Excess Comovement of Stock Returns: Evidence from Cross-Sectional Variation in Nikkei 225 Weights

Robin Greenwood
Harvard Business School

Relative to their weights in a value-weighted index, a number of stocks in Japan's Nikkei 225 stock index are overweighted by a factor of 10 or more. I document a strong positive relation between overweighting and the comovement of a stock with other stocks in the Nikkei index, and a negative relationship between index overweighting and comovement with stocks outside of the index. The cross-sectional approach resolves endogeneity problems associated with event study demonstrations of excess comovement. A trading strategy that bets on the reversion of stock prices of overweighted stocks generates economic profits, confirming that the observed comovement patterns are excessive, and providing further evidence that comovement of stock returns can be a consequence of commonality in trading behavior. (*JEL* G10, G14, G15)

Several recent papers show that security prices comove in excess of their common fundamentals, casting doubt on the completeness of the model in which comovement is fully explained by common variation in cash flows and discount rates. In their interpretation of the evidence, some authors have argued that excess comovement of stock returns may be explained by the price impact of correlated investor demand, or common liquidity shocks (e.g., Pindyck and Rotemberg, (1993), Lee, Shleifer and Thaler, (1991), Froot and Dabora, (1999)).[1]

Consistent with the intuition that common variation of stock returns may be explained by commonality in investor demand, there is growing empirical evidence that comovement of security returns is related to the trading patterns of groups of investors. Barberis, Wurgler and Shleifer (2005) find that stocks tend to comove more (less) with index stocks after they are added to (deleted from) the S&P 500 index. They argue that this is consistent with excess comovement being driven by correlated demand by investors who trade index stocks together. Kumar and Lee (2005) show that correlated trades of retail investors are related to patterns of comovement in stock returns. Boyer (2006) argues that correlated trading

I thank Taro Hornmark, Nathan Sosner, Jeff Wurgler, Tuomo Vuolteenaho, two anonymous referees and Campbell Harvey (the editor) for their helpful comments. Address correspondence to Robin Greenwood, Baker Library 267, Soldiers Field, Boston, MA 02163, or e-mail: rgreenwood@hbs.edu.

[1] See also Vijh (1994), Hardouvelis, La Porta and Wizman (1994), De Long et al. (1990), and Morck, Yeung and Yu (2000).

doi:10.1093/rfs/hhm052 Advance Access publication October 17, 2007

patterns may explain some of the comovement among stocks with similar book-to-market ratios.

While researchers have been able to identify the existence of commonality in investor demand, they have had less success determining whether the degree of commonality varies across securities, and if so, whether it bears any cross-sectional relation to comovement among security prices. This article lays out cross-sectional predictions from a model of excess comovement of stock returns, and tests them on a large panel of stocks between 1993 and 2003.

The basic idea is as follows. Consider a set of risky assets in fixed identical supply and a risk-free asset in elastic supply. There are limits to arbitrage, so that uninformed demand for securities by groups of investors affects prices. Suppose that a group of investors regularly, and arbitrarily, buys or sells three of the assets, A, B, and C in the ratio 1 : 1 : 2. Thus when these investors buy one share of A and one share of B, they buy two shares of C. Theory predicts that in the presence of limits to arbitrage, a positive demand shock increases the prices of all three assets, but with asset C experiencing the greatest price appreciation, holding all else equal. A negative demand shock generates opposite results. Therefore, when there are regular uninformed demand shocks of unpredictable sign and magnitude, all three securities experience comovement in excess of their common fundamentals, but because of the variation in the exposure to shocks, the three securities experience excess comovement in different degrees. Specifically, the covariance of returns between A and C should exceed the covariance of returns between A and B, after controlling for common fundamentals. This relation arises even though the level of uninformed demand in any period is random—the key insight is that *proportionality* between demand shocks generates a cross-sectional relation between the weighting vector and the comovement of stock returns.

While easy to envision in theory, cross-sectional variation in regular demand shocks of the type described above can be difficult to find in practice. I use the unusual weighting system of the Nikkei 225 index in Japan to identify variation in regular demand shocks for index stocks, and relate this variation to patterns in comovement. The Nikkei 225 index is effectively equal weighted, meaning that stocks exert influence on the index return in proportion to their lagged price. This system ensures that weight is essentially a function of the price at which the stock entered the index.[2] The index weights of some stocks exceed by an order of magnitude their weights in a value-weighted index. Thus, when Nikkei 225 index investor demand rises (perhaps because of demand for Nikkei stocks, or perhaps because of market-wide demand, or perhaps even due to hedging demand

[2] After addition to the index, index weights can go up or down because of stock returns. But this also happens in a value-weighted index, and therefore the relative overweighting is unaffected by stock returns.

by futures traders), investors purchase significantly more of some index stocks than they would if they were using the value-weighted market index as the benchmark. Conversely, when Nikkei 225 index investor demand falls, investors sell more of these stocks than they would if the index were value weighted. For example, on 1 September 2003, Advantest, a producer of testing equipment, had a weight of 3.5% in the Nikkei 225 index by virtue of its high price, while its share of the market capitalization of Nikkei stocks was only 0.28%. More generally, the Nikkei 225 weights of more than 50 stocks were less than their weight in the value-weighted Tokyo Stock Exchange index (TOPIX) at the end of 2003, while the weights of 48 stocks were more than five times their TOPIX weight. If excess comovement of index stock returns is the result of uninformed demand for index stocks, then the returns of overweighted stocks should comove more with the equal-weighted return of the other index stocks, while the returns of underweighted stocks should comove less.

I analyze the comovement among 298 Nikkei index stocks, and between these stocks and nonindex stocks for the 1993 to 2003 period. In cross-sectional regressions, I study the relation between comovement of a stock with other index stocks, and a lagged measure of its overweighting in the Nikkei 225 index relative to the value-weighted TOPIX. As the true cross-sectional structure of investor demand in each period is unobserved, index overweighting acts as an instrument.

The results provide strong support for the predictions. Stocks that are overweighted in the Nikkei 225 index have much higher betas. I also study the comovement of index stock returns with the returns of stocks outside of the Nikkei 225 index, and find converse results—index overweighting is significantly negatively related to the comovement of an index stock with stocks outside of the Nikkei 225, controlling for the Nikkei itself. Finally, I expand the universe of stocks to include all liquid traded stocks in Japan and study the cross-sectional determinants of their comovement with stocks inside the Nikkei 225 index. Controlling for index membership, index overweighting is a significant determinant of the comovement of returns with index returns.

The cross-sectional evidence in this article has important advantages relative to prior event-based demonstrations of excess comovement, such as changes in betas upon inclusion into a stock index (Barberis, Shleifer and Wurgler (2005), Greenwood and Sosner (2007)). The identifying assumption in these event-based studies is that entry into a stock index does not change a stock's loadings on fundamental factors. But, as Denis et al. (2003) show, analysts revise their expectations about future earnings upon entry into the S&P 500 index. It is conceivable that stocks may simultaneously experience shocks in their exposure to factor loadings, particularly factors related to distress risk. With a cross-sectional approach, however, one must only satisfy the requirement that factor loadings are not plausibly related to index overweighting. In the case of the Nikkei 225 index, this seems particularly unlikely—there is no reason that the price at which a stock entered the index long ago should be systematically related to its risk characteristics. To make this concrete, two otherwise identical index firms, each with market value of 1000, would have different Nikkei weights if one were comprised of 10 shares with a price of 100 and the other were comprised of 100 shares with a price of 10. In short, the setting in this article presents several advantages relative to event-based demonstrations of excess comovement.

In addition to being a clean experiment, the cross-sectional approach used in this article provides enough power to analyze crucial but previously untested predictions of models of excess comovement. In particular, if correlated demand shocks are related to mispricing in the cross-section, then one should, in theory, be able to generate abnormal returns by exploiting the predictable reversion of prices. Put simply, if stocks move together too much in the short run, they must diverge at intermediate horizons so that, in the long run, prices reflect fundamentals. This idea motivates my final set of tests. First, I show that overweighted stocks exhibit negative covariance with lagged index returns. The magnitude of this covariance is increasing in the degree of overweighting. Second, I construct a simple trading strategy that bets on the reversion of stock prices of overweighted stocks. A portfolio that takes short positions in overweighted stocks following index increases, and takes long positions in overweighted stocks following index declines, earns significant risk-adjusted profits. Thus, the high short-term comovement of overweighted stocks is truly "excessive" in that these stocks overreact to current index returns. Of course, one could probably construct portfolios that exploit the reversion of stock prices at horizons longer than one day. But it is telling that reversion can be detected in even the simplest strategy. The benefit of this final approach, which combines the time-series with the cross-section, is that it does not require one to take a stand on the fundamental factors driving long-horizon returns.

These final tests can also help distinguish between two nonfundamentals-based explanations of excess comovement articulated by Barberis, Shleifer and Wurgler (2005). In the first explanation, also adopted in the current article, comovement of stock returns comes from correlated investor demand shocks for a particular group of securities.[3] In the second explanation, dubbed "information diffusion," changes in short-term comovement come from differences in the speed at which security prices reflect new information. This explanation says that index members

[3] Barberis, Wurgler and Shleifer (2005) further break down this first theory into an "investor habitat" view and a "category" view. Since both have similar theoretical implications, I treat them as subsets of a demand-based theory of return comovement.

incorporate information about aggregate earnings today, while nonindex stocks incorporate this information with a lag. Both explanations are consistent with index additions (deletions) experiencing increases (decreases) in contemporaneous short-term comovement with other index stocks. However, my reversion tests can discriminate. In the information diffusion explanation, index membership increases the overall efficiency of the pricing process, because nonindex stocks do not incorporate new information quickly enough. In contrast, the investor demand theory states that index members exhibit *excess* comovement, meaning that they are subject to mispricing. To the extent that a trading strategy earns profits from the reversion of excess comovement, the results provide support for the demand view rather than information diffusion.

The results in this article relate to a growing literature on financial contagion, in which researchers have argued that linkages between national stock markets may be related to return comovement during financial crises. A series of empirical papers (Forbes and Rigobon (2002), Corsetti, Pericoli and Sbracia (2002), Bekaert, Harvey and Ng (2005), Bekaert, Hodrick and Zhang (2005)) ask whether increases in cross-market correlations are evidence of excess comovement or whether they can be justified within standard factor models. Following this work, a growing theoretical literature argues that liquidity and wealth constraints (Yuan (2005), Kyle and Xiong (2001), Pavlova and Rigobon (2005)), information asymmetries (King and Wadhwani (1990), Pasquariello (2007)), and combinations thereof may lead asset returns to comove excessively.[4] Recently, the insights of this literature have been applied to more pervasive forms of comovement, such as the form studied in this article. To a large extent, however, rational expectations models are unable to account for my findings because they postulate that investors behave optimally given liquidity and information constraints. This can be contrasted with my results, where an investor with relatively little capital and access only to information on past prices can earn economically significant profits by trading on the reversion of excess comovement.

The article proceeds as follows. Section 1 lays out the basic predictions. Section 2 describes the data and the Nikkei 225 index methodology. Section 3 examines determinants of contemporaneous comovement with index and nonindex stocks, using a large sample of stocks traded in Japan. Section 4 analyzes the relationship between excess comovement and predictability of future returns. Section 5 concludes.

[4] For related theoretical work, see also Kodres and Pritsker (2002) and Mondria (2006).

1. Security Demand and Comovement of Stock Returns: Cross-Sectional Predictions

This section outlines a set of cross-sectional predictions concerning the relation between index overweighting and the comovement of stock returns. These predictions come from a simple limits-to-arbitrage model in which uninformed investor demand shocks occur in proportion to a weighting vector. Although the *level* of index demand in any period is random, the proportionality between index demand shocks generates a cross-sectional relation between index weights and the comovement of stock returns.

The predictions that I test, although novel, can be generated within a theoretical framework that has been developed in other papers (e.g., Hong and Stein (1999), Barberis and Shleifer (2003), Barberis, Shleifer and Wurgler (2005), Greenwood and Sosner (2007)). Therefore, I outline only the basic setup here and describe the predictions in general terms.

1.1 Predictions

Barberis and Shleifer (2003), Barberis, Shleifer and Wurgler (2005), and Greenwood and Sosner (2007) model a capital market with a finite number N of risky securities and a risk-free asset in elastic supply. Each risky security represents a claim on an uncertain terminating dividend, to be paid at a point far in the future. From a fundamental perspective, these securities are identical in that they share exposure to a market factor and additionally have uncorrelated sources of unique risk.

Two types of agents operate in the capital market—index traders and arbitrageurs. Index traders arbitrarily transfer funds into and out of M ($M < N$) index stocks. In Barberis and Shleifer (2003), these traders tend to purchase index stocks when they are selling the remaining $N - M$ stocks. A simpler implementation is described in Barberis, Shleifer and Wurgler (2005), and Greenwood and Sosner (2007), in which these traders buy (sell) the risk-free asset when they sell (buy) the M stocks.

When index traders have positive demand for index assets, arbitrageurs trade against them, but their risk aversion and finite horizons force them to ask for higher prices. Thus arbitrageurs dampen, but do not eliminate, the effects of trading by index traders.

Because the subset of index stocks are subject to correlated demand shocks, their returns comove in excess of what would be implied by fundamentals. Thus, one central prediction is that when securities change categories (i.e., become a member of an index, when they were not one before), their comovement with the other securities in that category increases.

These models have obvious cross-sectional extensions. If the degree of commonality in demand shocks varies across securities, this variation is cross-sectionally related to the amount of excess comovement. To make this concrete, consider the following simple example. Suppose the economy contains two index assets, A and B, and one nonindex asset C, all in fixed and identical supply. Security A is overweighted in the index, so that when index investors buy (or sell) the index, they purchase (or sell) 10 times as many shares of A as of B. For the moment, assume that these investors trade into and out of assets A and B by purchasing the risk-free asset. After controlling for common fundamentals, the covariance of asset A with the equal-weighted return of A and B should exceed the covariance of asset B with the equal-weighted return of A and B, and both should exceed the covariance of asset C with the equal-weighted return of A and B. This prediction is a simple consequence of asset A having a higher loading on index demand than B. This leads to the first new testable hypothesis.

Hypothesis 1. *For an index security i, the covariance of returns with the returns of other index securities is increasing in index weighting.*

One can extend the intuition described above to a model in which index investors transfer assets between competing stock indexes. Suppose that in addition to the index of A and B, described above, there exists an index which contains asset C as its only security. Suppose also that index investors transfer assets between the index and asset C at random. As before, the covariance of returns of asset A with the equal-weighted return of A and B should exceed the covariance of returns of asset B with the equal-weighted return of A and B. However, after controlling for fundamentals, the returns of asset C should covary negatively with the equal-weighted return of assets A and B. Moreover, the returns of asset C should have higher negative covariance with the returns of asset A than with the returns of asset B, because of asset A's overweighting in the index. This is the second new testable hypothesis.

Hypothesis 2. *The covariance of returns between index security i and nonindex returns is decreasing in the index weighting of i.*

The third set of predictions relates to the covariance of returns with lagged index returns. Prices must revert predictably to fundamentals at longer horizons. Therefore, returns of index stocks that move together too much contemporaneously must move apart over extended periods. Cross-sectionally, this implies that stocks that comove more contemporaneously, due to higher exposure to demand shocks, should also display more reversion.

Hypothesis 3. *The covariance between the returns of index security i and lagged index returns becomes more negative with higher index weighting.*

This last hypothesis is not present in Barberis, Shleifer and Wurgler (2005), or in Greenwood and Sosner (2007), which are concerned primarily with the contemporaneous comovement of stock returns rather than the way in which the mispricing is eventually corrected. Nor does it feature prominently in the extensive literature on financial contagion, which has been focused on identifying excess comovement in *contemporaneous* returns. A corollary of Hypothesis 3 is that a portfolio that has long overweighted stocks and short underweighted stocks will exhibit positive abnormal performance following index declines, and negative abnormal performance following index climbs.

1.2 Empirical strategy

The basic empirical strategy is to relate estimates of comovement with index returns to lagged measures of overweighting in the Nikkei 225 stock index. Comovement is measured as the covariance of security returns with the equal-weighted return of the other securities in the Nikkei 225. I use the equal-weighted return, rather than the actual index return, for two reasons. First, covariance with the equal-weighted return provides a measure of the *average* comovement with all index stocks. Second, by weighting all returns equally, I avoid inducing a mechanical relation between index weight and comovement that comes from overweighted stocks contributing more to index return. For the purposes of notation, "index return" refers to the return of the actual Nikkei 225 index, while "equal-weighted index return" refers to the equal-weighted return of the 225 members of the Nikkei index. These two time series are 93% correlated during the sample period, and the main results are similar using either measure.[5]

To generate Hypotheses 1, 2, and 3, I assume that all securities are identical and present in equal supply. In this setting, any index security with a high weight, or exposure to index demand shocks, is said to be "overweighted." In practice, however, securities are available in different supplies and therefore overweighting should be measured relative to a security's ability to absorb demand without a large change in the price. Although I experiment with several measures of overweighting, the one used in most of the empirical tests is the log ratio of the index weight to the weight the stock takes in a value-weighted index.

2. Data

This section describes the data used in the study. The first part outlines sample construction. The second part discusses index methodology and the calculation of overweighting.

[5] An alternative approach would be to estimate comovement with respect to a surrogate Nikkei index computed using N-1 stocks.

Table 1
Descriptive statistics

	Mean	SD	Min	Max
	Main sample ($N = 298$)			
Length of time series (days)	2451	455	109	2609
Fraction of period in Nikkei (%)	75.50	32.16	4.18	100.00
Daily return (%)	1.07	3.95	−14.24	25.73
Daily turnover (%)	1.00	5.57	0.02	68.47
Size (¥ million)	910, 229	1, 797, 490	19, 759	18, 884, 907
Ln (P/B)	0.57	0.42	−0.45	2.70
Leverage	0.94	0.14	0.00	1.00
	Other stocks ($N = 1458$)			
Length of time series (days)	2154	780	0	2609
Daily return (%)	0.43	0.10	−81.19	219.57
Daily turnover (%)	0.14	0.37	0.00	12.64
Size (¥ million)	145, 698	581, 337	1, 633	9, 912, 280
Ln (P/B)	0.28	0.57	−2.52	3.62
Leverage	0.82	0.25	0.00	1.00

Cross-sectional mean, standard deviation, and extreme values of time series averages of selected variables. The main sample includes 298 stocks that were members of the Nikkei 225 index for at least one day during the period from 1 September 1993 through 1 September 2003. This period includes a total of 2609 trading days. The second sample includes 1458 stocks in Japan that provided at least two years of returns data and that were never members of the index between September 1993 and September 2003. The length of the time series is the number of days for which each stock provides volume and price data. The fraction of the sample in Nikkei 225 is the percentage of the time that the stock was a member of the Nikkei 225 stock index. Daily return is the time series average of returns for each stock. Turnover is the average daily trading volume expressed as a percentage of total shares outstanding. Size is equal to the time series average of market capitalizations, in millions of yen. The log price-to-book value is the natural log of the share price divided by the book value per share for the appropriate financial year end, adjusted for capital changes. Leverage is the ratio of long-term debt to long-term debt plus common (book) equity. All data are collected from Datastream. The history of index membership is constructed using the index membership changes given on the *Nihon Keizai Shimbun* webpage.

2.1 Sample Construction

The main sample consists of 298 stocks that were present in the index for at least 200 days between 1 September 1993 and 29 August 2003. This period is chosen because stock return and volume data is available for each stock using Datastream. Prior to 1993, I am unable to collect comprehensive data on returns for all index stocks.

Table 1 outlines the composition of this sample. Of the 298 stocks, 225 are in the stock index at any point in time.[6] Out of 225, there are 164 stocks in the index for the entire sample period. Seventy-three stocks were added to the index between 1993 and September 2003, and 73 were deleted.

Panel A of Table 1 shows the cross-sectional mean, standard deviation, and extreme values of time series averages of selected descriptors of the stocks in the sample. On average, each of the 298 stocks was in the Nikkei 225 stock index for 76% of the sample. Index stocks experience a moderate

[6] During the entire 10-year period, there are a total of 39 days during when less than 225 stocks were present in the Nikkei 225 index. These periods typically occur following an unexpected deletion.

amount of trading, with an average of 1% of shares trading daily. The table also reports data on various characteristics used as controls in the cross-sectional regressions. The average stock has a market capitalization of about ¥910 billion (approximately US$ 7.6 billion at the time of writing), a log price-to-book ratio (share price divided by book value per share) of 0.57, and a leverage ratio (long-term debt to long-term debt plus common equity) of about 0.90.

Although not part of the main sample, I collect data on 1458 other stocks that were not in the index at *any* point between 1993 and 2003. Although most of these stocks are smaller than the median index constituent, some are much larger. Panel B of Table 1 shows that these stocks have lower average returns, trade less frequently, and are somewhat smaller than the index constituents.

2.2 Nikkei 225 index methodology and overweighting

The value of the Nikkei 225 ($P_{N225,t}$) on day t is determined by adding the prices ($P_{i,t}$) of its constituents, divided by the face value (FV_i), times a constant, dividing the total by the index divisor (D_t):

$$P_{N225,t} = \frac{\sum_{i=1}^{225} \frac{P_{i,t}}{FV_i/50}}{D_t}. \tag{1}$$

Most stocks have a face value of 50, though some have face values of 500, 5000, or 50,000. The index divisor is adjusted daily to account for stock splits, capital changes, or stock repurchases. It is designed to preserve continuity in the index, though not necessarily in the index weights of its constituents. For example, following a two-for-one stock split of an index constituent, the effective weight of the stock falls by half, while the divisor is changed to keep the Nikkei index value unchanged.

After adjusting for face value, the index value is equal-weighted in prices. This means that the index return, denoted by R_{N225}, is the price-weighted average of the returns of its constituents

$$R_{N225,t} = \sum_{i=1}^{225} w_{N225,it} R_{it} \tag{2}$$

where index weights, w_{it}, are given by

$$w_{N225,it} = \frac{\frac{P_{it}}{FV_i/50}}{\sum_{j=1}^{225} \frac{P_{jt}}{FV_j/50}}. \tag{3}$$

$w_{N225,it}$ can be interpreted as the cash value of stock i held by an investor at time t who owns 1 yen worth of the index.

Table 2 describes index weights between 1993 and 2003. Because index composition varies over time, a single cross-section would leave out 73 of the 298 securities. Therefore, for the purposes of the table, I compute index weights for each stock at the time the stock enters the index, or on 1 September 1993, whichever comes later. The table shows that most stocks have a face value of 50, but four have a face value of 500, two have a face value of 5000, and 12 have a face value of 50,000.

Starting with the stocks of face value 50, the average price is 1491, yielding an average Nikkei 225 weight of 0.58%. As a measure of overweighting, I calculate the ratio of each stock's weight in the Nikkei to its weight in the market value-weighted index

$$\frac{w_{N225,it}}{w_{VW,it}} = \frac{\frac{P_{it}}{FV_i/50}}{\sum_{j=1}^{225} \frac{P_{jt}}{FV_j/50}} \Bigg/ \frac{MV_{it}}{\sum_{k=1}^{N^*} MV_{kt}} \tag{4}$$

where $w_{VW,it}$ denotes the weight of stock i in the value-weighted index, MV_{it} denotes the market value of stock i, and N^* denotes the total number of tradable stocks on section 1 of the TOPIX. The value-weighted TOPIX index is proportional to market capitalization of Section 1 stocks, so the overweighting measure in Equation (4) can be calculated by dividing Nikkei weights by TOPIX index weights. The table shows that on average,

Table 2
Nikkei 225 index composition and weights

Sample	N	Price	$\frac{\text{Price}}{\text{Face Value}}50$	w^{N225}	$\frac{w^{N225}}{w^{VW}}$	Overweighting $\log\left(1+\frac{w^{N225}}{w^{VW}}\right)$
Face value 50	280	1491	1491	0.58	6.00	1.62
Face value 500	4	3320	332	0.18	0.83	0.46
Face value 5000	2	1,098,500	10, 985	4.42	5.83	1.92
Face value 50,000	12	987,583	988	0.34	0.38	0.30
Full sample	298	48,586	1518	0.59	5.70	1.55

Relationship between prices and Nikkei 225 weights for the 298 stocks present in the Nikkei 225 stock index for at least one day between 1 September 1993 and 1 September 2003. The level of the Nikkei 225 stock index is given by

$$p_{Nikkei,t} = \frac{1}{D_t} \sum_{i=1}^{225} \frac{P_{i,t}}{F_i/50}$$

where D_t is the Nikkei 225 divisor, $P_{i,t}$ is the price of stock i on day t, and F_i is the face value of stock i ranging from 50 to 50,000. The table reports average prices of stocks, where the price is taken from the first day of the sample in which a stock is present in the index. The next column reports mean stock price normalized by face value—this is the form in which prices enter the Nikkei 225 index calculation. The next column reports the average weight in the Nikkei index, the ratio of the face value adjusted price to the sum of the face value adjusted prices. The second-to-last column reports the average ratio of the Nikkei index weight to the weight in the market value-weighted TOPIX index. The final column reports the average of the log of the ratio in the previous column. This is the measure of overweighting used in the article. Reported averages are broken down by face value (50, 500, 5000, or 50,000).

the Nikkei weight exceeds the TOPIX weight by a factor of 5.7. The average ratio of the Nikkei weight-to-market-value weight can exceed 1 for two reasons. First, the Nikkei index includes less than 100% of the traded securities in Japan. Second, smaller stocks tend to be overweighted in the Nikkei, but receive equal weighting in reported averages.

Ultimately, the quantity of interest is not the *average* over- or under-weighting of stocks in the Nikkei 225 in this article, but cross-sectional differences in weighting across stocks. Panel A of Figure 1 shows the histogram of index weights, using data from the 225 index constituents at the (exact) midpoint of the sample. The distribution is heavily left skewed, with most stocks taking index weights between 0 and 0.15%. Panel B shows that the skewness of the weight distribution is reduced substantially when I scale the weights by market-value weight, following Equation (4).

I define overweighting, OW, as the log of one plus the ratio of Nikkei weight to TOPIX index weight:

$$OW_{it} = \log\left(1 + \frac{w_{N225,it}}{w_{MV,it}}\right) \tag{5}$$

The distribution of overweighting for the 225 stocks in the Nikkei 225 index at the midpoint of the sample is shown in Panel C of Figure 1. Variation in this variable comes from two sources: differences in adjusted prices of the securities, and differences in market capitalizations. Empirically, cross-sectional variation in market capitalizations accounts for more than half the variation in this variable. Note that because Nikkei 225 weight is zero for nonindex stocks, overweighting is equal to zero for these stocks ($\log(1 + 0)$). The last column of Table 2 summarizes overweighting for the entire sample of 298 securities. Note that overweighting is not sensitive to stock returns—index stocks that perform well find their Nikkei 225 weight and market-value weight increasing in proportion. As a result, the cross-sectional ranking of overweighting is stable over the sample, despite wide variation in returns over the 10-year period.

For the remainder of the article, "overweighting" refers to the quantity computed in Equation (5). The overweighting vector OW_t, serves an instrument for the true vector of demand shocks, which is unobserved.

2.3 The practice of index investing in Japan

The Nikkei 225 stock average is Japan's most widely watched index of stock market activity, and one of two common benchmarks for institutional investors. Although aggregated data on institutional tracking of the index are not available, Nomura Securities estimates that 2.43 trillion yen were benchmarked to the Nikkei as of April 2000.[7] Consistent with the Nikkei

[7] "Potential impact of the change in stock selection criteria for the Nikkei Average: High-priced tech stocks to be included" Nomura Research Briefs, April 18, 2000.

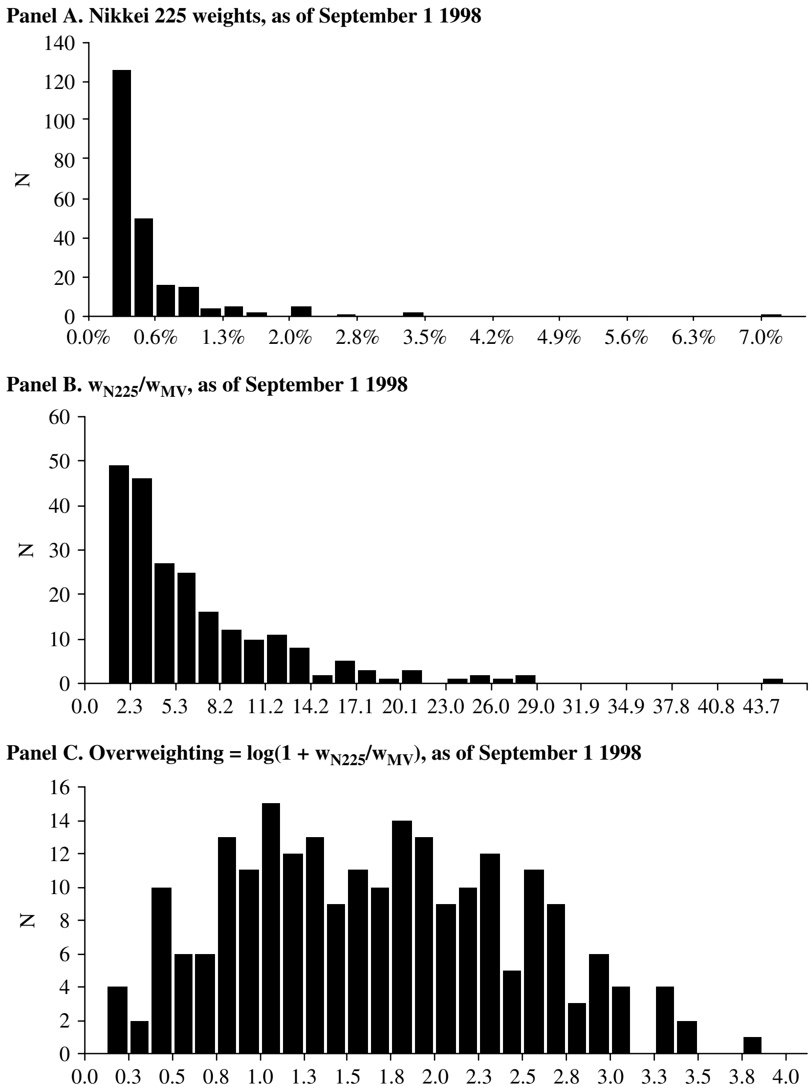

Figure 1
Nikkei 225 index weights and overweighting
Cross-sectional distribution of index weights of the stocks in the Nikkei 225 index on 1 September 1998, the midpoint of the panel. Panel A shows the distribution of raw index weights for the 225 stocks. Panel B shows the distribution of the ratio of index weights to the weight that each stock takes in the market value-weighted TOPIX index. A value greater than 1.0 indicates overweighting relative to the value-weighted TOPIX. Panel C shows the distribution of the overweighting measure, equal to the natural log of one plus the ratio of index weights to the weight that each stock takes in the value-weighted TOPIX index.

225 being an important trading universe for investors, Greenwood (2005) shows that additions to the Nikkei 225 experience average abnormal returns of over 10%.

One reason for the popularity of the index is that a liquid Nikkei 225 futures contract trades in both Osaka and Chicago, making it a convenient way for international investors to gain exposure to Japan. In contrast, the broader value-weighted TOPIX index is traded only in Tokyo. Although futures contract trading volume is not the sole indicator of the importance of the index, I estimate that an average of 581 billion yen of Nikkei 225 futures are traded per day, compared with 271 billion yen of the TOPIX contract.[8]

3. Security Demand and the Cross-Section of Comovement

This section analyzes the cross-sectional determinants of the comovement among index stock returns and between index and nonindex stock returns. I begin by analyzing the determinants of comovement for a single cross-section of stocks in the middle of the sample. A more systematic analysis of the full panel then follows.

3.1 Graphical analysis of comovement for a single cross-section

I start by studying the determinants of index comovement for the 225 stocks present in the Nikkei 225 index on 1 September 1998, the midpoint of my sample. For each of the 225 stocks, I estimate univariate time series regressions of stock returns on the equal-weighted return of stocks in the Nikkei:

$$R_{it} = \alpha_i + \beta_{EWN225,i} R_{EWN225,t} \tag{6}$$

To compute $R_{EWN225,t}$, the equal-weighted Nikkei return, I use only the 168 stocks that remained in the index throughout my sample period 1993 through 2003. This avoids the possibility of identifying comovement from changes in the index.[9]

The slope parameter from Equation (6) is a measure of the comovement of the stock return during the previous 100 days with the equal-weighted index return. An alternate measure of comovement is the R^2 from this regression, which captures the correlation between a stock's return and the return of all other stocks in the index. Panel A of Figure 2 plots equal-weighted index beta, β_{EWN225}, against overweighting, OW_{t-1}.

[8] These figures are based on average yen-denominated trading volume for the period September 2003 through August 2005. Yen-denominated volume is the number of contracts traded times the closing price of the contract, times the notional value of the contract.

[9] The results are somewhat stronger if I replace the equal-weighted index return with the actual Nikkei 225 return.

Overweighting is lagged by 100 days to ensure that the results are not driven by changes occurring during the estimation period for β. As a practical matter, this timing is unimportant, because there is very little time series variation in the cross-sectional ranks of weights.

Panel A shows a strong positive relation between index overweighting and index beta. In the cross-sectional regression that corresponds to this figure (unreported), the slope is 0.28, and the R^2 is 0.25. Combined with a cross-sectional standard deviation of index overweighting of 0.78, the figure shows that a 1 standard deviation increase in overweighting is associated with a change in index beta of 0.22.

Panel B plots R^2 against lagged overweighting. Again, the figure shows a strong positive relation between these two variables. This shows that the cross-sectional relation is driven by increased correlation of returns with index returns for overweighted stocks, rather than by their increased variance.

I next study the relation between index overweighting and the *conditional* comovement of stock returns with stocks inside and outside of the index. Using the same cross-section, I estimate bivariate time series regressions of stock returns on the equal-weighted index return and the return on the value-weighted TOPIX index:

$$R_{it} = \alpha_i + \beta^*_{EWN225,i} \cdot R_{EWN225,t} + \beta_{TOPIX,i} \cdot R_{TOPIX,t} + \varepsilon_{it}. \qquad (7)$$

The slope parameter β^*_{EWN225} is a measure of the conditional comovement of the stock return during the previous 100 days with the equal-weighted index return, and thus a more precise estimate of comovement that controls for possibly time varying exposure to the market risk factor. The slope parameter β_{TOPIX} is a measure of comovement with non-Nikkei stocks.

Panel A of Figure 3 plots β^*_{EWN225} against lagged overweighting. Consistent with the previous results, the figure shows a strong positive relation between these two variables. Figure 3 is stronger evidence for the basic predictions than Figure 2, because it controls for variation in firms' exposure to the market risk.

Panel B of Figure 3 plots β_{TOPIX} against lagged overweighting. Confirming *Hypothesis 2*, the figure shows a strong negative relation between index overweighting and the conditional comovement of index stocks with stocks outside of the index. The strength of the result is consistent with the index investors pulling funds out of other stocks in order to invest in index stocks.

3.2 Panel Analysis: comovement among Nikkei 225 stocks

I now expand the cross-sectional approach taken above to study comovement for the full panel of stock returns between 1993 and 2003. The basic approach is as follows. Every 100 days starting on 20 January 1994 (100 days after the first day of returns), I estimate time series regressions of

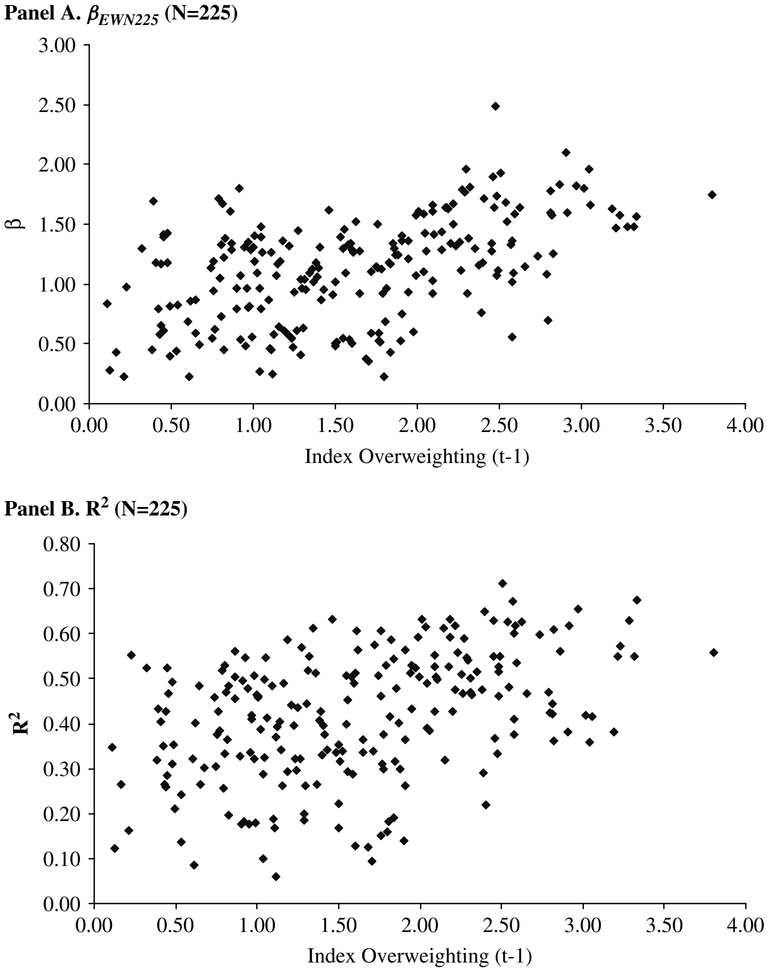

Figure 2
Comovement and Nikkei 225 index overweighting: Univariate regressions
Measures of comovement of stock returns with returns of other stocks in the Nikkei 225 are plotted against overweighting of the stock in the index. In Panel A, comovement is measured as the slope (β_{EWN225}) from a time series regression of stock returns on the equal-weighted return of stocks remaining in the Nikkei 225 throughout the sample period. In Panel B, comovement is measured as the R^2 from this regression. Index overweighting is defined as the natural log of one plus the ratio of a stock's weight in the Nikkei 225 to the weight of the stock in the value weighted TOPIX index. Index overweighting is measured one day before the start of the sample of returns used to estimate comovement.

$$R_{it} = \alpha_i + \beta_{EWN225,it} \cdot R_{EWN225,t} + \varepsilon_{it}$$

Both plots show comovement for the entire sample of Nikkei 225 stocks on 1 September 1998.

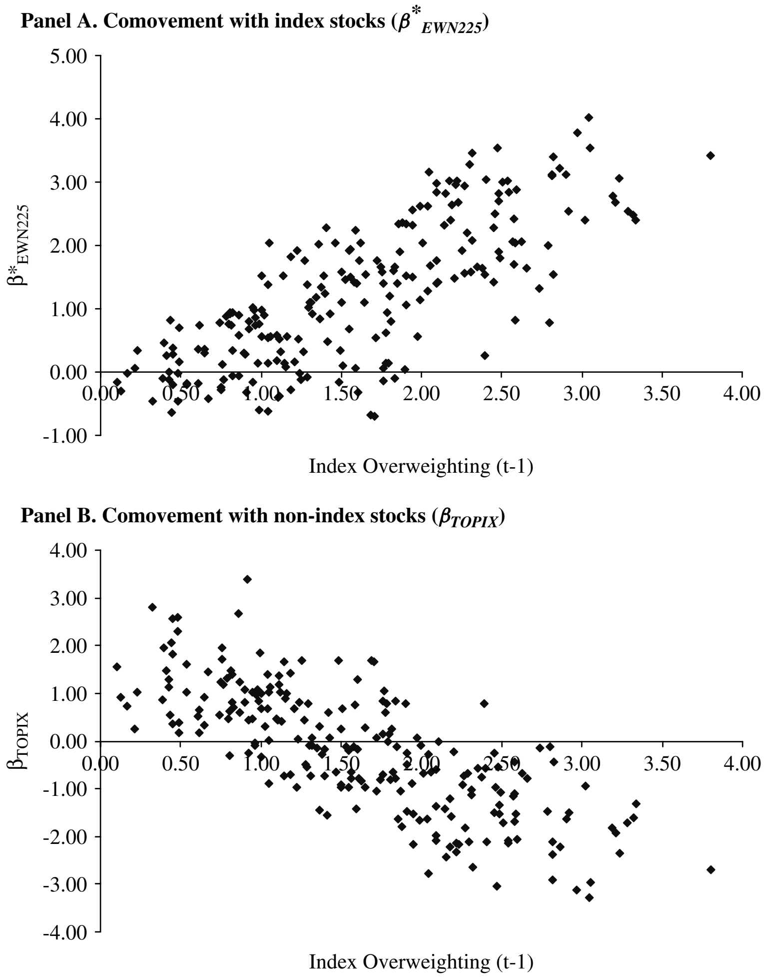

Figure 3
Comovement and Nikkei 225 index overweighting: Bivariate regressions
Measures of comovement of stock returns with returns of other stocks in the Nikkei 225 and with returns of stocks outside of the Nikkei 225 are plotted against overweighting of the stock in the index. The figure plots the slope coefficients β_{EWN225} and β_{EWN225} from the regression

$$R_{it} = \alpha_i + \beta^*_{EWN225,i} \cdot R_{EWN225,t} + \beta_{TOPIX,i} \cdot R_{TOPIX,t} + \varepsilon_{it}$$

against index overweighting, measured one day before the start of the sample of returns used to estimate comovement. Index overweighting is defined as the natural log of one plus the ratio of a stock's weight in the Nikkei 225 to the weight of the stock in the value-weighted TOPIX index. Regression slope parameters are estimated using 100 days of data. Panel A plots β_{N225} for every stock in the sample on 1 September 1998, the midpoint of the sample, against the index overweighting 100 days before. Panel B plots regression slope parameter β_{TOPIX} against the index overweighting 100 days before.

returns on the equal-weighted index return for the past 100 days, following Equation (6). I estimate these regressions for every security in the sample, provided it has complete returns data over the previous 100 days, and provided it is in the index during the entire estimation period. The latter restriction ensures that I only exploit cross-sectional variation within index securities, thereby avoiding the possibility of attributing results to differences between the index betas of index and nonindex stocks. The full 10-year panel contains 26 cross-sections, with an average of 222 stocks in each. For each security and each cross-section, I record the beta from this regression, as well as the associated R^2.

The next step is to relate these measures of comovement in each cross-section to the index overweighting at the start of the period. To do this, I run 26 cross-sectional regressions of the comovement measures from Equation (6) on index overweighting at the start of the period:

$$\beta_{EWN225,it} = a_t + b_t \cdot OW_{it-1} + u_{it}, \tag{8}$$

and

$$R^2_{it} = a_t + b_t \cdot OW_{it-1} + u_{it}. \tag{9}$$

Note that β and R^2 now have t (as well as i) subscripts, indicating that they are specific to a 100-day period of returns.

The first line of Table 3 shows the time series average of coefficients a and b from Equation (8), as well as their associated t-statistics, following Fama and MacBeth (1973). The average sensitivity of index beta to overweighting is a highly significant 0.218. Moreover, the table indicates that the time series average of R^2 from this cross-sectional regression is 0.21. Thus, on average, index overweighting explains more than a fifth of the cross-sectional variation in daily comovement among index stocks.

What is the economic significance of b, the coefficient on overweighting? Independent of an estimate of the variance of exogenous index trader demand, the coefficient does not imply an estimate of the slope of the demand curve for Nikkei stocks. Put differently, b can be high either because demand curves for Nikkei stocks are steep, or because these stocks frequently experience large index trader shocks. Previous studies on the demand curves for Nikkei stocks suggest that their demand curves are steep (e.g., Greenwood (2005)).

I next alter the second stage regressions to include controls for other common factors in returns. Banz (1981) and Fama and French (1992)

Table 3
Nikkei 225 overweighting and comovement measures

Y	*a*	[*t*-stat]	*b*	[*t*-stat]	*c*	[*t*-stat]	*d*	[*t*-stat]	*N*	R^2
β_{EWN225}	0.719	[24.38]	0.218	[10.17]					222	0.21
β_{EWN225}	2.401	[9.92]	0.048	[1.84]	0.005	[1.22]	−0.113	[−6.66]	222	0.30
R^2	0.269	[12.98]	0.065	[8.21]					222	0.15
R^2	0.340	[4.50]	0.053	[6.44]	0.005	[2.14]	−0.004	[−0.84]	222	0.19

Average parameter estimates and Fama and MacBeth (1973) *t*-statistics from rolling cross-sectional regressions of comovement measures on index overweighting at the start of the period

$$\beta_{EWN225,it} = a_t + b_t \cdot OW_{it-1} + c_t \cdot \ln(P/B)_{it-1} + d_t \cdot Size_{it-1} + u_{it}$$

$$R^2_{it} = a_t + b_t \cdot OW_{it-1} + c_t \cdot \ln(P/B)_{it-1} + d_t \cdot Size_{it-1} + u_{it}$$

β_{EWN225} denotes the slope parameter from a regression of daily stock returns on the equal-weighted return of stocks that were in the Nikkei 225 index for the entire sample period, and R^2 denotes the R^2 from this univariate regression. First-stage regression estimates for period t are computed using returns from the interval $[t-99, t]$ and are estimated for each stock for each 100-day interval. The table shows average coefficients from second stage cross-sectional regressions of β_{EWN225} and R^2 on lagged independent variables. These include Nikkei 225 overweighting at the start of the period (OW), defined as the natural log of one plus the ratio of index weight to the weight the stock would have taken in a value weighted index; the natural log price-to-book ratio (P/B); and the natural log of market value (*Size*). All independent variables are measured at $t-100$, with the exception of the log price-to-book ratio, which is measured in December of the previous year. Overweighting is defined to be zero for firms that were not in the index at the start of the period. The table also indicates the average number of firms in each cross-section (N), as well as the time series average of R^2 from the cross-sectional regressions.

find that commonality in average returns can be attributed to size and the book-to-market ratio. I thus run a second set of regressions that control for the (log of) size and the price-to-book ratio of each stock:

$$\beta_{EWN225,it} = a_t + b_t \cdot OW_{it-1} + c_t \cdot \ln(P/B)_{it-1} + d_t \cdot Size_{it-1} + u_{it} \tag{10}$$

where $\ln(P/B)$ is the log of price-to-book ratio and *Size* is the log of market capitalization. Comovement of a stock with other index members is negatively related to the price-to-book ratio, and negatively related to log size. As a large part of the variation in overweighting comes from market capitalization, it is not surprising that the size control lowers the direct effect of overweighting. The coefficient b on overweighting falls to 0.048, but remains significant. By including size as an independent variable, the regression is now able to identify the unscaled effect of price on comovement. Even by itself, lagged price is significantly related to the comovement of stock returns. The fact that size enters the regression with a negative sign is further support of the basic claim that overweighted securities tend to have higher betas.

The next two lines of the table show the time series average of coefficients from the cross-sectional regressions of R^2 on lagged index weight and controls. Like the previous results, index overweighting explains a significant fraction of the correlation between the returns of index stocks.

3.3 Panel analysis: comovement with nonindex stocks

Hypothesis 2 says that the covariance of returns between index security i and nonindex returns is decreasing in the index weighting of i. I test this proposition using a two-step procedure.

In the first stage, I generate repeated time specific cross-sectional estimates of conditional comovement of each stock with the equal-weighted index return and with return of the value-weighted TOPIX index. Every 100 days starting on 20 January 1994, I estimate bivariate time series regressions of returns on the equal-weighted index return and the value-weighted TOPIX return:

$$R_{it} = \alpha_i + \beta^*_{EWN225,it} \cdot R_{EWN225,t} + \beta_{TOPIX,it} \cdot R_{TOPIX,t} + \varepsilon_{it} \tag{11}$$

These regressions are estimated for every security in the panel, provided they have complete returns data over the estimation period, and provided their index status does not change during the estimation period. For each security and each cross-section, I record β^*_{EWN225}, the conditional comovement of stock returns with the equal-weighted index return, and β_{TOPIX}, the conditional comovement of stock returns with stocks outside of the index.

In the second stage, I relate these measures of comovement to lagged index overweighting and controls:

$$\beta^*_{EWN225,it} = a_t + b_t \cdot OW_{it-1} + c_t \cdot \ln(P/B)_{it-1} + d_t \cdot Size_{it-1} + u_{it} \tag{12}$$

These results are in Table 4. The univariate relation between conditional comovement and overweighting appears stronger than in the previous table, with the R^2 rising to 39%. Thus, controlling for fundamentals in a crude way increases the significance of the results.

Panel B shows average coefficients from the regression of comovement with nonindex stocks on overweighting and controls:

$$\beta_{TOPIX,it} = a_t + b_t \cdot OW_{it-1} + c_t \cdot \ln(P/B)_{it-1} + d_t \cdot Size_{it-1} + u_{it} \tag{13}$$

Line 1 shows that there is a strong negative relation between the comovement of a stock's returns with the returns of stocks outside of the Nikkei, and that stock's weight in the Nikkei index. Line 2 shows that this relation continues to hold after controlling for log price-to-book and log size.

Finally, in unreported regressions I repeat the regressions in Equations (12) and (13), substituting a small-minus-large stock portfolio, a momentum portfolio, and industry-matched portfolios for the value-weighted market return. The results of these regressions are similar to those reported in Table 4. Interestingly, in a final test, I include all these controls at once, in which case the relation between comovement and overweighting becomes stronger. It is not surprising that more elaborate controls for fundamentals increase the ability of demand-based variables to explain comovement.

3.4 Expanded sample analysis

So far, the sample has been constrained to include only index stocks. The benefit of this constraint is that I avoid assigning cross-sectional variation in index weights to differences in the average index weights between index and nonindex stocks. In this section, I formally identify the distinct roles of index weighting and index membership. To do this, I expand the basic sample with an additional 1458 stocks. With the expanded sample, I modify the second step of the two-step procedure to allow separate roles for index weighting and index membership in the cross-section.

The first stage remains unchanged: I estimate time series regressions of returns on the equal-weighted index return for the past 100 days, following Equation (6), collecting measures of comovement in each instance. The second stage regression is modified to include a dummy variable for index

Table 4
Nikkei 225 overweighting and conditional comovement

Y	*a*	[*t*-stat]	*b*	[*t*-stat]	*c*	[*t*-stat]	*d*	[*t*-stat]	*N*	R^2
Panel A: Daily conditional comovement with Nikkei 225 stocks										
β^*_{EWN225}	−0.159	[−2.30]	0.800	[16.28]					222	0.390
β^*_{EWN225}	6.748	[17.39]	0.136	[2.24]	−0.062	[−2.97]	−0.465	[−17.55]	222	0.614
Panel B: Daily conditional comovement with stocks outside of Nikkei 225										
β_{TOPIX}	1.178	[13.53]	−0.780	[−13.06]					222	0.330
β_{TOPIX}	−5.661	[−12.83]	−0.126	[− 1.81]	0.077	[2.54]	0.459	[15.29]	222	0.530

In the first stage, I jointly estimate the conditional comovement of a stock's return with the equal-weighted return of the stocks present in the Nikkei 225 index and the return on the TOPIX value-weighted index.

$$R_{it} = \alpha_i + \beta^*_{EWN225,it} \cdot R_{EWN225,t} + \beta_{TOPIX,it} \cdot R_{TOPIX,t} + \varepsilon_{it}$$

The table reports average parameter estimates and Fama and MacBeth (1973) t-statistics from second stage rolling cross-sectional regressions of conditional comovement on index overweighting at the start of the period as well as a set of control variables

$$\beta^*_{EWN225,it} = a_t + b_t \cdot OW_{it-1} + c_t \cdot \ln(P/B)_{it-1} + d_t \cdot Size_{it-1} + u_{it}$$

$$\beta_{TOPIX,it} = a_t + b_t \cdot OW_{it-1} + c_t \cdot \ln(P/B)_{it-1} + d_t \cdot Size_{it-1} + u_{it}$$

where β_{EWN225} and β_{TOPIX} are the conditional comovement estimates from the first-stage regressions. The independent variables in the second-stage regressions include Nikkei 225 overweighting at the start of the period (OW), defined as the log of one plus the ratio of index weight to the weight the stock would have taken in a value weighted index, the price-to-book ratio (P/B), and the log of market value (*Size*). All independent variables are measured at $t-100$, except for the price-to-book ratio, measured in December of the previous year. For each cross-section, the sample includes all firms present in the Nikkei index at the beginning and end of the estimation period. The table indicates the average number of firms in each cross-section (N), as well as the time series average of R^2 from the cross-sectional regressions.

membership, as follows:

$$\beta_{EWN225,it} = a_t + b_t \cdot OW_{it-1} + c_t \cdot 1_{OW_{it-1}>0} + d_t \cdot (P/B)_{it-1} + e_t \cdot Size_{it-1} + u_{it} \quad (14)$$

where OW denotes lagged index overweighting and the lagged indicator $1_{OW\,it-1>0}$ takes a value of one if the stock is in the Nikkei 225 index, and zero otherwise. The other controls are defined as before. I estimate this regression for each cross-section, reporting average coefficients in Table 5.

The table shows that even in the broader cross-section, index overweighting has a profound effect on the comovement of returns. The slope in the univariate regression between index beta and overweighting averages 0.296. When I add the control for index membership, the average slope is reduced to 0.218. After adding the additional controls, the slope remains approximately the same, at 0.202. Note that the size control is no longer significant. But for index stocks, size still enters the regression through its effect on the overweighting variable. The interpretation is therefore that smaller index stocks have higher index beta, while smaller nonindex stocks have lower index beta. This eliminates any remaining concerns that the result could be driven only by differences in firm size.

Finally, I apply this procedure to study the conditional comovement of returns with stocks outside of the index. For the expanded sample of stocks, I estimate bivariate time series regression of returns on the equal-weighted index return and the value-weighted TOPIX return, following Equation (7). Then, in repeated cross-sectional regressions, I study the determinants of these comovement measures:

$$\beta^*_{EWN225,it} = a_t + b_t \cdot OW_{it-1} + c_t \cdot 1_{OW_{it-1}>0} + d_t \cdot \ln(P/B)_{it-1} + e_t \cdot Size_{it-1} + u_{it} \quad (15)$$

and

$$\beta_{TOPIX,it} = a_t + b_t \cdot OW_{it-1} + c_t \cdot 1_{OW_{it-1}>0} + d_t \cdot \ln(P/B)_{it-1} + e_t \cdot Size_{it-1} + v_{it} \quad (16)$$

Like Equation (14), the regressions now include a dummy variable that takes on a value of one if the stock was a member of the Nikkei index, and zero otherwise. Average coefficients from Equations (15) and (16) are shown in the bottom panels of Table 5. Panel B shows the determinants of comovement with Nikkei 225 stocks (β^*_{EWN225}) and Panel C shows the determinants of comovement with stocks outside of the Nikkei 225 (β_{TOPIX}).

These results are similar to those in Panel A, only stronger. Index overweighting has a strong positive relation with comovement of a

Table 5
Nikkei 225 index overweighting and comovement: Expanded sample

Dependent variable (β)	First-stage regression	β Measures co-movement with:	a	[t-stat]	b	[t-stat]	c	[t-stat]	d	[t-stat]	e	[t-stat]	N	R^2
Panel A: Comovement with Nikkei 225 members (univariate)														
β_{EWN225}	Univariate	Nikkei stocks	0.577	[27.12]	0.296	[22.05]							1421	0.229
β_{EWN225}	Univariate	Nikkei stocks	0.572	[26.73]	0.218	[10.17]	0.147	[5.81]					1421	0.237
β_{EWN225}	Univariate	Nikkei stocks	0.656	[5.61]	0.202	[7.90]	0.186	[4.02]	0.030	[2.77]	−0.010	[−1.00]	1421	0.266
Panel B: Conditional comovement with Nikkei 225 members (multivariate)														
β^*_{EWN225}	Multivariate	Nikkei stocks	0.106	[2.54]	0.664	[20.50]							1416	0.295
β^*_{EWN225}	Multivariate	Nikkei stocks	0.115	[2.78]	0.800	[16.28]	−0.275	[−5.06]					1416	0.301
β^*_{EWN225}	Multivariate	Nikkei stocks	1.065	[3.81]	0.666	[9.52]	0.113	[1.07]	−0.030	[−1.80]	−0.091	[−3.92]	1416	0.368
Panel C: Conditional comovement with non-Nikkei 225 stocks (multivariate)														
β_{TOPIX}	Multivariate	Non-Nikkei stocks	0.634	[15.25]	−0.496	[−14.65]							1416	0.141
β_{TOPIX}	Multivariate	Non-Nikkei stocks	0.614	[14.71]	−0.780	[−13.06]	0.564	[7.14]					1416	0.158
β_{TOPIX}	Multivariate	Non-Nikkei stocks	−0.524	[−2.00]	−0.623	[−8.25]	0.099	[0.84]	0.082	[3.83]	0.106	[4.83]	1416	0.238

Average parameter estimates and Fama and MacBeth (1973) t-statistics from rolling nonoverlapping cross-sectional regressions of comovement of daily stock returns on index overweighting at the start of the period, a dummy variable for index membership, and controls

$$\beta_{EWN225,it} = a_t + b_t \cdot OW_{it-1} + c_t \cdot 1_{OW_{it-1}>0} + d_t \cdot \ln(P/B)_{it-1} + e_t \cdot Size_{it-1} + u_{it}$$

β denotes the comovement measure, estimated in one of three ways. In Panel A, β is the slope parameter from a univariate regression of stock returns on the equal-weighted return of stocks that were in the Nikkei 225 for the entire sample period. In Panel B, β is the coefficient on equal weighted Nikkei return, from a multivariate regression of stock returns on the equal-weighted return Nikkei 225 stocks and the broader value-weighted TOPIX index. In Panel C, β is the coefficient on the TOPIX return, from the same multivariate regression. Thus in Panel C, β measures comovement with stocks outside of the Nikkei stock index. In all cases, β is estimated using returns from the interval $[t-99,t]$, and the independent variables are measured at $t-100$, except for the price-to-book ratio, which is measured in December of the previous year. Each cross-section contains all the firms in Japan that contain complete data and for which β can be estimated. The independent variables include Nikkei 225 overweighting at the start of the period (OW), defined as the natural log of one plus the ratio of index weight to the weight the stock would have taken in a value-weighted index; a dummy variable indicating whether the firm is present in the Nikkei index (1_{OW}); the natural log of the price-to-book ratio (P/B); and the log of market value (*Size*). Overweighting is zero for firms that were not in the index at the start of the period. The last two columns indicate the average number of firms in each cross-section (N), as well as the time series average of R^2 from the cross-sectional regressions.

stock with stocks in the index, and a strong negative relation with the comovement of a stock with stocks outside of the index. More interestingly, the cross-sectional variation in index weights is enough to wipe out the importance of the index dummy, highlighting the benefit of my cross-sectional approach. In Panel B, for example, the coefficient on index membership is either insignificant or the wrong sign.

3.5 Robustness

This section examines whether altering any of the following—(i) the length of time over which comovement is estimated, (ii) the choice of control variables, (iii) the calculation of index returns, (iv) concerns about a nontrading bias—affects the results significantly.[10]

3.5.1 Length of time over which comovement is estimated. All the reported results are based on betas estimated using 100 days of daily return data. I expand the measurement window from 100 days to 250 days, and compute univariate equal-weighted Nikkei 225 betas according to Equation (6) and multivariate equal-weighted Nikkei 225 and TOPIX betas according to Equation (7). I then reestimate the first two lines from Table 3 and Table 4. The total number of cross-sections is now reduced from 26 to 10, causing the standard errors to increase; the main results remain.

3.5.2 Choice of control variables. Control variables are selected based on the two criteria. First, size and book-to-market ratio are both extensively documented as having pronounced effects on the cross-section of average returns in the USA (e.g., Banz (1981), Fama and French (1992)). Second, both are available from Datastream for the majority of the stocks in my sample.

A third control variable, not included in the main tests, is firm leverage. It follows from Modigliani and Miller (1958) that, controlling for asset risk, equity betas should be positively related to the degree of financial leverage. However, Hecht (2002) tests this proposition, finding that the US data do not confirm it. Nevertheless, I collect a proxy for leverage from Datastream (see Table 1 for summary information), and repeat the basic tests from Table 3. In Japan, leverage is positively related to the comovement of stock returns with the Nikkei, but the regression coefficient on lagged overweighting, *b*, is virtually unchanged after including this control.

3.5.3 Calculation of equal-weighted index return. All the empirical tests measure index comovement using an equal-weighted average of the returns of the stocks that remained in the index for the entire sample period. This

[10] Additional robustness tests on (i) alternate standard errors (ii) alternate measures of index overweighting were performed but are omitted here for brevity.

time series is 93% correlated with the actual Nikkei 225 return over the same period. It is also 99% correlated with the equal-weighted return of the Nikkei, and 88% correlated with the value-weighted return of index constituents. Although it may lack a compelling theoretical motivation, I experiment with replacing the actual Nikkei 225 return for the equal-weighted index return in Equations (6) and (7). I replicate the main results using each of these series, with no change in the results.

3.5.4 Nontrading and nonsynchronous trading bias. A nonsynchronous trading bias occurs when infrequently traded securities appear not to incorporate market information immediately, generating a positive correlation between security returns and lagged index returns, and a downward-biased estimate of contemporaneous beta. The bias is well-known and documented by Fisher (1966), Scholes and Williams (1977), Lo and MacKinlay (1990), and Ahn et al. (2002). In my sample, the returns of overweighted securities are negatively correlated with lagged index returns. Therefore, if one were to adjust for possible infrequent trading among these securities, it would further increase my contemporaneous measures of comovement.

Nevertheless, it is possible to replicate the results within groups of high- and low-volume Nikkei stocks. I find a stronger cross-sectional correlation between overweighting and index beta in the high volume group (but significant relations in both groups) inconsistent with the presence of a nonsynchronous trading bias. I also check the frequency of trading of all securities using a small sample of tick data in late 2004. For 99.536% of ticker-days, there is significant trading volume between 2 : 45 and 3 : 00 P.M., when the exchange closes. In short, a nontrading bias seems unlikely.

4. Excess Comovement and Predictability

While it is hard to think of plausible fundamentals-based explanations for the results, one is always left with the possibility that the cross-sectional variation in lagged index weights is related to underlying variation in economic factors driving returns. To address this concern, one can take two approaches. The first, taken earlier in the article, tries to isolate the fundamental component of returns by controlling for market returns when estimating comovement.

A second and more compelling approach is to argue that if returns comove too much today, then if they are to return to fundamental valuations, there must be implications for future returns. Put in the context of the Nikkei 225, this means that if overweighted stocks comove excessively with the index in period t, then during period $t + 1$ they must, on average, move in the opposite direction of the time t index return. Taken further, the degree of reversion should be related to their overweighting.

Consistent with this simple intuition, Barberis, Wurgler and Shleifer (2005), and Greenwood and Sosner (2007) document decreases in the correlation with lagged index returns following inclusion into a stock index. However, these results are difficult to interpret on their own, since they could simply reflect improvements in the speed at which firms incorporate information.

Another way of documenting reversion of excess comovement is to study comovement when returns are measured at different intervals. Taking this approach, Barberis, Wurgler and Shleifer (2005) show that changes in comovement are smaller upon addition to a stock index when the comovement measures are based on weekly or monthly returns. Table 6 shows the analog of their tests using my data. Consistent with the intuition above, the degree to which overweighting can explain variation in comovement is reduced, both economically and statistically, at longer horizons.

A cleaner test of reversion, however, is to exploit the cross-sectional variation in demand shocks. The basic intuition is that if overweighting is

Table 6
Comovement measured at longer horizons

Y	*a*	[*t*-stat]	*b*	[*t*-stat]	*N*	R^2
			Daily return comovement			
Nikkei only	0.719	[24.38]	0.218	[10.17]	0.21	222
All stocks	0.577	[27.12]	0.296	[22.05]	0.23	1421
			Weekly return comovement			
Nikkei only	0.773	[11.58]	0.180	[3.98]	0.15	218
All stocks	0.709	[17.92]	0.217	[8.50]	0.10	1450
			Monthly return comovement			
Nikkei only	0.694	[14.86]	0.237	[6.90]	0.19	212
All stocks	0.897	[10.94]	0.135	[4.88]	0.03	1488

Average parameter estimates and Fama and MacBeth (1973) t-statistics from rolling nonoverlapping cross-sectional regressions of comovement of daily stock returns on index overweighting at the start of the period

$$\beta_{EWN225,it} = a_t + b_t \cdot OW_{it-1} + u_{it}$$

β_{EWN225} denotes the slope parameter from regressions of daily, weekly, and monthly stock returns on the equal-weighted return of stocks that were in the Nikkei 225 index for the entire sample period. First-stage regression estimates for period t are computed using returns from the interval $[t-99, t]$ and are estimated for each stock for each 100-day interval. The table shows average coefficients from second-stage cross-sectional regressions of β_{EWN225} on overweighting. Overweighting is the natural log of one plus the ratio of index weight to the weight the stock would have taken in a value-weighted index. The table also indicates the average number of firms in each cross-section (N), as well as the time series average of R^2 from the cross-sectional regressions.

positively related to contemporaneous measures of comovement, it should be negatively related to measures of comovement reversal (*Hypothesis 3*).

To test *Hypothesis 3*, I estimate the cross-sectional relationship between overweighting and the covariance of stock returns with the lagged equal-weighted Nikkei return. In the first stage, I estimate conditional covariance of each stock return with the lagged equal-weighted index return:

$$R_{it} = \alpha_i + \beta_{EWN225,it} \cdot R_{EWN225,t} + \beta_{EWN225,it-1} \cdot R_{EWN225,t-1} + \varepsilon_{it} \quad (17)$$

using 100 days of prior returns data. As before, these regressions are estimated for each security in the panel, provided it has complete returns data over the estimation period, and provided its index status does not change during the estimation period.

Table 7 reports average coefficients from these regressions. For the full sample, average index beta is close to one, as expected. The covariance with lagged index returns is negative, on average, but the magnitude is small. When one sorts stocks according to the overweighting in the index, however, a clear pattern emerges. Overweighted stocks tend to have higher contemporaneous betas (1.18 compared to 0.92), but also exhibit more of a tendency to revert lagged index returns (coefficient on lagged returns of −0.066 compared to 0.05 for underweighted stocks). A test of means comparing $\beta_{EWN225,it-1}$ for under- and overweighted stocks rejects the hypothesis that these stocks have identical covariances with lagged index returns (unreported). Put differently, overweighted stocks tend to overreact to index returns, while underweighted stocks tend to underreact.

It is worth commenting on what the estimates in Table 7 say about the economic magnitude of excess comovement. The spread between the betas on lagged daily returns between the two groups is 0.12, slightly less than half the spread between the betas on contemporaneous daily returns of 0.26. Therefore, a large fraction of the excess comovement occurring at the daily level is reverted by the close of the next day, with the remainder occurring over longer horizons.

Although differences in the means of $\beta_{EWN225,it-1}$ between overweighted and underweighted securities are suggestive, one can apply a more formal cross-sectional test to these measures of reversion. Specifically, *Hypothesis 3* predicts a positive relation between the degree of reversion of index returns and lagged overweighting,

$$\beta_{EWN225,it-1} = a_t + b_t \cdot OW_{it-1} + u_{it} \quad (18)$$

The bottom panel of Table 6 shows results from regressions of $\beta_{EWN225,it}$ and $\beta_{EWN225,it-1}$ on overweighting. Consistent with the basic predictions, the panel shows a strong positive relation between contemporaneous beta and overweighting, and a strong negative relation between beta on lagged returns and overweighting. The strength of the cross-sectional relation is

Table 7
Reversion of excess comovement

Panel A: Sample averages

	$\beta_{EWN225,it}$	[t-stat]	$\beta_{EWN225,it-1}$	[t-stat]
Full sample	1.05	[105.12]	−0.008	[−2.26]
OW< Median(OW)	0.92	[74.63]	0.05	[6.99]
OW≥Median(OW)	1.18	[54.76]	−0.066	[−5.42]

Panel B: Cross-sectional relationships

	a	[t-stat]	b	[t-stat]
$\beta_{EWN225,it}$	0.72	[22.88]	0.22	[10.08]
$\beta_{EWN225,it-1}$	0.13	[6.39]	−0.09	[−6.47]

In the first stage, I estimate the contemporaneous comovement of the stock return with the equal-weighted return of the stocks in the Nikkei 225 index and the tendency to revert this comovement the next day

$$R_{it} = \alpha_i + \beta_{EWN225,it} \cdot R_{EWN225,t} + \beta_{EWN225,it-1} \cdot R_{EWN225,t-1} + \varepsilon_{it}$$

These regressions are estimated every 100 days for every stock in the sample. The first line of Panel A reports the time series average of the cross-sectional mean of contemporaneous and lagged betas from this regression. The second and third lines of Panel A report these averages sorted by the overweighting of each stock in the Nikkei index. Overweighting is defined as the natural log of one plus the ratio of index weight to the weight the stock would have taken in a value-weighted index. Stocks are sorted in each cross-section as being above or below the median overweighting. Panel B reports average coefficients from cross-sectional regressions of these beta coefficients on overweighting at the start of the period.

$$\beta_{EWN225,it} = a_t + b_t \cdot OW_{it-1} + u_{it}$$
$$\beta_{EWN225,it-1} = a_t + b_t \cdot OW_{it-1} + u_{it}$$

These regressions are estimated on each of the 100-day nonoverlapping cross-sections between 1993 and 2003. Fama–MacBeth t-statistics are reported in brackets.

apparent in Figure 4, where I plot $\beta_{EWN225,it}$, measured at the midpoint of the sample, against overweighting 100-days earlier. The most overweighted stocks revert as much as 30% of the index return the next day.

The results in Table 7 and Figure 4 imply that a trading strategy based on the reversion of excess comovement will generate positive abnormal returns. The intuition behind the strategy is to short overweighted stocks following increases in the Nikkei, and to go long overweighted stocks following declines in the Nikkei. One can implement this strategy in a variety of ways, but I form a simple zero investment portfolio that only exploits differences in comovement *between* index stocks. These objectives are met by the following set of portfolio weights:

$$w_{it} = -\left[OW_{it-1} - \frac{1}{N} \sum_{N} OW_{jt-1} \right] \cdot R_{EWN225,t-1} \tag{19}$$

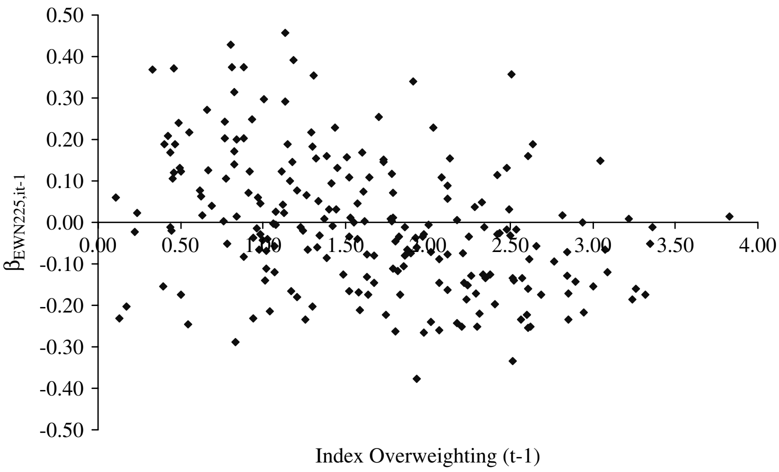

Figure 4
Reversion of excess comovement
A measure of the reversion of excess comovement with the Nikkei 225 stock index is plotted against lagged overweighting. The figure plots the slope coefficients on the lagged index return, $\beta_{EWN225,it-1}$, from the multivariate regression

$$R_{it} = \alpha_i + \beta_{EWN225,it} \cdot R_{EWN225,t} + \beta_{EWN225,it-1} \cdot R_{EWN225,t-1} + \varepsilon_{it}$$

against index overweighting, measured one day before the start of the sample used to estimate coefficients. Index overweighting is defined as the log of one plus the ratio of a stock's weight in the Nikkei 225 to the weight of the stock in the value-weighted TOPIX index. Regression slope parameters are estimated using 100 days of data ending at the midpoint of the panel, on 1 September 1998.

This portfolio takes a somewhat conservative approach because it relies on reversion in the cross-section, thereby ignoring the possibility that all index stocks comove excessively. Nevertheless, the portfolio produces profits in every calendar year in the sample and has an annualized Sharpe ratio of 2.33 between 1993 and 2003. In contrast, a portfolio formed in similar fashion that sorts Nikkei 225 stocks based on size (rather than overweighting) achieves a Sharpe ratio of only 0.83.

Figure 5 plots cumulative profits from the trading strategy described by (19). Note that it is difficult to trace these profits to any systematic factor exposure, as the portfolio frequently reverses the sign of its positions on individual stocks. Not surprisingly, trading profits are highest when market volatility is high, as this creates more opportunity to exploit overreaction of high-priced stocks.

An interesting feature of Figure 5 is that there are clear spikes in trading profits around October 1997 and October 1998. In fact, it is straightforward to connect these to market liquidity events: August 1998 is the Russian debt default, and Fleming (2003) shows that price impact was low in the Treasury market in October 1997.

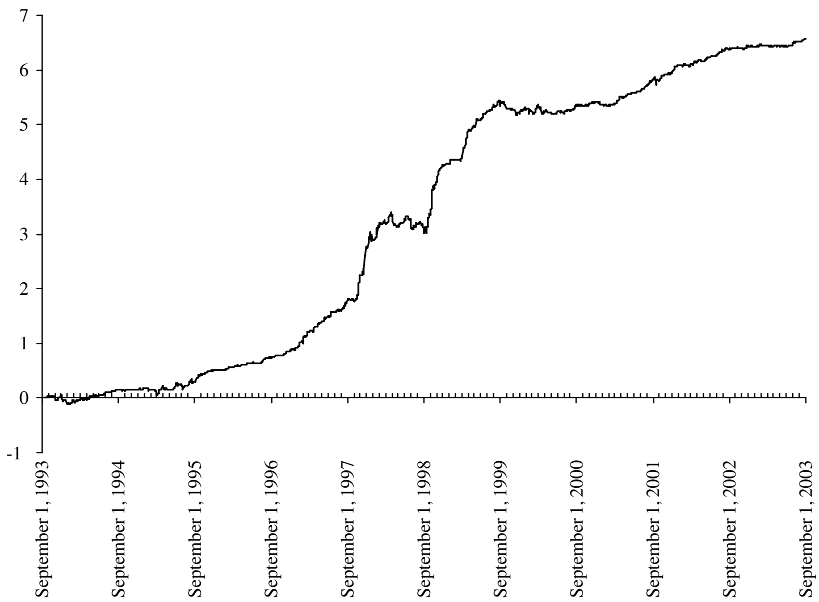

Figure 5
Cumulative profits from reversion of excess comovement trading strategy
Cumulative profits from a trading strategy that goes short overweighted Nikkei 225 stocks following index increases (hedged by offsetting long positions in underweighted stocks) and goes long overweighted Nikkei 225 stocks following index declines (hedged by offsetting short positions in underweighted stocks). The weights of the zero investment portfolio are given by

$$w_{it} = -\left[OW_{it-1} - \frac{1}{N}\sum_{N} OW_{jt-1}\right] \cdot R_{EWN225,t-1}$$

where OW_{it-1} denotes index overweighting at the end of the previous day, and $R_{EWN225,t-1}$ denotes the lagged return on the equal-weighted portfolio of stocks in the Nikkei 225 stock index. Overweighting is defined as the log of one plus the ratio of Nikkei 225 index weight to the weight the stock has in the value weighted TOPIX index. Panel A shows cumulative profits from this strategy, in yen. The annualized Sharpe ratio corresponding to the figure is 2.56.

The trading strategy bets that smaller members of the Nikkei 225 overreact to the index return. To be sure, the profitability of the strategy is evidence of mispricing. However, I verify that the profitability is evidence of correction of excess comovement, as another plausible story says that all small stocks (not just those in the Nikkei 225) overreact to Nikkei returns. This is easily checked by looking at the profits of a portfolio that goes short in small *nonindex* stocks following index increases, balanced by offsetting long positions in large *nonindex* stocks. Not surprisingly, this portfolio has negative profits, consistent with the idea that small stocks tend to lag the market in general, but overreact to the market when they are overweighted in the Nikkei 225.

To summarize, the excess comovement of stock returns documented in Section 3 implies predictability of returns at longer horizons. This implication receives strong support in the data.

5. Conclusions

In the presence of limits to arbitrage, periodic investor demand shocks that vary in degree across securities should have cross-sectional effects on the comovement of their returns. This article exploits the unique weighting system of the Nikkei 225 index to relate cross-sectional regularities in demand shocks to the comovement of stock returns. Using index overweighting as an instrument for the true variation in demand for Nikkei 225 stocks, I trace an empirical relation between overweighting and the comovement of a stock with other stocks in the index. The results suggest that overweighting accounts for a large fraction of the cross-sectional variation in comovement among index stocks, and comovement between index and nonindex stocks. By taking a cross-sectional approach, I avoid many of the endogeneity problems associated with single event studies.

The results seem most consistent with a demand theory of excess comovement, in which correlated investor demand across securities causes correlated mispricings. Supporting this explanation, a trading strategy based on the reversion of comovement over intermediate horizons generates economic profits. This is not predicted by friction-based explanations of excess comovement, in which index membership increases the speed at which stocks incorporate new common information. The link between excess comovement and predictability is novel relative to existing work on excess comovement, which focuses on contemporaneous correlations of returns. My results also explain why excess comovement is harder to detect at longer horizons.

More broadly, the magnitude of the results suggests that members of other arbitrarily weighted stock indexes may be subject to some degree of mispricing. Under these circumstances, it is not hard to understand the declining influence of the price weighted Dow Jones Industrial Average (DJIA) relative to the value weighted S&P 500, or the growing influence of the TOPIX relative to the Nikkei 225.[11] My results can justify why several sets of broadly used stock indices have moved toward "float adjusted" weightings, in which index weight is based on the tradable capitalization of each stock.[12] In Japan, even the value-weighted TOPIX recently announced a move toward free-float adjusted indices, citing a

[11] See Salomon, 1994 and Maeda, 2000.

[12] Recent conversions to float-weighted indices include the MSCI global indices, the FTSE (United Kingdom), certain Standard and Poor's indices, STOXX (Global), and SENSEXX (India).

desire to "avoid supply and demand distortion of share prices" arising in the trading of index shares from "passive funds."[13]

References

Ahn, D. H., J. Boudoukh, M. P. Richardson, and R. F. Whitelaw. 2002. Partial Adjustment or Stale Prices? Implications from Stock Index and Futures Return Autocorrelations. *Review of Financial Studies* 15:566–68.

Banz, R. 1981. The Relationship Between Return and Market Value of Common Stocks. *Journal of Financial Economics* 9:3–18.

Barberis, N., and A. Shleifer. 2003. Style Investing. *Journal of Financial Economics* 68:161–99.

Barberis, N., A. Shleifer, and J. Wurgler. 2005. Comovement. *Journal of Financial Economics* 75:283–317.

Bekaert, G., C. Harvey, and A. Ng. 2005. Market Integration and Contagion. *Journal of Business* 75:39–69.

Bekaert, G., R. C. Hodrick, and X. Zhang. 2005. International Stock Return Comovements, NBER Working Paper 11906.

Boyer, B. H. 2006. Comovement Among Stocks with Similar Book to Market Ratios. Working Paper, Brigham Young University.

Corsetti, G., M. Pericoli, and M. Sbracia. 2002. Some Contagion, Some Interdependence: More Pitfalls in Tests of Financial Contagion, CEPR Discussion Paper No. 3310.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann. 1990. Noise Trader Risk in Financial Markets. *Journal of Political Economy* 98:703–38.

Denis, D. K., J. J. McConnell, A. Ovtchinnikov, and Y. Yu. 2003. S&P 500 Index Additions and Earnings Expectations. *Journal of Finance* 58:1821–840.

Fama, E. F., and J. D. MacBeth. 1973. Risk, Return and Equilibrium: Empirical Tests. *Journal of Political Economy* 81:607–36.

Fama, E., and K. French. 1992. The Cross-Section of Expected Stock Returns. *Journal of Finance* 46:427–66.

Fisher, L. 1966. Some New Stock-Market Indexes. *Journal of Business* 39:191–225.

Fleming, M. J. 2003. Measuring Treasury Market Liquidity. *Federal Reserve Bank Economic Policy Review* 9(3):83–108.

Forbes, K., and R. Rigobon. 2002. No Contagion, Only Interdependence Measuring Stock Market Co-Movements. *Journal of Finance* 56:2223–261.

Froot, K. A., and E. Dabora. 1999. How are Stock Prices Affected by the Location of Trade? *Journal of Financial Economics* 53:189–216.

Greenwood, R. 2005. Short and Long Term Demand Curves for Stocks: Theory and Evidence on the Dynamics of Arbitrage. *Journal of Financial Economics* 75:607–49.

Greenwood, R., and N. Sosner. 2007. Trading Patterns and Excess Comovement of Stock Returns. *Financial Analysts Journal*, forthcoming.

Hardouvelis, G. A., R. La Porta, and T. A. Wizman. 1994. What Moves the Discount on Country Equity Funds? in *The Internationalization of Equity Markets*, eds., J. Frankel. Chicago: The University of Chicago Press.

[13] http://www.tse.or.jp/english/news/2004/200407/040723_a.html, 2004/7/23, Press release

Hecht, P. 2002. Do Equity Covariances Reflect Financial Leverage? Working Paper, Harvard Business School.

Hong, H., and J. C. Stein. 1999. A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets. *Journal of Finance* 54:2143–184.

King, M. A., and S. Wadhwani. 1990. Transmission of Volatility Between Stock Markets. *Review of Financial Studies* 3:5–33.

Kodres, L. E., and M. Pritsker. 2002. A Rational Expectations Model of Financial Contagion. *Journal of Finance* 57:769–99.

Kumar, A., and C. Lee. 2005. Retail Investor Sentiment and Return Comovements. Working Paper.

Kyle, A. S., and W. Xiong. 2001. Contagion as a Wealth Effect. *Journal of Finance* 56:1401–440.

Lee, C., A. Shleifer, and R. Thaler. 1991. Investor Sentiment and the Closed-End Fund Puzzle. *Journal of Finance* 46:75–110.

Lo, A. W., and C. A. MacKinlay. 1990. An Econometric Analysis of Nonsynchronous Trading. *Journal of Econometrics* 45:181–212.

Maeda, R. 2000. New Nikkei Reflects Japan Better, But Flaws Remain, Reuters News Service, April 25.

Modigliani, F., and M. Miller. 1958. The Cost of Capital, Corporation Finance, and the Theory of Investment. *American Economic Review* 48:261–97.

Mondria, J. 2006. Financial Contagion and Attention Allocation. University of Toronto Working Paper.

Morck, R., B. Yeung, and W. Yu. 2000. The Information Content of Stock Markets: Why do Emerging Markets have Synchronous Stock Price Movements? *Journal of Financial Economics* 58:215–60.

Pasquariello, P. 2007. Imperfect Competition, Information Heterogeneity, and Financial Contagion. *Review of Financial Studies*, 20:391–426.

Pavlova, A., and R. Rigobon. 2005. Wealth Transfers, Contagion, and Portfolio Constraints. NBER Working Paper 11440.

Pindyck, R. S., and J. J. Rotemberg. 1993. The Comovement of Stock Prices. *Quarterly Journal of Economics* 108:1073–104.

Salomon, R. S. Jr. 1994. Inaccurate Yardstick: The Dow Jones Industrial Average, Forbes Magazine.

Scholes, M., and J. Williams. 1977. Estimating Betas from Nonsynchronous Data. *Journal of Financial Economics* 5:309–28.

Vijh, A. M. 1994. S&P500 Trading Strategies and Stock Betas. *Review of Financial Studies* 7:215–51.

Yuan, K. 2005. Asymmetric Price Movements and Borrowing Constraints: A REE Model of Crisis, Contagion, and Confusion. *Journal of Finance* 60:379–411.