**	-2838.7
			(2.89)	(0.45)	**(-4.05)**	(1.82)	**(2.81)**	**(5.15)**	**(35.2)**	
Switzerland	Jan/1984—Dec/1994	998	-0.0013	-0.0099	**0.103**	**-0.0452**	**0.122**	**0.342**	**0.592**	-1405.5
			(-0.05)	(-0.28)	**(5.24)**	**(-3.53)**	**(2.43)**	**(5.51)**	**(7.76)**	
	Jan/1995—Mar/2006	2426	**0.0305**	**0.108**	0.0112	**-0.0474**	**0.0058**	**0.0382**	**0.951**	-2661.2
			(2.21)	**(5.26)**	(0.74)	**(-3.30)**	**(2.20)**	**(4.31)**	**(81.1)**	
U.K.	Jan/1984—Dec/1994	2553	0.0125	**0.0647**	**0.394**	**-0.245**	**0.0204**	**0.0692**	**0.917**	-3763.4
			(0.66)	**(3.32)**	**(10.2)**	**(-5.80)**	**(3.03)**	**(5.78)**	**(68.5)**	
	Jan/1995—Mar/2006	2794	0.0277	0.0294	**0.100**	**-0.129**	**0.0054**	**0.0277**	**0.965**	-3631.8
			(1.71)	(1.60)	**(5.61)**	**(-8.57)**	**(2.05)**	**(4.77)**	**(132.2)**	
U.S.	Jan/1984—Dec/1994	2627	-0.0142	**0.0736**	**0.329**	**-0.177**	**0.0463**	**0.0640**	**0.897**	-3859.2
			(-0.74)	**(3.90)**	**(6.97)**	**(-3.26)**	**(2.83)**	**(4.38)**	**(38.9)**	
	Jan/1995—Mar/2006	2737	0.0160	**0.0632**	**0.363**	**-0.407**	**0.0231**	**0.0370**	**0.943**	-4062.5
			(0.81)	**(3.33)**	**(12.9)**	**(-14.6)**	**(2.28)**	**(3.65)**	**(58.5)**	

A significant α_3 indicates time-varying comovement w.r.t. σ_t. Numbers in parentheses are Bollerslev and Wooldridge t-values.
The model is $B_t = \alpha_0 + \alpha_1 B_{t-1} + \alpha_2 S_t + \alpha_3 \ln(\sigma_t) S_t + \varepsilon_t$; $\varepsilon_t = z_t h_t$, $z_t \sim N(0,1)$, $h^2_t = \gamma_0 + \gamma_1 \varepsilon^2_{t-1} + \gamma_2 h^2_{t-1}$.

Table 5
Diagnostic Tests for Model (3.1)-(3.2)

	Germany	France	Italy	Japan	Switzerland	U.K.	U.S.
$MSFE_L$	**0.77538**	**1.48269**	**0.59283**	0.53214	**0.60714**	**0.82147**	**1.67505**
$MSFE_{N-L}$	**0.68364**	**1.32305**	**0.59365**	0.53362	**0.60560**	**0.71317**	**1.51311**
Δ	**-0.118**	**-0.108**	**0.001**	0.003	**-0.003**	**-0.132**	**-0.097**
C & McC	**129.27**	**88.826**	**3.8133**	-1.2844	**4.4440**	**194.02**	**79.759**
No rem. GARCH	**0.009**	0.055	0.377	0.341	0.245	0.336	0.363
D-W	1.971	2.027	1.974	1.938	2.017	1.938	1.996

The out-of-sample period is from Jan/2002—Mar/2006 for all countries. The parameters were updated each period in the out-of-sample tests.
"$MSFE_L$": Mean Squared Forecast Error for the Linear Model. The model is the same as for $MSFE_{N-L}$ below, but with the restriction $\alpha_3 = 0$.
"$MSFE_{N-L}$": Mean Squared Forecast Error for the Non-Linear Model. The model is $B_t = \alpha_0 + \alpha_1 B_{t-1} + \alpha_2 S_t + \alpha_3 \ln(\sigma_t) S_t + \varepsilon_t$.
"Δ": Change in MSFE, i.e., $(MSFE_{N-L} - MSFE_L)/MSFE_L$.
"C & McC": This is the Clark and McCracken statistic for equal forecast accuracy in nested models with parameter updating. A 95 % confidence interval for H_0: No improvement is given in Clark and McCracken (2001) to be 0.520. We reject H_0: No improvement for all countries except Japan.
"No rem. GARCH": p-values for H_0: No remaining GARCH structure in the errors, according to Lundbergh and Teräsvirta (2001). For Germany we reject H_0.
"D-W": Durbin-Watson statistic for the standardized residuals.

van Dijk (2000) for an overview of this type of model):

$$B_t = \begin{cases} \alpha_{01} + \alpha_1 B_{t-1} + \alpha_{21} S_t + \varepsilon_t, & \widehat{\sigma}_t \leq \delta \\ \alpha_{02} + \alpha_1 B_{t-1} + \alpha_{22} S_t + \varepsilon_t, & \widehat{\sigma}_t > \delta \end{cases} \tag{3.6}$$

$$\varepsilon_t \sim \mathrm{N}\left(0, \sigma_\varepsilon^2\right). \tag{3.7}$$

The model allows the marginal contribution from S_t to be different depending on whether $\widehat{\sigma}_t$ is higher or lower than some threshold δ. The intercept term is allowed to differ between regimes as well, but the autocorrelation parameter is restricted to be the same. The null hypothesis is then $\mathrm{H}_0 : \alpha_{21} = \alpha_{22}$.[5] The distribution is nonstandard, so to test the null I use 1000 bootstrap simulations for the likelihood ratio $T \times \log(SSE_r/SSE_u)$, where SSE_u is the sum of squared errors from the unrestricted regression above, and SSE_r is the same but under the restriction $\alpha_{21} = \alpha_{22}$. The results are summarized in Table 6. Japan has a high p-value, and we do not reject the null hypothesis for this country. Switzerland has a p-value of 0.8%. For the other countries we strongly reject the null hypothesis, and I conclude that the results from model (3.1)-(3.2) are unlikely to be model driven. We can also note that the estimates of α_{21} are positive for all countries, while the estimates of α_{22} are negative for all but Japan.

3.2. Changes in Volatility and Stock-Bond Comovement.

3.2.1. *A Linear Model.* The idea that changes in volatility are connected to stock-bond comovement is often mentioned in the financial press and is connected to the idea of flight-to-quality. I will discuss this more extensively in Section 5.2 but, briefly, the idea is that an increase in stock volatility makes stocks less attractive relative to bonds—and the larger the increase, the larger the change in relative attractiveness. As a consequence, stock and bond prices move in opposite directions, resulting in negative comovement. The opposite phenomenon, flight-from-quality, occurs when stock volatility decreases: stocks become more attractive relative to bonds, and the result is, similarly, negative comovement. Hence, it is the size of the change which is important and not whether it is positive or negative—positive and negative changes of similar sizes should have similar impact on comovement. In contrast to Connolly, Stivers and Sun, I therefore use the absolute value of changes, $\mathrm{abs}\left(\Delta\widehat{\sigma}_{t+1}\right)$, as the main explanatory variable. However, to make comparison possible I also include separate tests using raw changes, $\Delta\widehat{\sigma}_{t+1}$, but, as we will see, the results from those tests are less convincing.[6]

---

[5] The restriction on B_{t-1} does not change the results regarding H_0. We reject for the same countries even without this restriction.

[6] The time subscripts here demand a comment as it looks like the regression incorporates information from the future, $t+1$. If flight-to/from-quality is a cause for changes in comovement it should be a more or less instant reaction to a change in volatility. Suppose there is a large change in volatility in day t, possibly causing a change in time t comovement. Since we are using GARCH to estimate volatility, this change will show up in our estimate of volatility for the next day, $t+1$. The estimate is consequently done with information up to and including time t but is an estimate for time $t+1$, and the convention is to use the time subscript $t+1$ for the estimate (see Function (2.4)). Since we here want an estimate of the change in volatility over day t, we have to use the difference in estimated volatility for time t and $t+1$: $\Delta\widehat{\sigma}_{t+1} = \widehat{\sigma}_{t+1} - \widehat{\sigma}_t$.

Table 6
Likelihood Ratio Tests for a Regime Shift Depending on Level Volatility for Domestic Markets; Model (3.6)-(3.7)

Country	Period	#obs	α_{01}	α_1	α_{21}	α_{02}	α_{22}	δ	LR-stat	p-value
Germany	Jan/1984—Mar/2006	5233	-0.00982 (-0.77)	**0.0368 (2.75)**	**0.0808 (17.3)**	**0.067 (2.79)**	**-0.0207 (-4.80)**	3.46	244.90	0
France	Feb/1985—Mar/2006	4932	-0.00739 (-0.44)	**-0.0451 (-3.24)**	**0.100 (15.4)**	**0.0950 (2.89)**	-9.71e-5 (-0.015)	3.31	119.25	0
Italy	Apr/1991—Mar/2006	3731	0.0244 (1.40)	**0.105 (6.65)**	**0.0791 (14.0)**	**0.13 (2.64)**	**-0.0221 (-2.35)**	4.67	84.178	0
Japan	Jan/1984—Mar/2006	4731	-0.0658 (-1.29)	**0.0452 (3.12)**	**0.120 (4.05)**	**0.0298 (2.16)**	1.03e-5 (0.002)	1.60	12.977	0.071
Switzerland	Jan/1984—Mar/2006	3425	0.0119 (0.71)	0.00516 (0.30)	**0.0210 (3.50)**	0.0327 (0.89)	**-0.023 (-3.88)**	3.76	24.391	0.008
U.K.	Jan/1984—Mar/2006	5348	0.0079 (0.52)	**0.0390 (2.99)**	**0.137 (20.4)**	**0.107 (2.34)**	**-0.0938 (-9.62)**	3.23	362.64	0
U.S.	Jan/1984—Mar/2006	5366	-0.0100 (-0.58)	**0.0608 (4.63)**	**0.155 (18.0)**	0.033 (1.06)	**-0.0630 (-8.24)**	2.75	360.15	0

The model is: $B_t = \{\alpha_{01} + \alpha_1 B_{t-1} + \alpha_{21} S_t\}_{\sigma t \leq \delta} + \{\alpha_{02} + \alpha_1 B_{t-1} + \alpha_{22} S_t\}_{\sigma t > \delta} + \varepsilon_t$; $\varepsilon_t \sim N(0, \sigma^2_t)$.
The p-values are estimated by performing 1000 bootstrap simulations of the likelihood ratio $LR = T^*\log(SSE_r/SSE_u)$, where SSE_r and SSE_u are the sum of squared errors with and without the restriction $\alpha_{21} = \alpha_{22}$. We reject H_0 only for Japan.

Similar to before, one of the most straightforward ways to test if there is any impact from changes is:

$$B_t = \alpha_0 + \alpha_1 B_{t-1} + \alpha_2 S_t + \alpha_3 \operatorname{abs}(\Delta\widehat{\sigma}_{t+1}) S_t + \varepsilon_t, \tag{3.8}$$

$$\varepsilon_t \sim \mathrm{GARCH}(1,1). \tag{3.9}$$

Again, a significant estimate of α_3 would mean that we reject that comovement is independent of changes in volatility. As a comparison, I also include estimates for the model using $\Delta\widehat{\sigma}_{t+1}$, instead of $\operatorname{abs}(\Delta\widehat{\sigma}_{t+1})$.

There is little evidence of flight-to-quality in these regressions (see Table 7). For Italy and Japan, the estimates of α_3 are significantly positive, counter to our hypothesis. Only for the U.S., in both columns, do we have a significant value with the right sign. Furthermore, it appears that the choice between $\operatorname{abs}(\Delta\widehat{\sigma}_{t+1})$ or $\Delta\widehat{\sigma}_{t+1}$ is irrelevant: The estimates are, in most cases, almost identical.

3.2.2. *Testing for Changes in Regimes.* There is a potential problem in using a linear model such as (3.8)-(3.9). If the relation between comovement and changes in volatility is nonlinear, the power of such a test could be severely reduced.

Connolly, Stivers and Sun find evidence of reduced, and even negative, comovement using a rather crude method. They calculate the change in VIX ($\Delta VIX_t = VIX_t - VIX_{t-1}$) for each day, and then sort the data into percentiles, from the largest negative changes to the largest positive changes. They then calculate the stock-bond correlation within each percentile. For example, if $\{\Delta VIX_{t1}, \Delta VIX_{t2}, \Delta VIX_{t3}\}$ belong to the 25% smallest changes, the in-sample correlation for that percentile is calculated using $\{S_{t1}, S_{t2}, S_{t3}\}$ and $\{B_{t1}, B_{t2}, B_{t3}\}$.

Table 8 reports the results from performing the same exercise on all countries studied here, using both changes and absolute changes. For comparison, I also include estimates using the VIX on U.S. data, and some of the results from Connolly, Stivers and Sun. Remember that they use CRSP data, so the estimated correlations are not identical.

The use of raw changes does not produce a pattern that would indicate a flight-to-quality type of behavior for any of these countries. With the VIX, we obtain slightly different results with a lower correlation in the highest percentile, but we also obtain a lower correlation in the lowest percentile. Using absolute changes, however, clearly produces the hypothesized pattern for Germany and the U.S. For France and the U.K., there also seems to be some weak evidence in favor, in the sense that the second two percentiles have a lower correlation than the first two. For Italy and Japan, we still do not find the pattern.

If we use the VIX instead of GARCH-volatility, the absolute value of changes still seems to produce results that are more in line with the hypothesis: All the first three percentiles show more or less the same in-sample correlation, whereas the fourth shows a much lower correlation. If we use raw changes, the first percentile shows a low correlation that is not explained by the hypothesis. This pattern is more pronounced in the longer sample period than in the shorter one used by Connolly, Stivers and Sun. I conclude that there is more evidence supporting the use of absolute values of changes than there is for raw changes.

Table 7
Parameter Estimates for Model (3.8)-(3.9)

Country	Period	#obs	Model	α_0	α_1	α_2	α_3	γ_0	γ_1	γ_2	LLF
Germany	Jan/1984—Mar/2006	5233	A	**0.0329**	**0.0396**	**0.0373**	0.0050	**0.0109**	**0.0715**	**0.9130**	-6082
				(3.50)	**(2.80)**	**(10.1)**	(0.60)	**(4.44)**	**(8.70)**	**(97.4)**	
			B	**0.0328**	**0.0395**	**0.0372**	0.0049	**0.0109**	**0.0715**	**0.9130**	-6082
				(3.50)	**(2.86)**	**(9.59)**	(0.58)	**(4.43)**	**(8.70)**	**(97.4)**	
France	Feb/1985—Mar/2006	4932	A	0.0200	-0.0140	**0.0797**	-0.0100	**0.0247**	**0.0971**	**0.8820**	-6847
				(1.77)	(-0.94)	**(13.2)**	(-0.98)	**(4.96)**	**(8.37)**	**(70.0)**	
			B	0.0227	-0.0140	**0.0806**	-0.0110	**0.0248**	**0.0902**	**0.8820**	-6847
				(1.79)	(-0.94)	**(12.1)**	(-1.01)	**(4.96)**	**(8.37)**	**(69.9)**	
Italy	Apr/1991—Mar/2006	3731	A	**0.0400**	**0.1100**	0.0034	**0.0247**	**0.0050**	**0.0618**	**0.9340**	-4890
				(3.23)	**(6.53)**	(0.72)	**(3.19)**	**(2.66)**	**(7.31)**	**(116.2)**	
			B	**0.0399**	**0.1190**	0.0022	**0.0264**	**0.0049**	**0.0616**	**0.9350**	-4890
				(3.22)	**(6.53)**	(0.46)	**(3.30)**	**(2.65)**	**(7.33)**	**(116.7)**	
Japan	Jan/1984—Mar/2006	4731	A	**0.0430**	**0.0427**	**-0.0240**	**0.0174**	**0.0170**	**0.1140**	**0.8690**	-5653
				(4.28)	**(2.66)**	**(-5.65)**	**(2.90)**	**(3.63)**	**(8.11)**	**(53.1)**	
			B	**0.0428**	**0.0424**	**-0.0260**	**0.0190**	**0.0170**	**0.1140**	**0.8690**	-5652
				(4.26)	**(2.65)**	**(-5.69)**	**(3.07)**	**(3.63)**	**(8.10)**	**(53.1)**	
Switzerland	Jan/1984—Mar/2006	3425	A	**0.0373**	**0.0684**	**-0.0180**	0.0046	**0.0226**	**0.0840**	**0.8860**	-4151
				(2.66)	**(3.73)**	**(-3.85)**	(1.05)	**(4.02)**	**(5.37)**	**(46.3)**	
			B	**0.0330**	**0.0685**	**-0.0180**	0.0045	**0.0220**	**0.0840**	**0.8860**	-4152
				(2.65)	**(3.73)**	**(-3.78)**	(1.02)	**(4.02)**	**(5.36)**	**(46.3)**	
U.K.	Jan/1984—Mar/2006	5348	A	0.0249	**0.0594**	**0.0499**	-0.0080	**0.0125**	**0.0470**	**0.9420**	-7653
				(1.91)	**(4.16)**	**(7.32)**	(-0.37)	**(3.32)**	**(6.27)**	**(99.3)**	
			B	0.0248	**0.0593**	**0.0508**	-0.0100	**0.0124**	**0.0470**	**0.9420**	-7652
				(1.90)	**(4.16)**	**(6.67)**	(-0.42)	**(3.3)**	**(6.27)**	**(99.4)**	
U.S.	Jan/1984—Mar/2006	5366	A	0.0025	**0.0632**	**0.0840**	**-0.0240**	**0.0220**	**0.0470**	**0.9360**	-8166
				(0.17)	**(4.61)**	**(13.1)**	**(-2.80)**	**(3.78)**	**(6.34)**	**(94.4)**	
			B	0.0025	**0.0632**	**0.0800**	**-0.0240**	**0.0226**	**0.0460**	**0.9360**	-8166
				(0.17)	**(4.61)**	**(13.1)**	**(-2.79)**	**(3.77)**	**(6.34)**	**(94.5)**	

A significant α_3 indicates time-varying comovement w.r.t. $\Delta\sigma_{t+1}$ (marked A) and abs($\Delta\sigma_{t+1}$) (marked B). Numbers in parentheses are Bollerslev and Wooldridge t-values. The models are (rows A:) $B_t = \alpha_0 + \alpha_1 B_{t-1} + \alpha_2 S_t + \alpha_3(\Delta\sigma_{t+1})S_t + \varepsilon_t$, and (rows B:) $B_t = \alpha_0 + \alpha_1 B_{t-1} + \alpha_2 S_t + \alpha_3 \mathrm{abs}(\Delta\sigma_{t+1})S_t + \varepsilon_t$, $\varepsilon_t = z_t h_t$, $z_t \sim N(0,1)$, $h_t^2 = \gamma_0 + \gamma_1 \varepsilon^2_{t-1} + \gamma_2 h^2_{t-1}$.

Table 8
In-Sample Stock-Bond Correlations for Different Percentiles

Sorted by		$\Delta\sigma_{t+1}$				$abs(\Delta\sigma_{t+1})$			
Percentile		0-0.25	0.25-0.50	0.50-0.75	0.75-1	0-0.25	0.25-0.50	0.50-0.75	0.75-1
Germany	Jan/1984—Mar/2006	0.00743	0.106	0.219	0.124	0.240	0.162	0.142	0.0465
France	Feb/1985—Mar/2006	0.0594	0.103	0.177	0.197	0.175	0.143	0.199	0.135
Italy	Apr/1991—Mar/2006	0.0544	0.0960	0.179	0.240	0.164	0.154	0.100	0.217
Japan	Jan/1984—Mar/2006	0.00948	0.0542	-0.0997	0.0288	-0.091	0.0296	-0.0133	0.0453
Switzerland	Jan/1984—Mar/2006	-0.0803	0.0397	-0.0183	0.0117	-0.0755	0.0181	0.0101	0.0063
U.K.	Jan/1984—Mar/2006	-0.0276	0.128	0.204	0.189	0.206	0.182	0.116	0.140
U.S.	Jan/1984—Mar/2006	0.00378	0.0730	0.214	0.0976	0.225	0.146	0.0911	0.0399
U.S./VIX	Jan/1986—Mar/2006	0.0582	0.245	0.232	-0.0694	0.254	0.245	0.232	-0.0694
U.S./VIX	Jan/1986—Jan/2001	0.253	0.347	0.337	0.0317	0.338	0.369	0.370	0.144
CS&S*	Jan/1986—Jan/2001	0.287	0.361	0.352	0.112	–	–	–	–

The data were sorted w.r.t. $\Delta\sigma_{t+1}$ and abs($\Delta\sigma_{t+1}$), respectively. The in-sample correlation between stock and bond returns was then calculated for each of the four percentiles. For the rows U.S./VIX, the data were sorted w.r.t. ΔVIX_t and abs(ΔVIX_t), instead.
* Results from Connolly, Stivers and Sun (2005), Table 4.

These results lead me to introduce a different approach to model the impact. Instead of the linear model above, I use a Smooth Transition Regression (STR) model. This allows me to test for several different aspects, and not only for time-varying comovement. This, potentially, can help us make better sense of the results from (3.8)-(3.9), and of the Connolly, Stivers and Sun results, too.

The STR model is

$$\begin{aligned}
B_t &= \alpha_1 B_{t-1} + (\alpha_{01} + \alpha_{21} S_t)(1 - G_t) + (\alpha_{02} + \alpha_{22} S_t)(G_t) + \varepsilon_t, && (3.10)\\
G_t &\in \left\{G_t^E, G_t^L\right\}, && (3.11)\\
G_t^E &= 1 - \exp\left\{-\gamma\left(\Delta\widehat{\sigma}_{t+1} - c\right)^2\right\}, && (3.12)\\
G_t^L &= \left(1 + \exp\left\{-\gamma\left(\Delta\widehat{\sigma}_{t+1} - c\right)\right\}\right)^{-1}, && (3.13)\\
\varepsilon_t &\sim \mathrm{N}\left(0, \sigma_\varepsilon^2\right). && (3.14)
\end{aligned}$$

I first provide some intuition for how we can interpret this model, and then describe how it allows me to test for several central properties of the comovement-volatility change relation. For details on this model, see Teräsvirta (1994).

If we start by looking at (3.10), we see that if G_t is either 0 or 1, the model is just an ordinary linear regression. If $G_t = 0$, the parameters are α_1, α_{01} and α_{21}, and if $G_t = 1$, they are α_1, α_{02} and α_{22}. We can think of this as two different regimes, 1 and 2, where the stock-bond relation is linear within each regime, but with different parameters. Which regime we are in is determined by the function G_t, which is bounded on $[0, 1]$. This function is often referred to as a "transition function." As candidates for transition functions, I allow for two different ones: G_t^E is an exponential function, and G_t^L is the logistic function. For typical shapes of these transition functions, see Figure 6. (All plots are of G_t^E, except the last one.) The input to the transition function, G_t, is $\Delta\widehat{\sigma}_{t+1}$. Thus, on a certain day, there is a change in our GARCH estimate of volatility. The transition function determines how close to either regime this change in volatility brings us, and that, in turn, determines the stock-bond dynamics on that day.

The introduction of the STR framework allows us to study several aspects of the relation between comovement and changes in volatility:

- We can test the models against linearity. If we reject linearity, we also reject that comovement is independent of $\Delta\widehat{\sigma}_{t+1}$.[7]
- We can diagnostically test which transition function best fits the data, G_t^E or G_t^L. Here I use the Akaike information criterion (AIC) to select the model.[8] If G_t^E is the best choice, then small changes in volatility belong to one regime and large changes belong to the other (see the graphs in Figure 6), regardless of whether they are positive or negative (assuming that $c = 0$, see next bullet point). This would correspond to the use of $\mathrm{abs}\left(\Delta\widehat{\sigma}_{t+1}\right)$ in the linear model (3.8). If, on the other hand, G_t^L is the best choice, then negative changes belong to one regime and

[7] Luukkonen, Saikkonen and Teräsvirta (1988) show how it is possible to test the model against linearity through testing $\mathrm{H}_0 : \gamma = 0$. Note that using the t-value of $\widehat{\gamma}$ is not an appropriate way to test H_0.

[8] For a discussion of how to select model, see Teräsvirta (1994).

positive to the other (again, assuming $c = 0$). This would correspond to using $\Delta\widehat{\sigma}_{t+1}$ in the linear model. Consequently, we can interpret this test as evidence of whether flight-from-quality has a similar or different impact on comovement than does flight-to-quality.

- We can test if $c = 0$. The estimated constant c has different interpretations depending on whether we use G_t^E or G_t^L. In Figure 6, and if we use G_t^E, c corresponds to the location on the X-axis where G_t^E reaches its lowest value. If $c = 0$, the transition function is symmetric around 0, so that positive and negative changes have the same impact. If $c \neq 0$, then it is symmetric around some other point on the X-axis. If, on the other hand, we use G_t^L, then c corresponds to the mid-point of the transition between regimes, i.e., the point on the X-axis where G_t^L equals 0.5. If $c = 0$ in this case, then negative changes belong to one regime and positive to the other.
- As in Figure 6, we can also plot the G_t function that best fits the data. This gives us a picture of how different values of $\Delta\widehat{\sigma}_{t+1}$ affect comovement. The value of γ determines the steepness of the functions (a high value implies a dramatic shift, a low value a smoother shift). Plotting the shape of G_t then makes the interpretation of γ easier.

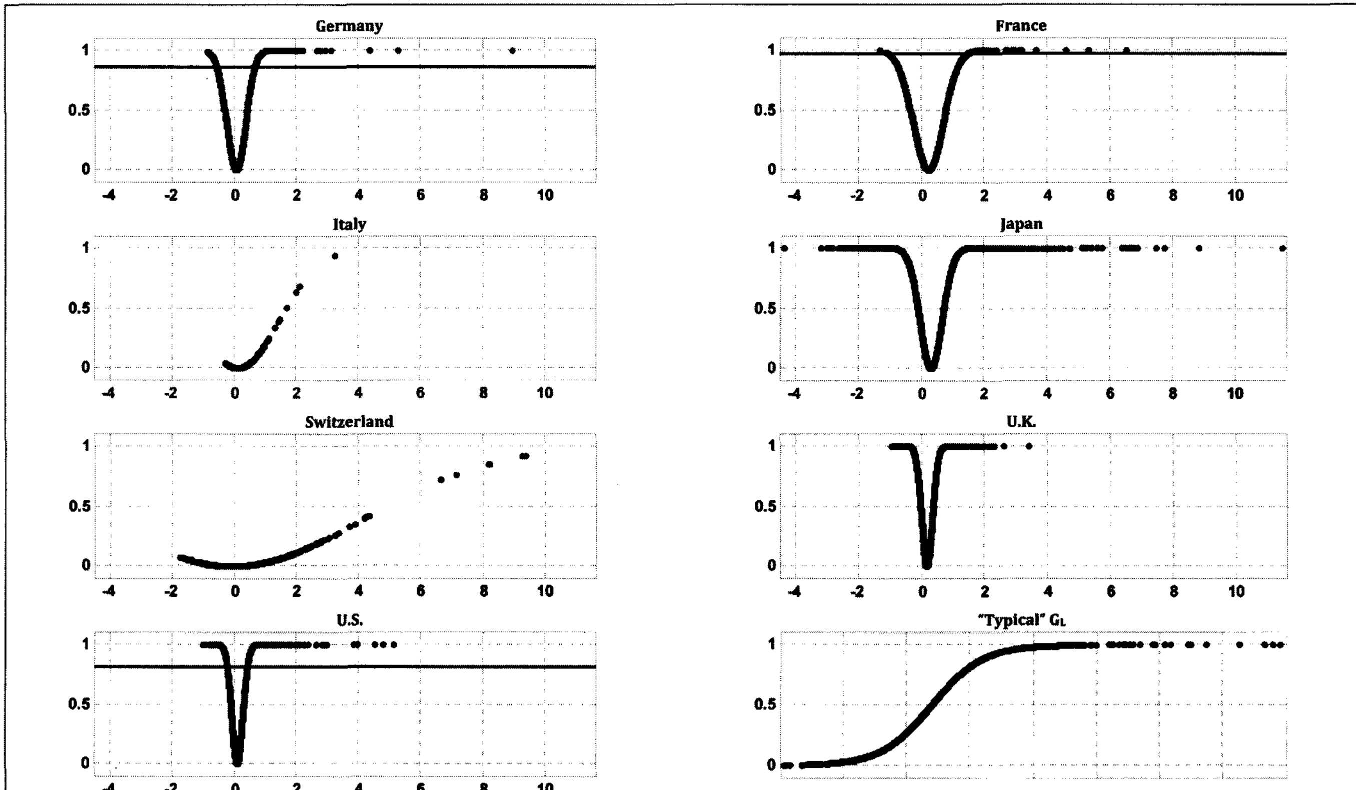
The results are summarized in Table 9. Note that the time period is changed to January 1988–April 2006 to avoid including the October 1987 crash, since the model is sensitive to outliers. The diagnostics favor G_t^E as the transition function in each case. However, some problems still remain. If we look at Figure 6, we see that the plots for Italy and Switzerland do not have the U-shape of a typical G_t^E transition function. Only 13 observations in the case of Italy, and 19 in the case of Switzerland are above 0.2. This indicates that even though the tests favor G_t^E, it is still not appropriate. This should come as no surprise, considering the results from Table 7. One interpretation could be that the model simply picks up a few outliers, and that this is the reason we reject linearity. For Germany, France, the U.K., and the U.S., G_t^E is favored. This means that negative changes have the same type of effect as do positive changes, much in the way outlined in Section 3.2.1. However, c does not seem to be exactly equal to zero. For France, the U.K., and the U.S., c is small but significantly positive, while for Germany it is positive but not significant. This means that a small positive shock to volatility is what is typical of the high correlation regime. Note that this could also be a reason for the weak results from model (3.8)-(3.9), as it implicitly assumes that $c = 0$.

Now, note that

$$\frac{\partial B_t}{\partial S_t} = \alpha_{21} + (\alpha_{22} - \alpha_{21})\, G_t.$$

This means that the marginal contribution from S_t to B_t varies as an affine function of G_t. Since we expect that the comovement in regime 2, the large change-regime, should be lower than in regime 1, our first hypothesis is that $\alpha_{22} - \alpha_{21} < 0$ or, equivalently, that $\alpha_{22} < \alpha_{21}$. For Germany, France, the U.K., and the U.S., this is the case. For Japan, the opposite is true, and we conclude once again that the hypothesis fails for this country.

Figure 6
Transition Functions for all Countries



For all countries, the chosen function is G^E. An example of a typical G^L is given in the last graph.

Table 9
Parameter Estimates for the STR Model (3.10)-(3.14)

Country	Period	#obs	AIC^E/AIC^L	G^E or G^L	p-val	α_{01}	α_1	α_{21}	α_{02}	α_{22}	γ	c
Germany	Jan/1988— Mar/2006	4442	**-2009.40**/-1995.90	E	**3.36e-10**	0.0237	0.0258	**0.0632**	-0.067	-0.0104	**5.31**	0.0571
						(1.41)	(1.63)	**(8.42)**	(-1.44)	(-0.99)	**(2.24)**	(1.37)
France	Jan/1988— Mar/2006	4276	**259.36**/290.60	E	**1.54e-5**	0.0149	**-0.0780**	**0.072**	0.0266	-0.00255	**2.17**	**0.215**
						(0.57)	**(-3.78)**	**(7.21)**	(0.40)	(-0.15)	**(2.09)**	**(2.11)**
Italy	Apr/1991— Mar/2006	3731	**125.04**/148.92	E	**2.87e-5**	0.00567	**0.112**	**0.045**	3.34	0.278	0.286	0.112
						(0.17)	**(5.10)**	**(7.08)**	(1.18)	(1.93)	(1.21)	(1.26)
Japan	Jan/1988— Mar/2006	4241	**-1553.20**/-1553.00	E	**2.26e-7**	0.0237	0.0286	**-0.0416**	0.0202	-0.00389	4.00	**0.303**
						(0.84)	(1.29)	**(-3.94)**	(1.06)	(-0.49)	(1.70)	**(2.40)**
Switzerland	Jan/1988— Mar/2006	3163	**-917.95**/-913.79	E	**2.05e-9**	0.0175	0.0344	**-0.016**	0.723	**0.110**	0.0281	-0.0736
						(1.14)	(1.21)	**(-2.28)**	(1.16)	**(4.59)**	(1.27)	(-0.069)
U.K.	Jan/1988— Mar/2006	4464	**148.03**/169.50	E	**2.74e-8**	-0.0657	0.0306	**0.114**	**0.0607**	0.00372	**17.0**	**0.154**
						(-1.71)	(1.66)	**(7.61)**	**(2.02)**	(0.25)	**(2.55)**	**(5.10)**
U.S.	Jan/1988— Mar/2006	4414	**816.89**/836.28	E	**3.54e-6**	0.0127	**0.0558**	**0.0931**	-0.00832	-0.021	**16.2**	**0.097**
						(0.40)	**(3.49)**	**(5.75)**	(-0.23)	(-1.30)	**(2.57)**	**(2.90)**

For low p-values we reject H_0: no regime shift. Numbers in parentheses are NLS t-values.
The model is $B_t = \alpha_1 B_{t-1} + (\alpha_{01} + \alpha_{21} S_t)(1 - G_t) + (\alpha_{02} + \alpha_{22} S_t)(G_t) + \varepsilon_t$, $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$.
$G^E_t = 1 - \exp\{-\gamma(\Delta\sigma_{t+1} - c)^2\}$, $G^L_t = (1 + \exp\{-\gamma(\Delta\sigma_{t+1} - c)\})^{-1}$
p-values are calculated according to Luukkonen, Saikkonen and Teräsvirta (1988) for H_0: $\gamma = 0$ (i.e., no regime shift) in the model chosen.
"AIC^E | AIC^L" are the Akaike Information Criteria using G^E or G^L.
"G^E or G^L" indicates the transition function that is chosen, i.e., the lower of AIC^E and AIC^L.

We are also interested in explaining negative comovement, not just lower comovement. For the total contribution from S_t to B_t to be negative, the value of $\alpha_{21} + (\alpha_{22} - \alpha_{21}) G_t$ must be negative. For Germany, France, and the U.S., this is possible (since $\alpha_{22} < 0$), and we can consequently explain negative comovement is those cases. I have inserted a horizontal bar for these three countries in Figure 6 to indicate where the transition function G_t is large enough to produce negative comovement. It should be noted, however, that there are large differences as to how often this occurs. While for Germany, 6% of the observations are above the threshold, for France only 1.7% are. And for the U.S., the figure is as high as 22%.

The results indicate that changes in volatility accompanied by negative stock-bond comovement primarily occur in the U.S. Lower stock-bond comovement can also be observed in Germany, France, and the U.K. The implications for hedging might be limited, though. In the next section we will see that the out-of-sample performance is quite unstable. Thus, U.S. bonds may not be such a good hedge against sudden changes in stock market volatility as one might expect–and as was suggested by Gulko (2002). More research on the economic value of this could prove interesting.

3.3. Out-of-Sample Test. We can test the out-of-sample performance of the STR model. The alternative linear (or single regime) model is

$$B_t = \alpha_0 + \alpha_1 B_{t-1} + \alpha_2 S_t + \varepsilon_t, \tag{3.15}$$

$$\varepsilon_t \sim N\left(0, \sigma^2\right). \tag{3.16}$$

The parameter estimates are not updated in the out-of-sample period, due to numerical problems. The results are reported in Table 10. In the cases of France, Italy, and the U.S., the STR model does not outperform the linear model (using the p-value according to Harvey, Leybourne and Newbold (1998)). These results are counter-intuitive and require a comment. The test does not reject that there is no improvement for France and the U.S., even though the in-sample estimates were favorable. And for Japan and Switzerland, the test does reject that there is no improvement, even though we rejected those earlier. In the case of Japan, it appears that the relation with the opposite sign is robust, even though this is contrary to our hypothesis. For the others, the test appears to be sensitive to whether a shock actually occurred in the out-of-sample period. If we look at the last diagram in Figure 4 again, we see that no large shocks occur in the U.S. after 2003. Instead, that period is unusually calm. When the nonlinear phenomenon does not occur, the model loses its advantage over the linear alternative.[9]

4. International Markets and the Level of Volatility

The tests so far have been carried out on a domestic level. Many investors, however, operate on an international level and are not restricted to choose government bonds in the same country as hedging instrument. A popular international choice of a low-risk

[9] The reader might be interested in what happens if we make the out-of-sample period longer. The test strongly rejects no improvement as soon as we reach the large changes around the year 2000.

asset is U.S. government bonds[10], so it is interesting to expand the previous model for the level of volatility and study the comovement between different domestic stock markets and the bond market in this country.[11]

We need to change the notation slightly to separate the different countries. The superscript S now indicates "safe market," and this will always be the U.S. market. The superscript U indicates "unsafe." To exemplify, say a manager invests in the German stock market and wants to hedge his investment in a low-risk asset. To do so, he buys bonds in the U.S. In this case, Germany would be the "unsafe" market and the U.S. would be the "safe" one. All returns are now in U.S. dollars.

We can extend model (3.1)-(3.2) for the new situation, and include the lagged bond return from the safe market, B^S, the stock returns from both markets, S^S and S^U, and a term controlling for the changing comovement between domestic U.S. stocks and bonds, $\log(\widehat{\sigma}_t) \times S_t$. This is necessary, since we have already seen that this domestic effect exists.

$$\begin{aligned} B_t^S &= \alpha_0 + \alpha_1 B_{t-1}^S + \alpha_2 S_t^S + \alpha_3 S_t^U + \alpha_4 \log\left(\widehat{\sigma}_t^S\right) \times S_t^S \\ &\quad + \alpha_5 \log\left(\widehat{\sigma}_t^U\right) \times S_t^U + \varepsilon_t, \qquad (4.1) \\ \varepsilon_t &\sim \mathrm{GARCH}(1,1). \qquad (4.2) \end{aligned}$$

The results for comparing each country pairwise with the U.S. are summarized in Table 11. To begin, we should expect $\widehat{\alpha}_2$ to be positive and significant, and $\widehat{\alpha}_4$ to be negative and significant, since these values refer to the domestic U.S. market in all cases above. (α_2 and α_4 concern the domestic U.S. comovements, i.e., we examined these parameters previously in the domestic level tests; see Section 3.1.) We can regard this as an additional robustness check of model (3.1)-(3.2). It is also robust to adding other countries. All estimates of these two parameters are in accordance with the previous results in Table 3.

The parameter of most interest now is α_5. For Germany, France, and the U.K. we reject $H_0 : \alpha_5 = 0$, whereas for Italy, Japan, and Switzerland we do not. In every case the estimates of α_5 are lower in absolute value than the estimates of α_4, which is intuitive: The domestic U.S. market should be more important than other countries' markets. The values of $\widehat{\alpha}_3$ are positive for all countries, except Italy, but most t-statistics are low.

Since we are mainly interested in the total marginal contribution from the "unsafe" market to U.S. bonds, I plot this against $\widehat{\sigma}_t^U$ for all countries in Figure 7. Remember that the slopes for Italy, Japan, and Switzerland are not significantly different from zero. For Germany, 99% of the values are below zero while for France and the U.K., about 10% are below zero.

The behavior of German, French, and U.K. stocks in relation to U.S. bonds is similar to the results from Section 3.1. When domestic stock market volatility is high, comovement with U.S. bonds is reduced, or negative. This makes U.S. bonds very useful for hedging, and particularly so during times of high stock market uncertainty.

[10] Swiss bonds are also sometimes mentioned as a safe market. I performed a similar study with Swiss bonds instead of U.S. bonds, but I could not find much evidence in favor of this hypothesis.

[11] I do not do this for changes in volatility, since the framework of (3.10)-(3.14) does not translate well to the international case.

Table 10
Test for Forecasting Accuracy for the STR Model (3.10)-(3.14)

	Germany	France	Italy	Japan	Switzerland	U.K.	U.S.
$MSFE_L$	0.79832	1.45103	0.74609	0.51010	0.67315	0.87783	1.62427
$MSFE_{N\text{-}L}$	0.76863	1.44498	0.73966	0.50522	0.65369	0.84923	1.62816
Δ	-0.037	-0.004	-0.009	-0.010	-0.029	-0.033	0.002
p-value	**3.86e-5**	0.134	0.104	**5.91e-5**	**6.62e-10**	**4.76e-5**	0.344

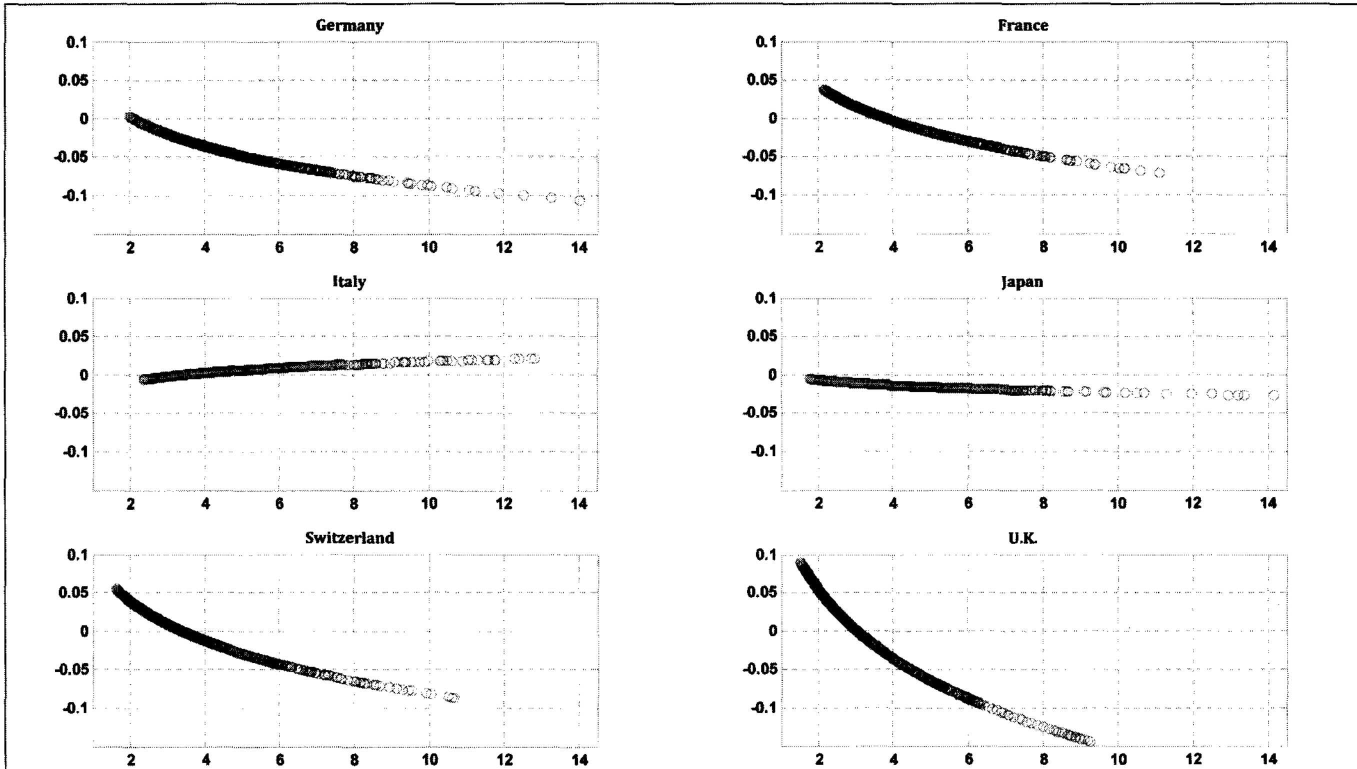
The parameters were not updated between periods. The out-of-sample period is Jan/2002—Mar/2006.
"$MSFE_L$": Mean Squared Forecast Error for the Linear Model. The model is the same as for $MSFE_{N\text{-}L}$ below, but with the restriction $\gamma = 0$.
"$MSFE_{N\text{-}L}$": Mean Squared Forecast Error for the Non-Linear Model. The model is
$B_t = \alpha_1 B_{t-1} + (\alpha_{01} + \alpha_{21} S_t)(1 - G_t) + (\alpha_{02} + \alpha_{22} S_t)(G_t) + \varepsilon_t$, $G_t = 1 - \exp\{-\gamma(\Delta\hat{\sigma}_{t+1} - c)^2\}$, $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$.
"Δ": Change in MSFE, i.e., $(MSFE_{N\text{-}L} - MSFE_L)/MSFE_L$.
p-values are according to Harvey, Leybourne and Newbold (1998) for H_0: No improvement in forecasting accuracy. For France, Italy, and the U.S. we do not reject H_0.

Table 11
Parameter Estimates for Model (4.1)-(4.2)

Country vs. U.S.	Period	#obs	α_0	α_1	α_2	α_3	α_4	α_5	γ_0	γ_1	γ_2	LLF
Germany	Jan/1984—	5357	0.0085	**0.0777**	**0.3069**	0.0413	**-0.2271**	**-0.0557**	**0.0346**	**0.0539**	**0.9197**	-8085.3
	Mar/2006		(0.59)	**(5.44)**	**(5.76)**	(1.39)	**(-3.85)**	**(-2.09)**	**(4.19)**	**(4.86)**	**(71.4)**	
France	Jan/1984—	5362	0.0105	**0.0730**	**0.3686**	**0.0882**	**-0.3277**	**-0.0660**	**0.0366**	**0.0612**	**0.9117**	8100.9
	Mar/2006		(0.68)	**(4.88)**	**(12.8)**	**(3.53)**	**(-14.4)**	**(-3.28)**	**(4.09)**	**(3.72)**	**(50.5)**	
Italy	Jan/1984—	5364	0.0112	**0.0773**	**0.4066**	-0.0198	**-0.3680**	0.0161	**0.0389**	**0.0643**	**0.9071**	-8113.2
	Mar/2006		(0.73)	**(5.18)**	**(15.4)**	(-0.86)	**(-14.3)**	(0.96)	**(3.74)**	**(3.26)**	**(40.5)**	
Japan	Jan/1984—	5322	0.0098	**0.0786**	**0.3272**	0.0005	**-0.2622**	-0.0103	**0.0293**	**0.0520**	**0.9263**	-8063.9
	Mar/2006		(0.65)	**(5.19)**	**(7.14)**	(0.02)	**(-5.90)**	(-0.69)	**(4.13)**	**(4.89)**	**(80.8)**	
Switzerland	Jan/1984—	5354	0.0042	**0.0680**	**0.3027**	0.0916	**-0.2352**	-0.0751	**0.0319**	**0.0539**	**0.9221**	-8094.2
	Mar/2006		(0.30)	**(4.93)**	**(5.52)**	(1.93)	**(-3.92)**	(-1.67)	**(4.20)**	**(4.79)**	**(73.0)**	
U.K.	Jan/1984—	5346	0.0038	**0.0708**	**0.3022**	**0.1441**	**-0.2381**	**-0.1296**	**0.0323**	**0.0543**	**0.9213**	-8075.7
	Mar/2006		(0.26)	**(5.13)**	**(5.66)**	**(2.24)**	**(-4.21)**	**(-2.00)**	**(4.22)**	**(5.29)**	**(73.9)**	

A significant α_5 indicates time-varying comovement between the country's stocks and U.S. 10-year bonds w.r.t. the volatility on the country's stock market. Numbers in parentheses are Bollerslev and Wooldridge t-values.
The model is $B^S_t = \alpha_0 + \alpha_1 B^S_{t-1} + \alpha_2 S^S_t + \alpha_3 S^U_t + \alpha_4 \ln(\sigma^S_t) S^S_t + \alpha_5 \ln(\sigma^U_t) S^U_t + \varepsilon_t$, $\varepsilon_t = z_t h_t$, $z_t \sim N(0,1)$, $h^2_t = \gamma_0 + \gamma_1 \epsilon^2_{t-1} + \gamma_2 h^2_{t-1}$.

Figure 7
Domestic Stocks' Comovement with U.S. Bonds w.r.t. Domestic $\hat{\sigma}_t$



A plot of the total marginal contribution from $S^U{}_t$ to $B^S{}_t$ as a function of $\hat{\sigma}_t^U$ (i.e., $\alpha_3 + \alpha_5 \ln(\hat{\sigma}_t^U)$ as a function of $\hat{\sigma}_t^U$); see model (4.1)–(4.2).

5. Discussion

We have established that GARCH estimates of domestic stock market volatility tend to predict stock-bond comovement, such that high volatility precedes low or negative comovement and vice versa. The only exception is Japan, where we found no significant relation. We have also established that for four countries (Germany, France, the U.K., and the U.S.), large changes in volatility tend to coincide with low or negative comovement. We have not found any such relation for Japan, Italy, and Switzerland. As a third result, we have established that the level of volatility in France, Germany, and the U.K. tends to predict the comovement between the stock market in that country and the bond market in the U.S.

Japan is a major exception among the countries we have studied. None of the relations that hold for some, or all, of the other countries holds for Japan. One circumstance that puts Japan in a category by itself is that this country has had a long and deep economic crisis since the early 1990s until the present. I have, however, run the tests concerning the level of volatility on a subsample for the period 1984–1990, but the results were still not significant; it appears that the Japanese economy works differently than the western economies in this study.

The results for Italy and, to some extent, Switzerland are rather mixed, while for Germany, France, the U.K., and the U.S. they are strong. It is noteworthy that the strongest results (highest values of point estimates as well as the highest t-statistics) are for the U.S. This is the largest stock market in the world. We could speculate that transaction costs are lower in the U.S., and that higher transaction costs tend to blur the picture in the other markets. Another fact that could explain some of the weak results for Italy and Japan is that these countries, as mentioned in Section 2, have lower credit ratings than the other five. Particularly the flight-to-quality argument below is less valid if the domestic bond is not of "high quality." Recall, also, from Section 2 that the summary statistics for Japan show that data from this country are very different from data from the others.

Now that we have established these results, what sense can we make of them? Here, I will discuss three different explanations: (1) Statistical problems drive the results; (2) The comovement changes because managers rebalance their portfolios as a response to changes in volatility; (3) The comovement changes because market timers are more active during times of high volatility.

5.1. Statistical Problems. A segment of the financial contagion literature has been devoted to possible biases in estimated conditional correlations. Forbes and Rigobon (2002) note that estimating correlations, conditional on variance, produces upward biased results in some cases. This type of bias is, as Connolly, Stivers and Sun similarly argue, not a serious challenge to our results here. This study focuses on whether we can explain a change in the sign of the comovement, i.e., if we can explain that positive comovement shifts to negative. The bias problem addressed by Forbes and Rigobon cannot have this effect, so I do not believe that it poses any threat to our conclusions. However, this does not mean that our estimates necessarily are unbiased, only that the change in sign is likely to be real.

5.2. Volatility, Portfolio Rebalancing, and Flight-to-Quality . For the further discussion, let us first review a simple asset pricing framework. The value of a financial asset can be viewed as the sum of all future payments that the owner is entitled to, discounted at some risk-corrected discount rate. Applying this to government bonds, which are considered to be risk-free, the future payments are the coupons and the nominal value, and the discount rates are the risk-free rates of return. The components of the risk-free rates are, in turn, the inflation rate, the term premium, and the real rate of return. For stocks, the future payments are an infinite stream of stochastic dividends, and the discount rates are the sum of the risk-free rates of return and risk-correction terms that are related to the risk in the dividends.

Since the risk-free rate is present in both the stock and the bond valuation formulas, variation in this component could be believed to cause positive comovement. However, this is not necessarily so. For instance, the stochastic dividends might be correlated to components in the risk-free rate. The most obvious example is that the inflation rate should also enter into the expected dividends, such that a rise in inflation would be largely offset by a corresponding rise in expected dividends. This idea is sometimes summarized as "stocks are real, bonds are nominal." Furthermore, investors could view a rise in inflation as negative news for the real economy. But they could also view it as positive news, or as neutral news. Hence, a change in inflation could cause either positive or negative comovement, or have no effect on comovement at all. See Li (2002) for a further discussion of this topic.

Now, regarding volatility and portfolio rebalancing, a very popular explanation for dynamic comovement between stock and bond returns is flight-to-quality (or cross-market hedging). There are frequent references to this concept in the popular press, and Connolly, Stivers and Sun, as well as Cappiello, Engle and Sheppard also refer to it. There is no agreement as to what this actually means, but the following captures the intuition: Consider a portfolio manager who wishes to keep a certain risk level in his portfolio. Suppose the risk in the stock market suddenly increases, causing the risk premium on stocks to increase and the price of stocks to decrease. The risk in the portfolio is now too high and bonds become relatively more attractive to the manager. The increased demand for bonds pushes up the price. So, the prices of stocks and bonds move in opposite directions and the correlation is lower, possibly negative. Occasionally, mention is made of the reverse, more unusual, phenomenon: flight-from-quality. This would occur if stock market volatility suddenly dropped, and would also, similarly, lower the correlation.

In this paper, the case that is closest to this idea of flight-to-quality is the test regarding changes in volatility and stock-bond comovement. As we just noted, however, this effect can only be spotted in Germany, France, the U.K., and the U.S. It also occurs very rarely and only when changes in volatility are very large in the non-U.S. markets. Furthermore, in the U.K. this does not seem to explain negative comovement at all. In the remaining three countries, Italy, Japan and Switzerland, we see no evidence of this type of phenomenon.

It is possible, perhaps, to have a looser notion of flight-to-quality. We could interpret it such that, when the stock market on average makes large jumps in a certain period, investors tend to value bonds higher and higher, thereby causing negative comovement. This interpretation is closer to the case regarding the level of volatility and stock-bond comovement. And for this case, we have strong evidence in favor in all countries, except Japan.

5.3. Volatility and Market Timing. Another type of investor behavior that could cause a decrease in stock-bond comovement is market timing. Consider an investor who has some market power, i.e., when he makes a large trade, prices move. In each period he forecasts whether the stock market will rise or decline before the next period. If he forecasts a rise, he invests part of his wealth in stocks; otherwise he invests it in bonds. Assuming that he is better at predicting the market than mere chance, this strategy will have a positive expected excess return. Since he has market power, his buying and selling stocks and bonds will push the prices of these securities in different directions. If we also assume that other fundamentals induce a positive correlation between stocks and bonds, then the behavior of the market timer pushes this positive correlation downwards. If he makes very large trades (possibly at the same time as other market timers), he will even push it into negative correlation.

Clearly, if the stock market variance is zero, the market timing strategy has no value. It is also intuitive that the value of market timing is increasing in volatility–the more prices change, the more money the investor makes. Merton (1981) derives an analytical expression for the value of the above-mentioned strategy, and for demand for a fund that uses such a strategy. Both expressions are, as we would expect, strictly increasing in stock market variance. Merton's results, however, assume that the market timer has no market power, so the analytical expressions cannot be directly translated to the present context.

The assumption that the fund has no impact on prices is hardly realistic, as Merton himself points out. The fund's size should be one of the determinants of its performance: The larger the fund, the more it moves prices–and the more it moves prices, the smaller the marginal profit. In equilibrium, the fund should be of an optimal size with price influences taken into account. A rise in volatility would still make the strategy more valuable, so we still expect demand to be increasing in volatility. A larger share of the market will then shift when the market timer changes his portfolio, and prices will change more when volatility is high. The impact of the market timer, consequently, is to push down the correlation. If other factors induce a modest positive correlation between stocks and bonds, we would get a positive correlation when volatility is low and the market timer does not influence prices so much, and lower, possibly even negative, correlation when volatility is high.

Note that the market timer need not actually have predictive power; it is enough that he acts as if he believes that he does. *Ex post*, it might be evident that he does not. But as long as he moves in and out of stocks and bonds on a regular enough basis, he will push the comovement towards zero, possibly even into negative territory. Many hedge funds advertise that they "chase the alpha," i.e., that they try to earn a return that is not related to market risk. In a CAPM world this is, obviously, impossible, but if these funds use a market-timing strategy similar to the one outlined above, and employ enough assets, this could explain some of the changes in stock-bond comovement.

5.4. Implications. The results from the analysis could probably be applied directly in financial decision-making. Comovement between stocks and bonds is a key input for portfolio managers and risk managers, among others, and better knowledge of the dynamics of comovement is likely to lead to better decisions. Since comovement tends to be lower, or even negative, at times of high volatility or when volatility makes a big jump, diversification benefits are largest when most needed. This might be a pleasant surprise if one is hit by bad times, and it would be tempting to congratulate those who are so lucky. But, pleasant or not, it should not have been a surprise at all; it should have been reckoned with beforehand. If surprised, the manager did not optimize his portfolio with respect to his hedging needs in the first place.

It seems important to incorporate information about volatility into models that deal with time-varying stock-bond comovement, both for financial theory, and for financial econometricians who develop models to mimic the behavior of real stock and bond returns. As the mimicking gets better, so should our understanding; and as our understanding gets better, so should our mimicking skills. And both should help us to make better decisions.

References

Bollerselev, T., and Wooldridge, J. M., 1992, "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances," *Econometric Reviews*, 11, 143-172.

Campbell, J.Y., and J. Ammer, 1993, "What Moves the Stock and Bond Markets? A Variance Decomposition for Long-Term Asset Returns," *Journal of Finance*, 48(1), 3-37.

Campbell, R.A.J., K.C.G. Koedijk, and P. Kofman, 2002, "Increased Correlation in Bear Markets: A Downside Risk Perspective," CEPR Discussion Paper No. 3172.

Cappiello, L., R.F. Engle, and K. Sheppard, 2003, "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns," ECB Working Paper No. 204.

Clark, T. E., and M. W. McCracken, 2001, "Test for Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, 105, 85-110.

Connolly, R. A., C. T. Stivers, and L. Sun, 2005, "Stock Market Uncertainty and the Stock-Bond Return Relation," *Journal of Financial and Quantitative Analysis*, 40, 161-194.

Engle, R., 2002, "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Model," *Journal of Business and Economic Statistics*, 20(3), 339-349.

Fleming, J., C. Kirby, and B Ostdiek, 1998, "Information and Volatility Linkages in the Stock, Bond, and Money Markets," *Journal of Financial Economics*, 49, 111-137.

Fleming, J., C. Kirby, and B Ostdiek, 2003, "The economic value of volatility timing using 'realized' volatility," *Journal of Financial Economics*, 67, 473-509.

Forbes, K. J., Rigobon, R., 2002, "No Contagion, Only Interdependence: Measuring Stock. Market Comovements," *Journal of Finance*, 57(5), 2223-2261.

Franses, P.H., and D. van Dijk, 2000, *Nonlinear Time Series Models in Empirical Finance*, Cambridge University Press.

Gulko, L., 2002, "Decoupling," *Journal of Portfolio Management*, 28, 59-66.

Harvey, D.I., S.J Leybourne, and P. Newbold, 1998, "Tests for Forecast Encompassing," *Journal of Business and Economic Statistics*, 16, 254-259.

Kodres, L. E., and M. Pritsker, 2002, "A Rational Expectations Model of Financial Contagion," *Journal of Finance*, 57(2), 769-799.

Kroner, K. F., and V. K. Ng, 1998, "Modeling Asymmetric Comovements of Asset Returns," *Review of Financial Studies*, 11, 817-844.

Li, L., 2002, "Macroeconomic Factors and the Correlation of Stock and Bond Returns," Yale ICF Working Paper No. 02-46.

Longin, F., and B. Solnik, 2001, "Extreme Correlation of International Equity Markets," *Journal of Finance*, 56(2), 649-676.

Lundbergh, S., and T. Teräsvirta, 2002, "Evaluating GARCH Models," *Journal of Econometrics*, 110, 417-435.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta, 1988, "Testing Linearity Against Smooth Transition Autoregressive Models," *Biometrika*, 75, 491-499.

Marten, M., and S. H. Poon, 2001, "Return Synchronization and Daily Correlation Dynamics between International Stocks Markets," *Journal of Banking and Finance*, 25, 1805-1827.

Merton, R.C., 1981, "On Market Timing and Investment Performance. I. An Equilibrium Theory of Value for Market Forecasts," *Journal of Business*, 54(3), 363-406.

Scruggs, J., and P. Glabadanidis, 2003, "Risk Premia and the Dynamic Covariance between Stock and Bond Returns," *Journal of Financial and Quantitative Analysis*, 38, 295-316.

Shiller, R. J., and A. E. Beltratti, 1992, "Stock prices and bond yields: Can their comovements be explained in terms of present value models?," *Journal of Monetary Economics*, 30, 25-46.

Teräsvirta, T., 1994, "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89(425), 208-218.

EFI, The Economic Research Institute

Published in the language indicated by the title. A complete publication list can be found at www.hhs.se/efi Books and dissertations can be ordered from EFI via e-mail: EFI.Publications@hhs.se

Reports since 2003

2007

Books

Wijkström, Filip och Torbjörn Einarsson. *Analysmodell för sektorsöverskridande statistik. Fallet vård och omsorg.*

Dissertations

Alexius, Susanna. *Regelmotståndarna – om konsten att undkomma regler.*

Berg, Bengt Åke. *Volatility, Integration and Grain Bank. Studies in Harvests, Rye Prices and Institutional Development of the Parish* Magasins *in Sweden in the 18th and 19th Centuries.*

Brodin, Karolina. *Consuming the Commercial Break: An Ethnographic Study of the Potential Audiences for Television Advertising.*

Hagberg, Axel. *Bankkrishantering: Aktörer, marknad och stat.*

Hjalmarson, Hanna. *En växande marknad: Studie av nöjdheten med konsumtionsrelaterade livsområden bland unga konsumenter.*

Hjelström, Tomas. *The Closed-End Investment Company Premium Puzzle – Model development and empirical tests on Swedish and British data.*

Kraus, Kalle. *Sven, inter-organisational relationships and control – a case study of domestic care of the elderly.*

Melian, Catharina. *Progressive Open Source*

Valiente, Pablo. *Re-innovating the Existing. A study of Wireless IS Capabilities to Support Mobile Workforces.*

2006

Books

Thodenius, Björn. *Organisering av kunskap. En studie av Wallenberg Consortium North.*

Wijkström, Filip och Torbjörn Einarsson. *Från nationalstat till näringsliv? Det civila samhällets organisationsliv i förändring.*

Wijkström, Filip, Einarsson, Stefan och Larsson, Ola. *Staten och det civila samhället. Idétraditioner och tankemodeller i den statliga bidragsgivningen till ideella organisationer.*

Östman, Lars. *Lysande ögonblick och finansiella kriser. Dramaten under ett sekel.*

Dissertations

Argenton, Cedric. *Quality provision in duopoly.*

Beckerman, Carina. *The Clinical Eye. Constructiong and Computerizing an Anesthesia Patient Record.*

Borglund, Tommy. *Aktievärden i fokus – internationell påverkan på intressentrelationer vid förvärv och fusion.*

Breman, Anna. *The Economics of altruism, paternalism and self-control.*

Edquist, Harald. *Technological breakthroughs and productivity growth.*

Eklund, Jana. *Essays on Forecasting and Bayesian Model Averaging.*

Frostenson, Magnus. *Legitimitetskontrollen – en studie av etiska värderingars roll i gränsöverskridande förvärv och fusioner.*

Gaspar, Raquel M. *Credit Risk and Forward Price Models.*
Gustafsson, Peter. *Essays on Trade and Technological Change.*
Hopkins, Elisabeth. *Is a higher degree of local currency pricing associated with lower exchange rate pass-through? A study of import pricing in 51 Swedish industries.*
Kling, Ragnar. *Developing Product Development in Times of Brutal Change.*
Langenskiöld, Sophie. *Peer Influence on Smoking: Causation or Correlation?*
Lychnell, Lars-Olof. *"Och fungerar det inte, gör vi på något annat sätt" — En klinisk fallstudie av IT-relaterat förändringsarbete i småföretag*
Meitz, Mika. *Five Contributions to Econometric Theory and the Econometrics of Ultra-High-Frequency Data.*
Mendicino, Caterina. *Financial Market Imperfections, Business Cycle Fluctuations and Economic Growth.*
Ovanfors, Anna. *Essays on Nonlinear Time Series Analysis and Health Economics.*
Paltseva, Elena. *Essays on Commitment and Inefficiency in Political Economy.*
Rogberg, Martin. *Den modeföljande organisationen. Om acceptansen av TQM och andra populära managementmodeller.*
Silvennoinen, Annastiina. *Essays on Autoregressive Conditional Heteroskedasticity.*
Sjögren, Ebba. *Reasonable Drugs. Making Decisions with Ambiguous Knowledge.*
Slinko, Irina. *Essays in Option Pricing and Interest Rate Models.*
Wilander, Fredrik. *Essays on Exchange Rates and Prices.*

2005

Books

Andersson, Per, Susanne Hertz and Susanne Sweet (eds). *Perspectives on market networks – boundaries and new connections.*
Charpentier, Claes. *IT inom omsorgen. Förväntade effekter av införande av IT-system för utförarna inom äldre- och handikappomsorgen.*
Dembrower, Maria. *Entreprenörskap i industriella nätverk.*
Lind, Johnny och Göran Nilsson (redaktörer). *Ekonomistyrningens metoder, sammanhang och utveckling – En vänbok till Lars A Samuelson.*
Samuelson, Lars A. *Organizational governance and control – a summary of research in the Swedish society.*
Svedberg Nilsson, Karin, Roger Henning och Karin Fernler (eds). *En illusion av frihet? Företag och organisationer i regelsamhället.* EFIs Årsbok 2005. EFI/Studentlitteratur.

Dissertations

Andersson, Martin. *Making a Difference – Project Result Improvement in Organizations.*
Arvidsson, Per. *Styrning med belöningssystem – Två fallstudier om effekter av belöningssystem som styrmedel.*
Bems, Rudolfs. *Essays in International Macroeconomics.*
Berg-Suurwee, Ulrika. *Nya trender, nya nämnder – effekter av en stadsdelsnämndsreform inom kultur och fritid.*
Björkman, Hans. *Learning from members – Tools for strategic positioning and service innovation in trade unions.*
Bodnaruk, Andriy. *Essays in Empirical Corporate Finance and Portfolio Choice.*
Clapham, Eric. *Essays in Real Estate Finance.*
Dareblom, Jeanette. *Prat, politik och praktik – Om individers möten med strukturer i en kommunal satsning på kvinnors företagande.*
Fromm, Jana. *Risk Denial and Neglect: Studies in Risk Perception.*
Hjelström, Anja. *Understanding International Accounting Standard Setting – A Case Study of IAS 12, Accounting for Deferred Tax.*
Hortlund, Per. *Studies on Swedish Banking 1870-2001.*

Lindahl, Therese. *Strategic and Environmental Uncertainty in Social Dilemmas.*

Linnarsson, Håkan. *Alliance for Innovation. A structural perspective on new business development in cooperative ventures.*

Madestam, Andreas. *Developing Credit Markets.*

Nilsson, Roland. *The Market Impact of Short-Sale Constraints.*

Nordfält, Jens. *Is consumer decision-making out of control? Non-conscious influences on the consumer decision-making process for fast moving consumer goods.*

Nordin, Fredrik. *Externalising Services – Walking a Tightrope between Industrial and Service Logics.*

Parmler, Johan. *Essays in Empirical Asset Pricing.*

Simbanegavi, Witness. *Price Discrimination, Advertising and Competition.*

Thodenius, Björn. *Användning av ledningsinformationssystem: en longitudinell studie av svenska storföretag.*

Tolis, Christofer. *Framing the Business – Business Modelling for Business Development.*

Östberg, Per. *Corporate Disclosure and Investor Recognition.*

2004

Books

Ahrne, Göran och Nils Brunsson (red). *Regelexplosionen.*

Lind, Johnny. *Strategi och ekonomistyrning. En studie av sambanden mellan koncernstrategi, affärsstrategi och ekonomistyrning.*

Lind, Johnny och Walter Schuster (red). *Redovisningens teori, praktik och pedagogik. En vänbok till Lars Östman.*

Sevón, Guje och Lennart Sjöberg (red). *Emotioner och värderingar i näringslivet.* EFIs Årsbok 2004.

Wijkström, Filip and Stefan Einarsson. *Foundations in Sweden – Their scope, roles and visions.*

Dissertations

Anderson, Anders. *Essays in Behavioral Finance.*

Balsvik, Gudrun. *Information Technology Users: Studies of Self-Efficacy and Creativity among Swedish Newspaper Journalists.*

Blix, Magnus. *Essays in Mathematical Finance – Modelling the Futures Price.*

González Gómez, Andrés. *Nonlinear dynamics and smooth transition models.*

Grönqvist, Erik. *Selection and Moral Hazard in Health Insurance: Taking Contract Theory to the Data.*

Ivarsson Westerberg, Anders. *Papperspolisen – varför ökar administrationen i moderna organisationer.*

Jutterström, Mats. *Att påverka beslut – företag i EUs regelsättande.*

Jönsson, Kristian. *Macroeconomic Aspects of Capital Flows to Small Open Economies in Transition.*

Larsson, Pär. *Förändringens villkor. En studie av organisatoriskt lärande och förändring inom skolan.*

Lagerwall, Björn. *Empirical Studies of Portfolio Choice and Asset Prices.*

Malmsten, Hans. *Properties and Evaluation of Volatility Models.*

Marshall, Cassandra. *Dating for Innovation. Creating and Recognizing Opportunities through Collaborative Interorganizational Relationships in Fluid Environments.*

Mattsson, Susanna. *På gränsen mellan ordning och oordning – tingens betydelse vid marknadsombildningar. En studie av svenska postväsendets ombildning under 1990-talet.*

Nilsson, Charlotte. *Studies in Environmental Economics: Numerical Analysis of Greenhouse Gas Policies.*

Nilsson, Hans. *Medborgaren i styrsystemet – beskrivning av VAD och HUR i styrning av kommunal verksamhet.*

Nystedt, Jens. *Competition, Regulation and Integration in International Financial Markets.*

Pajuste, Anete. *Corporate Governance and Controlling Shareholders.*

Richtnér, Anders. *Balancing Knowledge Creation. Organizational Slack and Knowledge Creation in Product Development.*

Salabasis, Mickael. *Bayesian Time Series and Panel Models – Unit Roots, Dynamics and Random Effects.*

Sandberg, Rickard. *Testing the Unit Root Hypothesis in Nonlinear Time Series and Panel Models.*

Skallsjö, Sven. *Essays on Term Structure and Monetary Policy.*

Strikholm, Birgit. *Essays on Nonlinear Time Series Modelling and Hypothesis Testing.*

Söderström, John. *Från Produkt till Tjänst. Utveckling av affärs- och miljöstrategier i produktorienterade företag.*

Talia, Krim. *The Scandinavian Currency Union, 1873–1924 – Studies in Monetary Integration and Disintegration.*

2003

Books

Lundahl, Mats (editor). *Globalization and Its Enemies.* EFIs Årsbok 2003.

Sundgren, Bo, Pär Mårtensson, Magnus Mähring and Kristina Nilsson (editors). *Exploring Patterns in Information Management. Concepts and Perspectives for Understanding IT-Related Change.*

Dissertations

Andersson, Henrik. *Valuation and Hedging of Long-Term Asset-Linked Contracts.*

Bergman, Malin. *Essays on Human Capital and Wage Formation.*

Damsgaard, Niclas. *Deregulation and Regulation of Electricity Markets.*

Eklund, Bruno. *Four Contributions to Statistical Inference in Econometrics.*

Hakkala, Katariina. *Essays on Restructuring and Production Decisions in Multi-Plant Firms.*

Holgersson, Charlotte. *Rekrytering av företagsledare. En studie i homosocialitet.*

Ivaschenko, Iryna. *Essays on Corporate Risk, U.S. Business Cycles, International Spillovers of Stock Returns, and Dual Listing.*

Lange, Fredrik. *Brand Choice in Goal-derived Categories – What are the Determinants?*

Le Coq, Chloé. *Quantity Choices and Market Power in Electricity Market.*

Magnusson, Peter R. *Customer-Oriented Product Development – Experiments Involving Users in Service Innovation.*

Meisiek, Stefan. *Beyond the Emotional Work Event Social Sharing of Emotion in Organizations.*

Mårtensson, Anders. *Managing Mission-Critical IT in the Financial Industry.*

Nilsson, Göran. *Processorientering och styrning – Regler, mål eller värderingar?*

Sandberg, Robert. *Corporate Consulting for Customer Solutions Bridging Diverging Business Logics.*

Sturluson, Jon Thor. *Topics in the Industrial Organization of Electricity Markets.*

Tillberg, Ulrika. *Ledarskap och samarbete – En jämförande fallstudie i tre skolor.*

Waldenström, Daniel. *Essays in Historical Finance.*

Wallén, Ulrika. *Effektivitet i grundskolan i anslutning till en stadsdelsnämndsreform.*

Ögren, Anders. *Empirical Studies in Money, Credit and Banking – The Swedish Credit Market in Transition under the Silver and the Gold Standards, 1834–1913.*

STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

ISBN 978-91-7258-726-7