Available online at www.sciencedirect.com
ScienceDirect
Journal of Financial Economics 83 (2007) 367–396
JOURNAL OF Financial ECONOMICS
www.elsevier.com/locate/jfec

Do industries lead stock markets?☆

Harrison Hong[a,*], Walter Torous[b], Rossen Valkanov[b]

[a] *Princeton University, 26 Prospect Avenue, Princeton, NJ, 08540-5296, USA*
[b] *UCLA, 110 Westwood Plaza, Los Angeles, CA, 90095-1481, USA*

Received 16 June 2004; received in revised form 28 July 2005; accepted 9 September 2005
Available online 22 September 2006

Abstract

We investigate whether the returns of industry portfolios predict stock market movements. In the US, a significant number of industry returns, including retail, services, commercial real estate, metal, and petroleum, forecast the stock market by up to two months. Moreover, the propensity of an industry to predict the market is correlated with its propensity to forecast various indicators of economic activity. The eight largest non-US stock markets show remarkably similar patterns. These findings suggest that stock markets react with a delay to information contained in industry returns about their fundamentals and that information diffuses only gradually across markets.

JEL classifications: G12; G14; G15; E44

Keywords: Asset pricing; Information and market efficiency; Financial markets and macroeconomy; International financial markets

☆We are grateful to Jeremy Stein and a referee for many insightful comments. We also thank John Campbell, Kent Daniel, Ken French, Owen Lamont, Toby Moskowitz, Sheridan Titman, Rob Engle, Matthew Richardson, David Hirshleifer, Matthew Slaughter, Will Goetzmann, Mark Grinblatt, and participants at the 2004 AFA meetings, Berkeley-MIT-Texas Real Estate Research Conference, Goldman Sachs, Columbia, Dartmouth, Harvard Business School, NYU, Rice, and Yale for helpful comments. Hong acknowledges support from an NSF grant. Valkanov acknowledges support from the Anderson School at UCLA.

*Corresponding author.
E-mail address: hhong@princeton.edu (H. Hong).

0304-405X/$ - see front matter
doi:10.1016/j.jfineco.2005.09.010

1. Introduction

In this paper, we investigate whether the returns of industry portfolios are able to predict the movements of stock markets. We begin our analysis with the US stock market. Over the period 1946–2002, we find that 14 out of 34 industries, including commercial real estate, petroleum, metal, retail, financial, and services, can predict market movements by one month. A number of others such as petroleum, metal, and financial can forecast the market even two months ahead. After adding a variety of well-known proxies for risk and liquidity in our regressions as well as lagged market returns, the predictability of the market by these 14 industry portfolios remains statistically significant. We have also done numerical simulations to gauge just how many industries will (with statistical significance) forecast the market simply by chance, and in only 0.04% of the simulations are 14 or more industries able to forecast the market and on average, only five (in contrast to the 14 we find) are able to do so at the 10% level of significance.

In addition, we provide a few metrics regarding the statistical and economic significance of the documented predictability. First, we examine the ability of these industries to predict the market in comparison to well-known predictors such as inflation, default spread, and dividend yield and find comparable forecasting power. Second, we show that a portfolio incorporating information in past industry returns can lead under certain circumstances to a higher Sharpe ratio than simply holding the market. And third, we extend our analysis to each of the largest eight stock markets outside of the US, including Japan, Canada, Australia, UK, Netherlands, Switzerland, France, and Germany. In contrast to the US, these time series are limited to the period of 1973–2002 and we are unable to obtain the same set of controls (e.g., market dividend yield, default spread). With these caveats in mind, we find that the US results hold up remarkably well for the rest of the world.

Our investigation is motivated by recent theories that explore the implications of limited information-processing capacity for asset prices. Many economists have recognized for some time that investors, rather than possessing unlimited processing capacity, are better characterized as being only boundedly rational (see Shiller, 2000; Sims, 2001). Even from casual observation, few traders can pay attention to all sources of information, much less understand their impact on the prices of the assets that they trade. Indeed, a large literature in psychology documents the extent to which even attention is a precious cognitive resource (see, e.g., Kahneman, 1973; Nisbett and Ross, 1980).

More specifically, our investigation builds on recent work by Merton (1987) and Hong and Stein (1999). Merton develops a static model of multiple stocks in which investors have information about only a limited number of stocks and trade only those that they have information about. As a result, stocks that are less recognized by investors have a smaller investor base (neglected stocks) and trade at a greater discount because of limited risk-sharing. Hong and Stein develop a dynamic model of a single asset in which information gradually diffuses across the investment public and investors are unable to perform the rational expectations trick of extracting information from prices. As a result, the price underreacts to the information and there is stock return predictability.[1]

The hypothesis that guides our analysis of whether industries lead stock markets is that the gradual diffusion of information across asset markets leads to cross-asset return

[1]For other models of limited market participation, see Brennan (1975) and Allen and Gale (1994). For related models of limited attention, see, e.g., Peng and Xiong (2002) and Hirshleifer, Lim, and Teoh (2002).

predictability. The basic idea is that certain investors, such as those that specialize in trading the broad market index, receive information originating from particular industries such as commercial real estate or commodities like metals only with a lag. As a result, the returns of industry portfolios that are informative about macroeconomic fundamentals will lead the aggregate market.

This hypothesis relies on two key assumptions. The first is that valuable information that originates in one asset market reaches investors in other markets only with a lag, i.e., news travels slowly across markets. The second assumption is that because of limited information-processing capacity, many (though not necessarily all) investors might not pay attention to or be able to extract the information from the asset prices of markets that they do not specialize in. These two assumptions taken together lead to cross-asset return predictability.

Our hypothesis appears to be a plausible one for a few reasons. To begin with, even among equity money managers, there is specialization along industries such as sector or market-timing funds. Investors have their hands full trying to understand the markets that they specialize in. As a result, they are unable to devote the attention needed to process potentially valuable information from other markets in a timely manner. Moreover, as pointed out by Merton (1987) and the subsequent literature on segmented markets and limited market participation, few investors trade all assets. Put another way, limited participation is a pervasive feature of financial markets and could be another rationale for why investors in one market are slow to adjust to information emanating from another. Individual investors also participate in a limited number of markets as they hold very undiversified portfolios (see Blume and Friend, 1978; King and Leape, 1984).

We recognize that this gradual-information-diffusion hypothesis might not be the only explanation for our findings. However, it does provide a key auxiliary prediction that we can take to the data—namely, that the ability of an industry to predict the market ought to be strongly correlated with its propensity to forecast market fundamentals such as industrial production growth or other measures of economic activity. We test this prediction by first using individual industry returns to separately forecast industrial production growth and the growth rate of the Stock and Watson (1989) coincident index of economic activity. Many of the same sectors that lead the market can also forecast these two proxies of market fundamentals. Indeed, industry returns that are positively (negatively) cross-serially correlated with the market are also positively (negatively) cross-serially correlated with future economic activity. For instance, high returns for some industries like retail mean good news for future economic activity and the market, while high returns for other industries such as petroleum mean just the opposite.

Beyond the standard statistical inference techniques, we have also performed various numerical simulation exercises to rule out that such a relation is due purely to chance. This finding strongly supports our hypothesis that the documented cross-predictability is due to the market reacting with a delay to information contained in industry returns about its fundamentals. It also distinguishes our gradual-information-diffusion hypothesis from other behavioral explanations of stock return predictability (see Section 2.3). Importantly, we have also extended this analysis to the eight largest stock markets outside of the US. The key relation, that the propensity of industries to forecast the market is correlated with their propensity to forecast industrial production growth, is also present in seven of the eight countries. The only country in which this relation does not emerge is Japan.

Our paper is related to the literature on lead-lag relations among stocks, epitomized by the finding of Lo and MacKinlay (1990) that large stocks lead small stocks. A number of other papers followed trying to rationalize this finding (see, e.g., Brennan, Jegadeesh, and Swaminathan, 1993; Badrinath, Kale, and Noe, 1995; Jegadeesh and Titman, 1995). These studies typically find that stocks that are in some sense more liquid (e.g., have more analysts following or have institutional ownership) tend to lead less liquid stocks. However, Boudoukh, Richardson, and Whitelaw (1994) argue that many of these lead-lag relations are due to the own-autocorrelation of portfolios and a high contemporaneous correlation among portfolios. In other words, a large stock portfolio does not significantly lead a small stock portfolio once the lagged returns of the small stock portfolio are included in a multiple regression.

Our findings differ from those in the above papers in a few respects. First, we focus on predicting the aggregate stock market, whereas the papers in this literature have focused on stocks of different characteristics leading or lagging one another. We are able to link this predictability to information about fundamentals such as industrial production growth. Second, in contrast to the lead-lag relations between small stock and large stock portfolios, our findings do not have an obvious liquidity interpretation since there is not an obvious difference in liquidity between stocks in the typical industry portfolio (e.g., oil) and stocks in the value-weighted market portfolio. Third, our findings are not due to the own-autocorrelation of portfolios since we control for lagged market returns in our predictive regressions and each industry portfolio constitutes only a small fraction of the market portfolio.[2]

Our paper proceeds as follows. In Section 2, we develop a simple model to make clear the assumptions behind our hypothesis and generate some testable predictions. We describe the data in Section 3. We present our empirical findings for the United States in Section 4. We extend our analysis to the eight largest stock markets outside of the US in Section 5. We conclude in Section 6.

2. Model

2.1. Basic setup

Our model considers the pricing of two assets (stocks) in a three-date economy, $t = 0, 1, 2$. We assume for simplicity that the risk-free rate is zero. The two assets, X and Y, have terminal values at $t = 2$ given by D_X and D_Y, which are jointly normal with means of zero and variances of $\sigma^2{}_{X,D}$ and $\sigma^2{}_{Y,D}$ and covariance $\sigma_{XY,D}$.

Investors participate in either market X or market Y. This limited market participation assumption can be motivated by exogenous reasons such as taxes or regulations. Alternatively, we can motivate it by introducing a fixed cost of participation in each market so that investors will only want to participate in one. As long as the fixed costs of participating in X and Y are not too different, there will be some investors participating in each market. However, this assumption is not crucial to our results. Investors may

[2]A few other papers have also confirmed that some industry portfolios lead the stock market (see Eleswarapu and Tiwari, 1996; Pollet, 2002). Our work is also related to the recent work on stock price momentum (see, e.g., Jegadeesh and Titman, 1993; Hong, Lim, and Stein, 2000; Grinblatt and Moskowitz, 1999). Our focus is on whether and why industries lead the aggregate market index.

participate in both markets as long as there are some investors who are slow to optimally combine signals from both markets in forming expectations.

At $t = 1$, investors in market X receive signal $S_X = D_X + \varepsilon_{X,S}$ about the terminal value of X, investors in market Y receive signal $S_Y = D_Y + \varepsilon_{Y,S}$ about the terminal value of Y, and these signals are known to all participants at $t = 2$. This is our gradual-information-diffusion assumption. The noise in the signals $\varepsilon_{X,S}$ and $\varepsilon_{Y,S}$ is normally distributed with means of zero and variances of $\sigma^2{}_{X,S}$ and $\sigma^2{}_{Y,S}$, respectively. We assume that $\varepsilon_{X,S}$ and $\varepsilon_{Y,S}$ are independent of each other and of all other shocks in the economy. The supply of assets is assumed to be Q_X and Q_Y shares outstanding for assets X and Y, respectively.

Investors in asset X cannot process information pertaining to asset Y, and vice versa—this is our limited information-processing capacity assumption. This assumption is a simple way of capturing the idea that investors, due to limited cognitive capabilities, have a hard time processing information from asset markets that they do not participate in. This could be because information from other markets is less salient. Alternatively, investors could be too busy trying to figure out the market that they are in to process this information in a timely fashion.

We assume that investors have CARA preferences with a risk aversion coefficient of a. Given the price function $P_{k,t}$, the investor in asset market k ($k = X, Y$) solves the following optimization problem:

$$\begin{aligned} &\operatorname{Max} \mathrm{E}_{k,0}[-\exp(-aW_{k,2})], \quad k = X, Y, \\ &\{\theta_k\} \\ &\text{s.t. } W_{k,t} = W_{k,t-1} + \theta_{k,t-1}(P_{k,t} - P_{k,t-1}), \end{aligned} \tag{1}$$

where $W_{k,t}$ and $\theta_{k,t,}$ are the wealth and share holdings of a representative investor in asset market k at time t (we do not index different investors in the same asset market for simplicity) and $P_{k,2} = D_k$. The solution to this problem and the equilibrium prices are obtained using standard techniques.

The equilibrium price in market k is given by

$$P_{k,t} = \mathrm{E}_{k,t}[D_k] - b_{k,t}Q_k, \quad k = X, Y, \tag{2}$$

where $\mathrm{E}_{k,t}$ $[D_k]$ is the conditional expectation of the terminal payoff of asset k at time t, $b_{k,t}>0$ is the standard risk discount at time t, and Q_k is the supply of the asset.

2.2. Serial and cross-serial correlations

Given the equilibrium prices described in Eq. (2), it is straightforward to calculate the serial and cross-serial correlations for assets X and Y. Let $R_{k,t} = P_{k,z} - P_{k,t-1}$ be the date t return for asset k. The two propositions that follow are self-explanatory and are given without proof.

Proposition 1. *The own serial return correlations are zero, i.e., $Corr(R_{k,2},\ R_{k,1}) = 0$ for $k = X, Y$. The cross-serial return correlations, $Corr(R_{Y,2},\ R_{X,1})$ and $Corr(R_{X,2},\ R_{Y,1})$, are non-zero and can be positive or negative depending on the sign of the covariance of asset payoffs, $\sigma_{XY,D}$.*

Intuitively, investors in market k rationally condition on all information associated with market k. As a result, the price is efficient with respect to own asset information. Hence the own serial correlation is zero. However, investors in asset market Y ignore or cannot process the information from X, including past returns. As a result, the time-1 return in market X predicts the time-2 return in market Y. If investors in market Y (X) condition on the time-1 return in market X (Y), then the cross-serial correlations would be zero.

Moreover, the results in Proposition 1 would remain even if we enriched the model to include the following sets of traders. First, even if some fraction of the investors in each market pays attention to and can process information from the other market, there will still be cross-predictability, though it will be smaller in magnitude. Second, if there are limits to arbitrage (Shleifer and Vishny, 1997), then cross-predictability will remain in equilibrium even if there are arbitrageurs who try to profit from the cross-asset return predictability. We state this more formally in Proposition 2.

Proposition 2. *Even if there are arbitrageurs who trade in both markets to exploit the cross-predictability, as long as there are limits to arbitrage, some cross-predictability will remain in equilibrium.*

While our model is designed to generate positive cross-serial correlations when own serial correlations are zero, it is important to note that the model can be easily augmented to simultaneously generate own and cross-serial correlations. If we additionally assume that some investors in asset k do not pay attention to or cannot process S_k, then along with cross-serial correlation, there is positive serial correlation, i.e., $Corr(R_{k,2},\ R_{k,1})>0$ for $k = X, Y$. Intuitively, if investors in the same market pay attention to (or wake up to) information at different points in time, then information gradually diffuses across investors in the same market (Hong and Stein, 1999), resulting in positive serial correlation as well as non-zero cross-serial correlation in asset returns.

2.3. Testable predictions

In our empirical work, we test two specific predictions that are implied by our model. In the context of our model, think of the broad market index as asset Y and an industry portfolio that is informative about market fundamentals as asset X. Proposition 1 implies the following prediction.

Prediction 1. *The broad market index can be predicted by the returns of industry portfolios, controlling for lagged market returns and well-known predictors such as inflation, default spread, and dividend yield.*

Note that our model only implies that an industry will lead the market if it has information about market fundamentals. In other words, an industry with little information about economic activity will not forecast the market whether or not investors are paying attention to it. Indeed, it follows from this logic that an industry's ability to predict the market is correlated with the information that it has about market fundamentals. As a result, we have Prediction 2.

Prediction 2. *The ability of an industry to forecast the market is related to its ability to forecast changes in market fundamentals such as industrial production growth or changes in other indicators of economic activity.*

This prediction also distinguishes our gradual-information-diffusion hypothesis from other behavioral explanations of stock return predictability due to biased inferences on the part of a representative investor (see, e.g., Daniel, Hirshleifer, and Subrahmanyam, 1998; Barberis, Shleifer, and Vishny, 1998). For instance, Daniel, Hirshfleifer, and

Subrahmanyam would attribute our cross-asset return predictability finding to continuing overreaction to industry returns on the part of overconfident investors trading the aggregate market index. Their model, however, is silent on why this cross-predictability is strongly related to the informativeness of an industry about market fundamentals.

By the same token, Barberis, Shleifer, and Vishny would argue that the documented cross-predictability is due to a conservatism bias, i.e., a representative investor updates on industry news about market fundamentals a bit slower than a Bayesian would. The degree to which prices underreact depends on how slowly the investor updates information. As such, their model would also not be able to generate this prediction unless there is an additional assumption that investors are slower to adjust to news from certain industries than others.

3. Data

Our data on industry portfolios from the US stock market come from two sources. From Ken French's website, we obtain monthly returns to 38 value-weighted industry portfolios for the years 1946–2002 (see Fama and French, 1997). We have to drop five of these industries from our analysis because they have missing observations.[3] Since commercial real estate is not a separate portfolio and is likely to provide a good setting to test our hypothesis, we augment this sample by constructing a real estate industry portfolio from an index of REIT returns obtained from the NAREIT website (www.nareit.com). We use the comprehensive, value-weighted REIT index of equity, mortgage, and hybrid REITs. The REIT data only go back to January 1972. So, counting real estate, we consider 34 industry portfolios in all.

In addition to these indices, we also utilize the following variables. We use the returns of the Center for Research in Security Prices (CRSP) value-weighted portfolio in excess of the risk-free rate (denoted by RM) as the proxy for the broad market index. Inflation (INF), measured as the growth rate of the Consumer Price Index, is obtained from the DRI database. Also obtained from the DRI database is the default spread (DSPR), defined as the difference between the yield of BAA-rated and AAA-rated bonds. The market dividend yield (MDY) is the one-year dividend from the CRSP value-weighted market portfolio divided by the current price. We also calculate a time series of monthly market volatility from daily returns to the CRSP value-weighted portfolio as in French, Schwert, and Stambaugh (1987), which is denoted by MVOL.

We will also use the following two macroeconomic variables. From the DRI database, we obtain a time series of the level of industrial production, which is available at a monthly frequency. From Mark Watson's web page, we obtain a time series of the Stock and Watson (1989) coincident index of economic activity, which is also available at a monthly frequency. Their experimental coincident index is a weighted average of four broad monthly measures of US economic activity: industrial production, real personal income, real manufacturing and trade sales, and total employee hours in non-agricultural establishments. We denote the monthly growth rates of industrial production and the Stock and Watson coincident index of economic activity by IPG and SWG, respectively.

Table 1 provides summary statistics of these variables. The means and standard deviations are in monthly percentage points. Panel A of Table 1 lists the 34 industry portfolios for the US

[3]The five industries that we exclude from our analysis are GARBAGE (sanitary services), STEAM (steam supply), WATER (irrigation systems), GOVT (public administration), and OTHER (everything else).

Table 1
Summary statistics

The table presents summary statistics of the variables of interest. In Panel A, the variables are the returns of the 34 US industry portfolios in excess of the risk-free rate. Panel B contains data about the US stock market and state of the economy. RM is the CRSP value-weighted market portfolio return in excess of the risk-free rate. INF is the CPI inflation rate. DSPR is the default spread between BAA-rated and AAA-rated bonds. MDY is the dividend yield of the market portfolio. MVOL is the market volatility computed from daily return data. IPG is industrial production growth and SWG is the growth rate of the Stock and Watson (1989) coincident index of economic activity. Panel B also contains data about the stock market return (RM) and industrial production growth (IPG) of the eight largest equity markets outside of the US, namely, the United Kingdom, Australia, Canada, France, Germany, Japan, the Netherlands, and Switzerland. In Panel C, we provide summary statistics for 35 industry returns for the eight stock markets outside of the US. Rather than listing the means and standard deviation for all countries and industries, we provide, for each industry, the mean industry return, its standard deviation, minimum, maximum, and number of observations across countries. For the US, all variables are from January 1946 to December 2002 with the exception of RLEST, which is from January 1972 to December 2002, and SWG, which is from March 1959 to December 2002. For the rest of the world, all variables are from January 1973 to December 2002. The data are at monthly frequency and in monthly percentage points.

Industry	Mean	Std. dev.	Industry	Mean	Std. dev.	Industry	Mean	Std. dev.
Panel A: Industry portfolio returns (US)								
RLEST	0.435	4.373	PAPER	0.699	5.341	CARS	0.628	5.406
AGRIC	0.433	7.239	PRINT	0.647	5.342	INSTR	0.704	5.367
MINES	0.416	6.259	CHEMS	0.660	4.593	MANUF	0.595	6.344
OIL	0.691	6.671	PTRLM	0.765	4.952	TRANS	0.504	5.728
STONE	0.843	7.654	RUBBR	0.610	6.019	PHONE	0.402	4.696
CNSTR	0.523	6.934	LETHR	0.691	6.232	TV	0.897	6.714
FOOD	0.669	4.247	GLASS	0.515	5.866	UTILS	0.471	3.868
SMOKE	0.872	5.664	METAL	0.434	6.109	WHLSL	0.619	5.542
TXTLS	0.488	5.934	MTLPR	0.587	4.876	RTAIL	0.634	5.120
APPRL	0.402	6.573	MACHN	0.616	5.821	MONEY	0.673	4.835
WOOD	0.661	7.268	ELCTR	0.650	6.236	SRVC	0.657	6.517
CHAIR	0.496	5.522						
Variable	Mean	Std. dev.	Variable	Mean	Std. dev.	Variable	Mean	Std. dev.
Panel B: Other variables (US and rest of the world)								
US			UK			Germany		
RM	0.567	4.120	RM	0.669	5.867	RM	0.381	5.258
INF	0.295	0.905	IPG	0.074	1.382	IPG	0.151	1.401
DSPR	0.075	0.035	Australia			Japan		
MDY	0.210	0.256	RM	0.643	6.169	RM	0.264	5.309
MVOL	3.546	1.803	IPG	0.187	1.265	IPG	0.113	2.706
IPG	0.303	1.180	Canada			Netherlands		
SWG	0.226	0.565	RM	0.550	4.663	RM	0.569	5.038

IPG	0.113	1.432	IPG	0.134	0.470
France			Switzerland		
RM	0.715	6.378	RM	0.468	4.832
IPG	0.095	1.742	IPG	0.218	0.868

Industry	Mean	Min	Max	Std. dev.	Nobs.	Definition
Panel C: Industry portfolio returns (rest of the world)						
AERSP	0.657	0.632	0.681	9.268	2	Aerospace and Defense.
AUTMB	0.378	0.000	0.805	8.217	8	Automobiles, Auto parts, Tires & Rubber, Vehicle distribution.
BANKS	0.594	0.087	0.866	7.148	8	Banks.
BEVES	0.419	−0.490	1.058	6.858	8	Beverages & Brewers & Distillers & Vintners Distillers & Vintners & Soft Drinks.
CHMCL	0.577	0.245	0.922	6.830	8	Chemicals, Commodity Chemicals, Specialty Chemicals, Advanced Chemicals.
CNSBM	0.423	0.058	0.809	6.568	8	Builders Merchants, Building & Building Materials, Construction Materials, Building, Other Construction.
DIVIN	0.418	0.039	0.742	7.472	6	Diversified, Industrials.
ELECT	0.289	−0.815	0.633	7.413	7	Electricity.
ELTNC	0.591	0.249	1.238	8.520	7	Electronic & Electrical Equipment.
ENGEN	0.201	−0.198	0.546	8.050	8	Engineering, Commercial Vehicles, Machinery Engineering, Contractors, Fabricators.
FDRET	0.812	−0.118	1.310	7.540	8	Food & Drug, Food & Drug Retailers.
FOODS	0.663	0.329	1.132	6.436	8	Food, Farming & Fishing Producers, Food Processors.
FSTPA	0.311	0.000	0.671	8.884	6	Forestry, Paper & Paper Processing.
HLTHC	0.897	0.260	1.520	7.877	8	Health, Health Maintenance, Organizations, Hospital Management, Medical Equipment, Medical Equipment & Supplies, Other Health Care.
HHOLD	−0.284	−6.239	1.423	10.013	8	Clothing & Footwear, Clothing & Footwear Goods, Furnishings & Floor, Consumer Electronics, Household Appliances, Leisure Equipment, Textiles & Leather.
INFOH	−0.056	−2.007	1.569	15.112	8	Information, Computer Hardware, Semiconductors, Hardware Telecom Equipment.
INSUR	0.744	0.231	1.194	8.402	8	Insurance, Brokers, Insurance Brokers, Insurance Non-life, Re-insurance, Other.
INVSC	−1.532	−12.702	0.886	10.887	8	Investment, Investment Companies, Investment Trust, International Investment. Trusts, Emerging Markets, Venture Investment Companies, Exchange Traded Funds.
LESUR	0.518	−2.096	2.429	10.731	8	Leisure, Gambling & Hotels, Leisure Facilities, Restaurants & Pubs.
LIFEA	0.145	−2.324	0.950	7.368	8	Life Assurance Companies.

Table 1 (*continued*)

Industry	Mean	Min	Max	Std. dev.	Nobs.	Definition
Panel C: Industry portfolio returns (rest of the world)						
MEDIA	0.421	−1.057	1.378	8.344	8	Media, Television, Radio, Entertainment, Filmed Entertainment, Entertainment Networks, Media Agencies, Photography, Publishing & Printing.
MNING	0.226	−0.841	0.701	8.952	4	Mining, Gold Mining, Mining Finance, Other Mineral, Extractors.
OILGS	0.588	0.178	0.860	7.853	6	Oil & Gas, Exploration, Production, Oil Services, Oil Integrated.
PERSH	0.718	0.405	1.215	7.044	5	Personal Care Household Products, Personal Products.
PHARM	0.251	−5.861	2.286	10.734	8	Pharmaceuticals, Biotechnology.
RLEST	0.350	−0.188	1.484	7.210	8	Real Estate Development, Development Property Agencies, REITs.
RTAIL	0.366	−0.033	0.858	7.465	8	Retailers, Discount & Super, Stores & Warehouses, Retailers E-commerce, Retailers, Hardline Retailers, Multi Department, Retailers-Soft Goods.
SFTCS	0.958	0.266	2.700	11.894	7	Software & Computer Services, Internet Services, Software.
SPFIN	0.474	−0.694	1.591	8.592	8	Speciality & Asset Managers, Asset Managers, Other Consumer Finance, Consumer Finance, Finance, Investment Banks, Mortgage Finance, Other Financial.
STLOM	−0.026	−0.767	0.526	10.044	7	Steel & Non-Ferrous Metals, Other Metals.
SUPSV	0.383	−0.714	0.979	7.627	8	Business Support, Business Support Services, Delivery Services, Education, Training, Environmental Control, Environmental Control, Transaction & Payroll Services, Security & Alarm Services.
TELCM	0.397	−0.272	1.331	9.866	8	Telecom, Telecom Fixed Line Services, Telecom Wireless.
TOBAC	0.619	−0.342	1.130	7.501	3	Tobacco.
TRNSP	0.285	−0.665	1.553	7.344	8	Transport, Airlines & Airports, Rail, Road & Freight, Shipping & Ports.
UTILO	0.521	0.198	0.902	6.643	6	Gas Distribution, Multi-Utilities, Water.

(by their abbreviated names) along with the means and standard deviations of their returns. All returns are in excess of the one-month T-bill rate. The acronyms of the industry portfolios are taken from Fama and French (1997). In most cases, the acronyms are self-explanatory. Precise definitions of these indices are available on Ken French's website. Notice that some of the industries are very related. For instance, OIL and PTRLM (petroleum) are treated as two different industries. The main difference between them is that OIL covers oil and gas extraction, while PTRLM covers petroleum refining and petroleum products. Two other industries that are also related are MINE and STONE, with the difference being that STONE covers non-metallic minerals except fuels. MTLPR or metal products is treated differently from METAL, which covers primary metal industries. Importantly, MONEY includes financial, insurance, and real estate stocks. However, real estate constitutes a minuscule part of MONEY. As such, we create a separate real estate portfolio (RLEST) by using the REIT index as a proxy.[4] Panel B lists the statistics for the remaining variables by country. We obtain from the Datastream database monthly market returns and monthly industrial production growth figures for the eight largest non-US stock markets from 1973 to 2002.

We also obtain from the Datastream database monthly industry returns for each of these eight markets for the period 1973–2002. The summary statistics for industry returns are summarized in Panel C, where the returns are calculated in each country's local currency. These returns are raw or unadjusted since data on risk-free rates for these countries are difficult to obtain. There are a total of 35 industries.[5]

The industry categorizations from the Datastream database are somewhat different from the Fama-French industry portfolios. While Datastream provides descriptions of their industry portfolios, it is not possible to convert the Fama-French industry portfolios into Datastream portfolios and vice versa. We could have featured the Datastream portfolios for the US stock market, but we decided on the Fama-French portfolios because they are more readily accessible and are available over a much longer period of time (Datastream data for the US also only go back to 1973). The results for the US stock market hold regardless of the database that we use.

4. Evidence from the US stock market

4.1. Predictive regressions involving industry and market returns

We begin by exploring the ability of industry returns to lead the market using US data. To see whether industries can forecast the market (Prediction 1), we estimate the following specification separately for each of the 34 portfolios:

$$\mathrm{RM}_t = \alpha_i + \lambda_i \; R_{i,t-1} + \boldsymbol{A_i Z_{t-1}} + e_{i,t}, \tag{3}$$

[4]Interestingly, we have replicated our findings using an alternative real estate industry portfolio from Ken French's website. This portfolio consists of small stocks such as realty companies and real estate brokers but excludes REITs, spanning 1970–2002. It is correlated with our real estate index but might not be as informative as RLEST since REITs are required to invest most of their resources in properties.

[5]A number of the stock markets outside of the US do not have companies in particular industries. Notice that a few of the industry portfolios have huge monthly means and standard deviations (e.g., INVSC in Japan has a mean of −12.7% and standard deviation of 22%). Some of this is due to industries having only a few stocks. We have experimented with a number of exercises to account for the effect of outliers on our international evidence below, such as winsorizing the returns or dropping industries with few companies. Our findings are robust to outliers.

where RM_t is the excess return of the market in month t, $R_{i,t-1}$ is the excess return of industry portfolio i lagged one month, and $\boldsymbol{Z_{t-1}}$ is a vector of additional market predictors. For each of these 34 time-series regressions, there are a total of 684 monthly observations.

We include a number of well-known market predictors in $\boldsymbol{Z_{t-1}}$ to address alternative explanations for why industry returns might forecast the market. Among them are the lagged excess market (RM_{t-1}), inflation (Fama and Schwert, 1977), the default spread (Fama and French, 1989), and the market dividend yield (Campbell and Shiller, 1988). These variables are typically thought to proxy for time-varying risk. To the extent that our results hold even with these predictors in the regressions, we conclude that our findings are not due to time-varying risk. Additionally, we worry that industry returns are forecasting market volatility, so we also include lagged market volatility in our set of control variables.

In an earlier draft (Hong, Torous, and Valkanov, 2002), we use a shorter sample beginning in 1973 and include as additional controls the term spread and changes in the Fed Funds rate. The term spread has been documented to predict the market (Fama and French, 1989). And since some industries such as financials could proxy for changes in the liquidity of the aggregate market or be especially sensitive to monetary policy variables, we also include lagged changes in the Fed Funds rate for good measure. The data for these two variables do not go back to 1946 and as a result, we are unable to include them in the current paper. However, these controls have little effect on our findings for the shorter sample and as such are not likely to change our results here.[6]

The coefficients of interest are the 34 λ_i's, which measure the ability of each of the industry portfolios to lead the market. Since many of these industries are likely to contain valuable information about market payoffs, we expect a significant number of these coefficients to be non-zero to the extent that our gradual-information-diffusion hypothesis holds.

Alternatively, rather than investigating whether different industries lead the market separately, we can augment specification (3) by simultaneously including all 34 industry returns. The cost of doing this is that the standard errors on our estimates will be larger since we only have a limited number of observations and so we cannot estimate the effect of each industry on future market returns very precisely. The benefit of doing this is that since industry returns are contemporaneously correlated, we worry about issues related to omitted variables. In other words, some of our results could be biased by not simultaneously including all other industries.

It turns out that our results are not significantly affected by whether we run the forecasting regressions separately or by pooling all the lagged industry returns. So, for the sake of precision, we present the results using specification (3). We discuss the results when we pool all the industries into one regression below.

We first present the results for the case of metal. This allows us to thoroughly describe all the regression specifications used in our analysis without reporting the results of all specifications for every industry. Much of the discussion for this case also applies to the other industries. In the following table and all subsequent tables, the standard errors are formulated to account for correlation of the residuals across the 34 industry returns at a point in time as well as for serial correlation. More specifically, we use the estimated residuals from the 34 regressions at a point in time to form a consistent estimate of their

[6]Similar results hold when we augment the specification to include the returns of the small stock minus the large stock portfolio and the high book-to-market minus the low book-to-market portfolio.

Table 2
Predictive regressions between metal and market portfolios (US)

This table presents the results from forecasting the market return in month t using variables at month $t-1$. METAL is the return on the primary metal industry portfolio. RM is the CRSP excess value-weighted market portfolio return. INF is the CPI inflation rate. DSPR is the default spread between BAA-rated and AAA-rated bonds. MDY is the dividend yield of the market portfolio. MVOL is the market volatility. In all columns, the least squares estimates, t-statistics (in parentheses), adjusted R^2, and number of observations are displayed. The standard errors (used in the computation of the t-statistics) are adjusted for cross-industry correlation in the error terms using an estimate of the covariance matrix computed with estimated residuals from the 34 regressions at a point in time. The standard errors also include a Newey–West serial correlation and heteroskedasticity correction with three monthly lags. The sample period is January 1946–December 2002. *Significant at 10% level. **Significant at 5% level.

	Dependent variable —RM (US)		
	(1)	(2)	(3)
CONST	0.005	−0.002	−0.007
	(3.074)**	(−0.451)	(−1.555)
METAL(−1)	−0.096	−0.077	−0.085
	(−2.261)**	(−2.107)**	(−2.308)**
RM(−1)	0.053	0.061	0.052
	(2.426)**	(1.752)*	(1.526)
INF(−1)		−0.578	−0.624
		(−3.292)**	(−3.549)**
DSPR(−1)		0.888	0.526
		(1.832)*	(1.069)
MDY(−1)		1.418	1.412
		(2.342)**	(2.211)**
MVOL(−1)			0.241
			(2.670)**
R^2	0.009	0.032	0.044
T	684	684	684

cross-industry covariance matrix, which is then used in the formulation of the standard errors. This approach is standard when dealing with systems of equations with possible cross-equation correlation in the residuals (Hayashi, 2000). It is similar to Vuolteenaho (2002) and Rogers (1993), with the only exception that we also include a Newey and West (1987) correction with three monthly lags for possible serial correlation in the residuals.

In Table 2, we report the results of various regressions that establish the predictive ability of the metal industry portfolio. In Column (1), we run a forecasting regression of market return on a constant, the lagged values of the metal portfolio, and RM. (We also experiment with adding in multiple lags of the market: past month, two months previous and three months previous; our results are unchanged). The coefficient on lagged metal is –0.096 and is statistically significant. Surprisingly, this coefficient is still statistically significant even after we control for other predictors such as INF, DSPR, and MDY in Column (2). In Column (3), we augment the specification in Column (2) by adding in lagged market volatility, but the coefficient of interest remains statistically significant. Across Columns (1)–(3), the usual market predictors such as INF, DSPR, and MDY have predictive power in this sample.[7]

We now see how many of these industries lead the market in Table 3. Our regression specification includes a constant, the lagged one-month industry return, lagged one-month market return, inflation, default spread, market dividend yield, and market volatility (all lagged by one month). Note that while we are running this regression separately for each industry, the standard errors are constructed to take into account cross-industry correlation in the residuals, as explained above. Rather than report the coefficient of each of the independent variables for every one of the 34 regressions, we report just the coefficient of the particular lagged one-month industry return along with the R^2 of the regression.

Table 3
Predictive regressions involving various industry and market portfolios

This table presents forecasts of the market return using various industry portfolio returns (separately) at various horizons: month t, month $t+1$, and month $t+2$. The other forecasting variables are lagged RM (the market return), INF (the CPI inflation rate), DSPR (the default spread between BAA-rated and AAA-rated bonds), MDY (the dividend yield of the market portfolio), and market volatility (MVOL). We only report the coefficients in front of the lagged industry return. The standard errors (used in the computation of the t-statistics) are adjusted for cross-industry correlation in the error terms using an estimate of the covariance matrix computed with estimated residuals from the 34 regressions at a point in time. The standard errors also include a Newey–West serial correlation and heteroskedasticity correction with three monthly lags. We also report Wald tests of the joint significance of the coefficients estimated in a GMM system (joint test of the null hypothesis that the lambdas are zero). The Wald tests (34 d.f.) are computed using Newey–West heteroskedasticity and serial correlation robust variance covariance matrix. The sample period is January 1946–December 2002, except for RLEST, which is from 1972 to 2002. *Significant at 10% level. **Significant at 5% level.

	Forecast of RM (US) using industry returns at various horizons (H)					
	Month t		Month $t+1$		Month $t+2$	
Industry	IND(−1)	R^2	IND(−1)	R^2	IND(−1)	R^2
RLEST	0.173	0.053	−0.029	0.021	0.009	0.008
	(2.643)**		(1.034)		(0.324)	
AGRIC	0.027	0.039	−0.010	0.024	−0.028	0.020
	(0.964)		(0.546)		(0.998)	
MINES	−0.067	0.044	−0.041	0.033	−0.040	0.031
	(−1.999)**		(1.708)*		(1.659)*	
OIL	−0.009	0.038	−0.021	0.020	−0.006	0.002
	(−0.259)		(0.932)		(0.435)	
STONE	−0.038	0.041	−0.039	0.032	0.004	0.001
	(−1.722)*		(1.669)*		(0.352)	
CNSTR	0.018	0.038	−0.020	0.024	−0.029	0.021
	(0.474)		(1.170)		(1.133)	
FOOD	0.011	0.038	0.039	0.028	−0.020	0.017
	(0.162)		(1.633)		(0.594)	
SMOKE	−0.025	0.038	−0.001	0.012	−0.002	0.001
	(−0.702)		(0.247)		(0.102)	
TXTLS	0.066	0.042	−0.005	0.009	−0.020	0.018
	(1.574)		(0.197)		(0.604)	
APPRL	0.093	0.042	−0.008	0.004	−0.018	0.016
	(1.996)**		(0.305)		(0.564)	
WOOD	−0.013	0.038	−0.014	0.010	−0.005	0.002
	(−0.369)		(0.982)		(0.242)	
CHAIR	0.030	0.038	0.001	0.006	−0.023	0.018

Table 3 (*continued*)

	Forecast of RM (US) using industry returns at various horizons (H)					
	Month t		Month $t+1$		Month $t+2$	
Industry	IND(−1)	R^2	IND(−1)	R^2	IND(−1)	R^2
	(0.656)		(0.157)		(0.734)	
PAPER	−0.024	0.038	−0.022	0.013	−0.050	0.036
	(−0.407)		(0.687)		(1.943)*	
PRINT	0.140	0.048	0.049	0.035	−0.031	0.022
	(2.455)**		(1.867)*		(1.232)	
CHEMS	−0.056	0.038	0.050	0.034	0.035	0.024
	(−0.704)		(1.895)*		(1.497)	
PTRLM	−0.105	0.045	−0.040	0.031	0.023	0.017
	(−2.169)**		(1.659)*		(0.643)	
RUBBR	0.007	0.038	0.005	0.009	−0.018	0.014
	(0.136)		(0.495)		(0.546)	
LETHR	0.077	0.042	−0.002	0.007	−0.015	0.010
	(2.252)**		(0.294)		(0.412)	
GLASS	0.024	0.038	0.018	0.010	−0.026	0.020
	(0.492)		(0.648)		(0.984)	
METAL	−0.085	0.044	−0.045	0.031	−0.042	0.033
	(−2.308)**		(1.674)*		(1.694)*	
MTLPR	0.073	0.039	−0.023	0.020	−0.025	0.020
	(1.001)		(1.054)		(0.878)	
MACHN	0.067	0.041	0.021	0.021	0.032	0.024
	(1.389)		(1.167)		(1.324)	
ELCTR	0.035	0.038	0.003	0.004	−0.010	0.008
	(0.687)		(0.390)		(0.435)	
CARS	0.006	0.038	0.002	0.003	−0.031	0.023
	(0.109)		(0.278)		(1.342)	
INSTR	0.049	0.039	0.003	0.004	−0.005	0.002
	(0.885)		(0.305)		(0.293)	
MANUF	0.043	0.040	0.033	0.021	−0.015	0.009
	(1.106)		(1.280)		(0.453)	
TRANS	0.090	0.040	−0.010	0.009	−0.033	0.024
	(2.142)**		(0.592)		(1.348)	
PHONE	−0.018	0.038	0.033	0.023	0.043	0.035
	(−0.419)		(1.378)		(1.721)*	
TV	0.067	0.042	0.043	0.032	0.009	0.007
	(1.769)*		(1.653)*		(0.193)	
UTILS	0.088	0.039	−0.035	0.025	0.023	0.018
	(2.228)**		(1.568)		(0.970)	
WHLSL	0.047	0.039	−0.025	0.019	−0.033	0.023
	(0.838)		(1.328)		(1.400)	
RTAIL	0.091	0.041	0.003	0.004	−0.029	0.018
	(2.506)**		(0.246)		(1.102)	
MONEY	0.180	0.046	0.058	0.035	−0.019	0.015
	(2.187)**		(1.854)*		(0.751)	
SRVC	0.086	0.043	0.051	0.033	−0.012	0.011
	(2.223)**		(1.721)*		(0.403)	
Wald test	75.830		47.017		27.193	
(p-value)	<0.01		0.068		0.789	

The dependent variable in Column (1) of Table 3 is next month's market return. The industries that have significant coefficients in this column are denoted by asterisks. There are 12 industries including commercial real estate (RLEST), mines (MINES), apparel (APPRL), print (PRINT), petroleum (PTRLM), leather (LETHR), metal (METAL), transportation (TRANS), utilities (UTILS), retail (RTAIL), money or financial (MONEY), and services (SRVC) that have *t*-statistics of the corresponding lagged industry return that are greater than 1.96 in absolute value (or significant at the 5% level). Two additional industries, non-metallic minerals (STONE) and television (TV), have *t*-statistics of about 1.7. So at the 10% level of significance (or *t*-statistics greater than 1.65 in absolute value), there are a total of fourteen industries that can significantly predict the market return.

The signs on the predictability coefficients for these 14 industries also seem to make economic sense. For instance, the lagged returns of petroleum and metal industry portfolios are negatively related to next period's market return as one might suspect since these are commodity (input) prices whose shocks have historically led the economy into a downturn. In contrast, retail and apparel are sectors that, when they are booming, are generally thought to be signs of a thriving economy. The fact that the signs of these predictive relations are consistent with conventional wisdom on the relation of these industries to the macroeconomy reassures us that these predictive regressions are indeed capturing the slow diffusion of sector information into the broad market index as opposed to being the result of chance (see also Section 4.2 below).

Finally, note that our findings are not simply an artifact of industry returns being serially correlated. First, most of the industries represent a small fraction of the market. So it is not likely that they forecast the market simply because their returns are serially correlated and part of the market portfolio. Second, the time series of most of the industry portfolios that can lead the market such as commercial real estate, apparel, petroleum, metal, and utilities are not (statistically significantly) serially correlated at a monthly frequency. They are also not an artifact of the market portfolio being auto-correlated since we control for lagged market returns in our predictive regressions. However, a number of industry portfolios, such as construction, smoke, textiles, retail, and money exhibit positive serial correlation (see also Grinblatt and Moskowitz, 1999). We omit these results for brevity.

An interesting empirical question is whether industries lead the stock market by more than one month. Our model predicts that there is such cross-predictability but is silent on the number of months by which industries ought to lead the stock market. However, we know from the literature on stock market predictability that being able to predict next month's return is already quite an achievement, as it is notoriously difficult to predict the market at long horizons. Indeed, Valkanov (2003) and Torous, Valkanov, and Yan (2004) argue that previous findings on long-horizon predictability are an artifact of not properly adjusting standard errors for the near random walk behavior of various predictors.

In Columns (2) and (3) of Table 3, we investigate whether these industry portfolios are able to lead the market by more than one month. In Column (2), the dependent variable is the market return in month $t+1$. The coefficient in front of the lagged industry return is now statistically significant at the 10% level for nine industries and at the 5% level for no

[7]Results for regressors such lagged RM, INF, DSPR, and MDY are the same when we use other industry returns to forecast the market. Again, this is why we present detailed results for only one of the industries.

industries. RLEST, APPRL, LETHR, TRANS, UTILS, and RTAIL are no longer statistically significant. In Column (3), the dependent variable is the market return in month $t+2$. Of the 14 industries that are significant in Column (1), only MINES and METAL are still statistically significant. Importantly, notice that at the 10% level of significance, only four industries have a statistically significant coefficient in front of lagged industry return and none have such a coefficient at the 5% level. In other words, the evidence is consistent with information taking about two months to be completely incorporated from industries into the broad market index.

Finally, in Table 3, we present a joint test of the null hypothesis that all the λ_i's are equal to zero ($\lambda_1 = \lambda_2 = \ldots = \lambda_{34} = 0$) for each of the three dependent variables (month t, month $t+1$, and month $t+2$ market returns). To do this, we stack the regressions given by (3) into a GMM system and calculate the Wald test of the joint significance of the coefficients. Similar to the t-statistics, the Wald tests (and all other tests reported in the paper) take into account cross-industry correlation in the residuals which is then used to compute a Newey–West heteroskedasticity and serial correlation robust variance covariance matrix with three monthly lags. Values of the Wald tests and corresponding p-values are reported in the final row of Table 3. For month-t returns, we can reject the null hypothesis that all the λ_i's are equal to zero at the 5% level of significance. For month $t+1$ returns, we can only reject the null at the 10% level of significance. And for month $t+2$ returns, we cannot reject the null.

We have also looked at cross-predictability at horizons of up to six months and find that there is virtually no predictability at longer horizons. This is a comforting finding since it suggests that our predictive regressions are informative and not subject to some bias that mechanically yields significant results.

Table 4

Industry predictive regressions—economic significance

In Panel A, the column "Economic Significance" computes the response of the market return to a two-standard-deviation shock of the corresponding industry return using the point estimates from column (1) of Table 3. Lower and upper bounds are given in parentheses below. The column "Absolute relative significance" computes the absolute value from dividing "Economic significance" by the standard deviation of the market return. Same calculations are done for the other market predictors in Panel B. In Panel C, we provide performance statistics for portfolios (market and risk-free rate) formed based on the various market predictors.

Industry	Economic significance	Absolute relative significance	Industry	Economic significance	Absolute relative significance
Panel A: By industries					
MONEY	1.741	0.422	MANUF	0.546	0.132
	(0.149, 3.332)			(−0.441, 1.532)	
RLEST	1.513	0.367	INSTR	0.526	0.128
	(0.368, 2.658)			(−0.663, 1.714)	
PRINT	1.496	0.363	WHLSL	0.521	0.126
	(0.277,2.714)			(−0.722, 1.764)	
APPRL	1.222	0.297	CHEMS	−0.514	0.125
	(−0.002, 2.447)			(−1.976, 0.947)	
SRVC	1.121	0.272	ELCTR	0.436	0.106
	(0.112, 2.130)			(−0.834, 1.707)	
METAL	−1.039	0.252	AGRIC	0.391	0.095
	(−1.939, −0.139)			(−0.420, 1.202)	

Table 4 (*continued*)

Industry	Economic significance	Absolute relative significance	Industry	Economic significance	Absolute relative significance
Panel A: By industries					
PTRLM	−1.040 (−1.999, −0.081)	0.252	CHAIR	0.331 (−0.679, 1.342)	0.080
TRANS	1.031 (0.068, 1.994)	0.250	SMOKE	−0.283 (−1.090, 0.524)	0.069
LETHR	0.960 (0.107, 1.812)	0.233	GLASS	0.282 (−0.863, 1.426)	0.068
RTAIL	0.932 (0.188, 1.675)	0.226	PAPER	−0.256 (−1.516, 1.003)	0.062
TV	0.900 (0.117, 1.917)	0.218	CNSTR	0.250 (−0.804, 1.303)	0.061
MINES	−0.839 (−1.678, 0.000)	0.204	WOOD	−0.189 (−1.213, 0.835)	0.046
TXTLS	0.783 (−0.212, 1.779)	0.190	PHONE	−0.169 (−0.976, 0.638)	0.041
MACHN	0.780 (−0.343, 1.903)	0.189	OIL	−0.120 (−1.047, 0.807)	0.029
MTLPR	0.712 (−0.711, 2.134)	0.173	FOOD	0.093 (−1.060, 1.247)	0.023
UTILS	0.681 (0.070, 1.292)	0.165	RUBBR	0.084 (−1.155, 1.323)	0.020
STONE	−0.582 (−1.257, 0.094)	0.141	CARS	0.065 (−1.125, 1.255)	0.016

	Economic significance	Absolute relative significance
Panel B: By other market predictors		
RM(−1)	0.428 (0.990, −0.133)	0.104
INF(−1)	−1.129 (−1.765, −0.493)	0.274
DSPR(−1)	0.036 (0.104, −0.032)	0.009
MDY(−1)	0.723 (0.069, 1.377)	0.175
MVOL(−1)	0.869 (0.218, 1.520)	0.211

	Market	Timing market and risk-free using only $\boldsymbol{Z_{t-1}}$	Timing market and risk-free using $\boldsymbol{Z_{t-1}}$ and R_{t-1}
Panel C: Performance statistics of various portfolios (figures in annualized %, with the exception of the relative MSE)			
Relative MSE	—	1.000	0.979
Mean	13.61	12.03	10.34
Standard deviation	15.62	11.27	7.40
Sharpe ratio	0.46	0.50	0.53
Certainty equivalent (relative risk aversion = 1)	8.24	9.44	11.07
Certainty equivalent (relative risk aversion = 5)	7.93	8.40	8.91

In Table 4, we take a more careful look at the ability of these industries to forecast the next month's market return. The goal is to compare the forecasting power of industry returns to those of well-known predictors such as inflation, dividend yield, term spread, and default spread as a gauge of the economic relevance of our findings. Toward this end, we calculate the effect of a two-standard-deviation shock to an industry's lagged monthly return on the next month's market return (using the coefficient estimates from Column (1) of Table 3). We report right below a lower and upper bound on the estimate obtained by using the standard errors of the coefficient estimates in Column (1) of Table 3. In addition, we report the absolute value of this magnitude as a fraction of market volatility. For instance, the coefficient in front of METAL from column (1) of Table 3 is -0.085 and plus or minus two standard errors gives us -0.159 as the lower bound and -0.011 as the upper bound. A two-standard-deviation shock in the monthly return of the metal industry is ($2*6.1\%$). So it leads to a change in next month's market return of 1.039% ($-0.085*2*6.1\%$). The lower and upper bounds on this estimate are obtained by using -0.159 and -0.011 instead of -0.085. The estimate of 1.039% is roughly 25% of market volatility.

The industries are listed in descending order, with the most economically significant industry first. As one might expect, the 14 industries that have a statistically significant ability to predict the market are also among the leaders in terms of economic significance. MONEY is very significant, with a two-standard-deviation shock in its returns resulting in a movement of market returns that is 40% of market volatility. The next most economically significant is real estate (RLEST). Print (PRINT), apparel (APPRL), and services (SRVC) round out the top five.

We perform the same exercise for the other market predictors (INF, DSPR, MDY, and MVOL) and report the results in Panel B. The coefficients in front of these other market predictors change very little across the separate industry-by-industry regressions. To save on space, we only discuss the coefficient estimates in front of these other market predictors from the METAL regression, which are already reported in Table 2, specification (3). INF is by far economically the strongest of the usual market predictors. A two-standard deviation shock in inflation (INF) leads to a 1.129% ($-0.624*2*0.905$) movement in the market, which is roughly 27% of market volatility. Our industry portfolios such as METAL do a comparable job of forecasting the market. These results must be interpreted with caution as the predictors have been shown to forecast expected returns over longer horizons. They are slower moving and might not capture short-term fluctuations as do the industry returns. Nonetheless, comparing Panel A to Panel B indicates that the industry returns are economically significant leading indicators of the market returns.

Finally, we see in Panel C whether our industry portfolios help us to better allocate between the market and the risk-free rate. That is, we form a portfolio comprising the market and the risk-free rate based on information from lagged industry returns and we investigate whether this portfolio does better than just passively owning the market. This exercise provides additional evidence about the usefulness of these predictors, though the comparison is not perfect since there are transaction costs associated with moving in and out of the market into the risk-free instrument that we do not take into account.

We do this market-timing exercise using rolling forecasts in the following way. Using the first ten years of data, from January 1946 to December 1955, we estimate two predictive regressions. The first is of RM_t on $\boldsymbol{Z}_{t-1}$ (INF, DSPR, MDY, and MVOL). The second is of RM_t on $\boldsymbol{Z}_{t-1}$ and simultaneously all the industry returns. Using these estimated coefficients, we form one-period-ahead forecasts of RM_t. Then, we re-estimate the two regressions for every subsequent month and use the estimates to form one-period-ahead forecasts of RM_t. The mean squared errors (MSE) of these rolling forecasts are computed. We report the relative MSE (Relative MSE) for the case in which we only use the other market predictors ($\boldsymbol{Z_{t-1}}$) and the case in which we use both $\boldsymbol{Z_{t-1}}$ and all the industry returns as predictors. We observe an improvement in MSE of about 2.1%.

We then consider the following portfolio strategies, which we implement using data from the rolling regression forecasts. For each month, if the forecasted RM is higher than the risk-free asset (in the last period), we invest everything in the market for a return RM. Otherwise, we invest everything in the risk-free asset. Thus, we obtain returns from two investment strategies, one using only $\boldsymbol{Z_{t-1}}$ and the other using $\boldsymbol{Z_{t-1}}$ as well as all the lagged industry returns. The mean, standard deviation, Sharpe ratio, and certainty equivalent returns are reported in Table 4 for these two strategies along with the strategy of passively holding the market. The certainly equivalent returns are computed as follows. We compute the utility $\mathrm{U}(1+R_\mathrm{p})$, where R_p is the return from the strategy and $\mathrm{U}(x)$ is either $\log(x)$ or $x^{1-\mathrm{RRA}}/(1\text{-RRA})$, where RRA is the coefficient of relative risk aversion. Then, we compute the average utility over the period and invert it to obtain the certainly equivalent return.

Notice that the means of these two strategies are lower than simply just holding the market. The standard deviations of these two strategies are also lower, leading to higher Sharpe ratios than just simply holding the market. The Sharpe ratio of the market during the latter half of the sample is 0.46, while the strategy incorporating only $\boldsymbol{Z_{t-1}}$ information yields a Sharpe ratio of 0.50 and the one using both $\boldsymbol{Z_{t-1}}$ and all the industry returns yields a Sharpe ratio of 0.53.

Moreover, it appears that industry information also leads to higher certainty equivalent returns. The interpretation of the certainty equivalent return is that an investor with an RRA parameter of one will be willing to pay up to 1.6% on an annual basis to use the lagged industry returns in forming a portfolio. Naturally, as the RRA parameter increases, the investor becomes more risk averse and the advantage of the strategy for RRA $= 5$ is only about 0.5%.

The point of this comparison is to show that, relative to other predictors, industry returns do have some additional information about the market. However, we caution against jumping to the conclusion that this is a more profitable strategy because of higher frequency of trading. For instance, using the strategy with $\boldsymbol{Z_{t-1}}$ only, we are rebalancing between the risk-free and risky assets 84 times and are in the risk-free asset 204 months out of a total of 564 months. In contrast, we are re-allocating between the two assets 158 times and are in the risk-free asset 279 months out of a total of 564 months using the strategy that also incorporates industry information. This suggests that the latter strategy could involve substantially more transaction costs.

4.2. Industry returns and market fundamentals

In this section, we attempt to test Prediction 2, that an industry's ability to predict the market ought to be correlated with its ability to forecast indicators of economic activity (i.e., market fundamentals). We begin by specifying the regression for forecasting market fundamentals:

$$X_t = \eta_i + \gamma_i R_{i,t-1} + \boldsymbol{C_i ZX_{t-1}} + v_{i,t}, \tag{4}$$

where X_t is the month t realization of the indicator of economic activity, $R_{i,t-1}$ is the previous month's return of industry i, and the $\boldsymbol{ZX_{t-1}}$ is the same as $\boldsymbol{Z_{t-1}}$ in Eq. (3) except that we also include three monthly lags of the indicators of economic activity.

The coefficients of interest are the γ_i's, which measure the ability of the various industry returns to predict the economic activity indicator of interest.

To the extent that Prediction 2 holds, we expect the relation between the λ_i's and γ_i's to be positive. In other words, the industries that can strongly forecast the market ought to also forecast market fundamentals. For instance, industries such as commercial real estate that have a positive λ_i ought to also have a positive γ_i. And industries such as metals or petroleum that have a negative λ_i ought to also have a negative γ_i.

To implement the regression specified in Eq. (4), we need to identify proxies for economic activity. We use two well-known measures that have been previously studied in the literature. The first is industrial production growth, IPG. We use this measure because it is one of the few measures of economic activity that are available at a monthly frequency. Industrial production growth is contemporaneously correlated with the aggregate market. Over the period 1946–2002, IPG and RM have a contemporaneous correlation of 0.147.

The second measure of economic activity that we use is SWG, the monthly growth rate of the Stock and Watson (1989) coincident index of economic activity. SWG is also contemporaneously correlated with the aggregate stock market. Over the period of March 1959–December 2002, SWG and RM have a (monthly) contemporaneous correlation of 0.030.

In Table 5, we determine which of the 34 industries can forecast industrial production growth. The regression specification is Eq. (4) but we only report the coefficient of lagged industry return. As in Eq. (3), the standard errors are formulated to account for correlation of the residuals from the 34 industry returns at a point in time as well as for serial correlation. Twelve of the 34 industries are statistically significant at the 10% level and nine are significant at the 5% level. Our finding that industries contain valuable information about future economic fundamentals is consistent with Lamont (2001), who finds that portfolios formed from industry returns can track various economic variables like industrial production growth, inflation, and consumption growth.

Finally, we present a joint Wald test of the null hypothesis that all the γ_i's are equal to zero ($\gamma_1 = \gamma_2 = \cdots = \gamma_{34} = 0$). To do this (as we did at the end of Table 3) we stack the regressions given by (4) into a GMM system and calculate the Wald statistic of the joint significance of the coefficients. Values of the Wald tests and corresponding p-values are reported in the final row of Panel A. We can reject the null hypothesis that all the λ_i's are equal to zero at the 5% level of significance.

More importantly, it appears that the industries that forecast the market (from Table 3) also forecast industrial production growth. Recall from Table 3 that MINES, PTRLM, and METAL negatively forecast the market: higher returns in these industries in month t lead to lower returns in the market the next month. Interestingly, these three industries also forecast industrial production growth with a negative coefficient: higher returns in these industries in month t lead to lower industrial production growth the next month. This is exactly what we would expect with the slow incorporation of information into the broad market index. Moreover, RETAIL, MONEY, and RLEST, which are positively cross-serially correlated with the market, also forecast industrial production growth with a positive coefficient.

To formally see that an industry's ability to forecast the market is indeed correlated with its ability to forecast industrial production growth, we plot λ_i on the y-axis and γ_i on the x-axis in Fig. 1 (US:IPG). As a benchmark, recall that in an efficient market, we would expect all the λ_i's to be around zero, i.e., the slope of the scatter plot ought to be zero.

Table 5

Predictive regressions of measures of economic activity using industry portfolios

This table presents results of forecasting IPG, industrial production growth in month t, using various industry portfolio returns at month $t-1$ separately and other information available at month $t-1$. The other forecasting variables are lagged RM, INF, DSPR, MDY, MVOL, and three monthly lags of IPG. We only report the coefficient in front of the lagged industry return. The standard errors (used in the computation of the *t*-statistics) are adjusted for cross-industry correlation in the error terms using an estimate of the covariance matrix computed with estimated residuals from the 34 regressions at a point in time. The standard errors also include a Newey–West correction with three monthly lags. We also report Wald tests of the joint significance of the coefficients estimated in a GMM system (joint test of the null hypothesis that the lambdas are zero). The Wald tests (34 d.f.) are computed using Newey–West heteroskedasticity and serial correlation robust covariance matrix. The sample period for the IPG regressions is January 1946 to December 2002, with the exception of RLEST, which is from January 1972 to December 2002. *Significant at 10% level. **Significant at 5% level.

Forecast of IPG (US) using industry returns					
Industry	IND(−1)	R^2	Industry	IND(−1)	R^2
RLEST	0.020	0.209	LETHR	0.016	0.109
	(1.889)*			(1.675)*	
AGRIC	0.002	0.106	GLASS	0.012	0.107
	(0.331)			(0.934)	
MINES	−0.001	0.106	METAL	−0.023	0.111
	(−0.137)			(−1.999)**	
OIL	0.002	0.106	MTLPR	0.006	0.106
	(0.171)			(0.346)	
STONE	−0.009	0.108	MACHN	0.017	0.108
	(−1.330)			(1.363)	
CNSTR	−0.021	0.112	ELCTR	0.031	0.113
	(−2.227)**			(2.438)**	
FOOD	−0.037	0.112	CARS	0.038	0.116
	(−2.140)**			(2.868)**	
SMOKE	−0.012	0.108	INSTR	−0.026	0.110
	(−1.309)			(−1.934)*	
TXTLS	0.027	0.115	MANUF	0.006	0.106
	(2.649)**			(0.563)	
APPRL	0.012	0.108	TRANS	0.014	0.107
	(1.237)			(1.062)	
WOOD	0.010	0.107	PHONE	−0.004	0.106
	(1.202)			(−0.382)	
CHAIR	0.005	0.106	TV	0.006	0.106
	(0.475)			(0.623)	
PAPER	0.009	0.106	UTILS	0.042	0.116
	(0.604)			(2.827)**	
PRINT	−0.021	0.108	WHLSL	0.007	0.106
	(−1.489)			(0.499)	
CHEMS	−0.041	0.111	RTAIL	0.009	0.106
	(−2.087)**			(0.606)	
PTRLM	−0.002	0.106	MONEY	0.052	0.114
	(−0.175)			(2.532)**	
RUBBR	0.014	0.107	SRVC	0.011	0.107
	(1.150)			(1.029)	
Wald test	61.292				
p-value	<0.01				

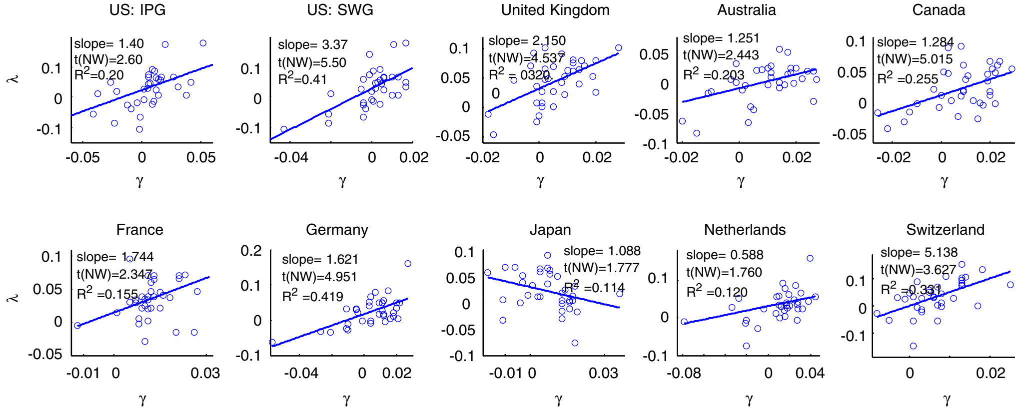

Fig. 1. The relation between an industry's ability to lead the market and its ability to predict economic activity (US and rest of the world) The first plot in the figure, US: IPG, presents a scatter-plot of the coefficients λ_i obtained by forecasting the market (RM) using 34 industry returns and other conditioning information on the coefficients γ_i obtained by forecasting industrial production growth (IPG) using the same 34 industry returns. The linear relation between the two sets of coefficients is plotted with a solid line. The slope of the line, Newey–West t-statistic, and R^2 are also presented. The second plot, US: SWG, displays the same relation between the coefficients λ_i and γ_i, where the γ_i are obtained by forecasting the growth rate of the Stock and Watson (1989) index of economic activity (SWG). The next eight plots present similar scatter-plots, for each of the eight largest stock markets outside of the US, of the coefficients λ_i obtained by forecasting RM using industry returns and other conditioning information on the coefficients γ_i obtained by forecasting industrial production growth (IPG) using the same industry returns.

In contrast, we see a distinctly positive relation between the λ_i's and the γ_i's. In Fig. 1 (US:IPG), we also plot the fitted values from a linear regression of λ_i's on γ_i's. The slope coefficient is 1.40 with a t-statistic of 2.60. In other words, there is a strong positive correlation between the ability of industries to forecast the market and their ability to forecast industrial production growth. (We have also conducted other statistical inference exercises to check that our results are not due purely to chance; see Section 4.3 below.)

We obtain similar results when we use SWG, the percentage change in the Stock and Watson coincident index of economic activity. We omit the results corresponding to Table 5 for brevity. However, in Fig. 1(US:SWG), we plot the fitted values from a linear regression of the λ_i's on γ_i's. The slope coefficient is 3.37 with a t-statistic of 5.50. In other words, there is a strong positive correlation between the ability of industries to forecast the market and their ability to forecast economic activity when we use an alternative measure of the change in market fundamentals.

One potential worry is that the results in Fig. 1 are due to coincidence/measurement error. Suppose that, for whatever accidental reason, high returns in our sample period for, say, the retail industry just happen to be followed by increases in industrial production. Since increases in industrial production will be naturally contemporaneously correlated with positive returns in the stock market, it is likely that retail will then also lead the market. In other words, in a given sample, the measurement error in our estimate of the coefficient of industry x returns on future industrial production is going to be correlated with the measurement error in our estimate of the coefficient of industry x returns on future market returns. This can generate a pattern like that in Fig. 1 even if the underlying true coefficients are not related.

One conservative way to deal with this issue is to split the sample period in half and, for every industry, estimate one of the coefficients in the first half and the other in the second half. (This is very conservative because there might be genuine time variation in the parameters.) We can then do similar exercises to Fig. 1. We use the 1946–1973 subsample to estimate the λ_i's for RM and the 1974–2002 subsample to estimate the γ_i's for IPG. (We get similar results if we use the 1946–1973 subsample to estimate the γ_i's and the 1974–2002 subsample to estimate the λ_i's.) We then re-do the analysis in Fig. 1(a). The coefficient is 2.3 with a *t*-stat of 1.97. We repeat the same exercise for SWG. We use the March 1959–December 1981 subsample to estimate the λ_i's for RM and the 1982–2002 subsample to estimate the γ_i's for SWG. We then re-do the analysis in Fig. 1(b). The coefficient is 2.9 with a *t*-stat of 2.3. These findings are similar to those obtained in Fig. 1. For brevity, we omit the figures. Hence, we can conclude that our findings in Fig. 1 are not due to coincidence/measurement error.

4.3. Additional analysis

4.3.1. Numerical simulations

Rather than preserving the time series ordering of the market returns, we shuffle them and pretend that the shuffled series is the true time series of the market. We then run regression (3) for all 34 industries and save the coefficients using this randomized market series. We repeat this procedure 10,000 times. Each time, we also keep track of the number of coefficients that are significant at the 5% and 10% levels of significance. The number of simulations in which we have 14 significant λ_i's is three out of 10,000. The number of simulations in which we have 15 significant λ_i's is one. There are no simulations with 16 or more significant λ_i's. Hence, in four out of 10,000 simulations, or 0.04%, we find 14 or more significant results.[8] We have repeated a similar analysis for regression (4) as well as the regressions presented in Fig. 1 and conclude that our findings are not spurious.

4.3.2. Subperiods

We also arbitrarily divide our sample into two equal subsamples (1946–1973 and 1974–2002) to see whether the ability of these industries to lead the market differs across these subperiods. There is not a marked difference in the results across the two subsamples. We omit these results for brevity.

4.3.3. Forecasting the market and indicators of economic activity using all industries simultaneously

What happens when we simultaneously include all industries in our forecasting specifications? In Panel A of Table 6, we answer this question by forecasting the market and the indicators of economic activity using all industries simultaneously along with the control variables specified in Tables 3 and 5, respectively. (Note that since RLEST only goes back to the early 1970s, this industry portfolio is excluded from the current analysis.)

[8] On average (across the 10,000 trials), about 5.4 industries have significant coefficients at the 10% level of significance and 2.6 industries at the 5% level of significance. These small numbers indicate that our findings are not due to chance.

Table 6
Predictive regressions of market returns and economic activity using all industry returns simultaneously (US)

This table presents results of forecasting the market and various indicators of economic activity in month t using all industry portfolio returns at month $t-1$ (excluding RLEST) and other information available at month $t-1$. RM is the market return, IPG is industrial production growth, and SWG is the growth rate of the Stock and Watson (1989) index of economic activity. The other control variables are lagged RM, INF (the CPI inflation rate), DSPR (the default spread between BAA-rated and AAA-rated bonds), MDY (the dividend yield of the market portfolio), MVOL (the market volatility), and in the cases of IPG and SWG, three lags of the dependent variable. In Panel A, for each of the three regressions, we test three hypotheses: (1) whether the coefficients on the lagged industry returns are jointly zero; (2) whether the coefficients on the other lagged control variables are jointly zero; and (3) whether the coefficients on the lagged industry returns and the lagged controls are jointly zero. The *F*-tests and *p*-values are reported below. The degrees of freedom of the *F*-tests are determined as in Hayashi (2000, p. 43). In all tests, we use Newey–West serial correlation and heteroskedasticity robust standard errors calculated with three monthly lags. In Panel B, we report the parameter of proportionality obtained from estimating the predictive regressions of RM and IPG using all industry returns jointly and testing whether the predictive coefficients from these two regressions are proportional. *Significant at 10% level. **Significant at 5% level.

	RM		IPG		SWG	
Null Hypothesis	*F*-test	*P*-value	*F*-test	*P*-value	*F*-test	*P*-value
Panel A: Forecasting using all industry returns simultaneously						
Coefficients of all industries equal zero	1.947	<0.01	2.085	<0.01	1.652	0.015
Coefficients of all controls equal zero	2.236	<0.01	8.482	<0.01	8.446	<0.01
Coefficients of all industries and controls equal zero	2.321	<0.01	2.828	<0.01	2.450	<0.01

	RM and IPG	RM and SWG
Panel B: Alternative method of estimating the relation between an industry's ability to lead the market and its ability to predict fundamentals		
Parameter of Proportionality	2.142 (2.644)**	3.620 (2.488)**

The values of the *F*-tests (with 33 restrictions) under the various null hypotheses are reported. The first row reports the *F*-tests and *p*-values for three separate null hypotheses. The first is whether the industry returns jointly do not forecast the market (RM). The *p*-value is less than 0.01, which means that we can strongly reject this null at the 5% level. The other two hypotheses are whether the industry returns can jointly forecast the various indicators of economic activity (IPG and SWG). Again, in the cases of IPG and SWG, we can reject the null at the 5% level. The degrees of freedom of the *F*-tests are determined as in Hayashi (2000, p. 43).

The second row reports the *F*-tests and *p*-values for the null hypotheses that the control variables jointly do not forecast the market and the indicators of economic activity. In each case, we can reject these null hypotheses at the 5% level of significance. The third row reports the *p*-values for the null hypotheses that all industry returns and all control variables do not jointly forecast the market or indicators of economy activity. The results are similar to those in the first and second rows in that we can reject that these variables do not jointly have forecasting power.

We do not report the coefficients from the regressions in Table 6 for brevity. However, we want to point out that the coefficients from these forecasting regressions are similar to the coefficients obtained in Tables 3 and 5. Moreover, we have also re-done the analysis in Fig. 1 using the coefficients from these alternative regression specifications and obtain similar results. So even though the coefficients are estimated much more imprecisely when we include all industries simultaneously, the economic messages that we obtain from Table 3, Table 5, and Fig. 1 are confirmed.

4.3.4. Alternative methodology for testing Prediction 2

Finally, we attempt to test Prediction 2 in a more parameterized manner. Rather than estimating the regression specifications in (3) and (4) separately and then analyzing the relation between the estimated coefficients, the λ_i's and the γ_i's, as in Fig. 1, we attempt to impose the restriction that these coefficients are proportional to each other, $\gamma_i = \kappa\lambda_i$, and estimate the coefficient κ and λ_i's from the following specification:

$$\begin{aligned} \mathrm{RM}_t &= \alpha + \lambda_1 R_{1,t-1} + \cdots + \lambda_N R_{N,t-1} + \boldsymbol{AZ_{t-1}} + e_t \\ X_t &= \eta + \kappa(\lambda_1 R_{i,t-1} + \cdots + \lambda_N R_{N,t-1}) + \boldsymbol{CZX}_{\boldsymbol{t-1}} + v_t. \end{aligned} \tag{5}$$

where N is the number of industry portfolios in the estimation, $\boldsymbol{Z_{t-1}}$ and $\boldsymbol{ZX_{t-1}}$ are as in Eqs. (3) and (4), and e_t and v_t are random errors. We can estimate the system in (5) using GMM. Note that this estimation excludes the real estate portfolio since we can only obtain data back to 1973. The null hypothesis of interest is that $\kappa = 0$, whereas if Prediction 2 holds true, we expect to find that $\kappa > 0$. The results are reported in Panel B of Table 6. We find that the coefficient $\kappa = 2.142$ with a t-statistic of 2.644 for IPG. The comparable numbers of SWG are, respectively, 3.620 and 2.488. These numbers are similar to those reported in Fig. 1. So the results of this estimation strongly confirm that there is a statistically significant positive relation between the predictive content of an industry for the market and the information it has for economic activity.

4.3.5. Alternative measures of economic activity

We have also experimented with how past industry returns forecast the deviations of these two macroeconomic variables from a potentially stochastic trend. We use the popular band-pass filter developed by Baxter and King (1999), the codes for which are available on Marianne Baxter's website. In practice, different industries can forecast fluctuations in economic activity at different frequencies. For the sake of parsimony, however, we set the parameters of this filter to capture fluctuations in industrial production and the Stock and Watson index at frequencies between two and 12 months.[9] The results are similar (see Hong, Torous, and Valkanov, 2002).

5. Evidence from the rest of the world

In this section, we extend the empirical analysis documented in Section 4 (for the US stock market) to each of the eight largest stock markets outside of the US In Table 7, we present the results of estimating the regression specifications given in (3) and (4) to the eight largest stock markets outside of the US. For regression specification (3), we now include as a control the lagged monthly market return. We are unable to control for other market predictors used in the US stock market sample due to lack of data. For regression

[9]More specifically, the parameter k for the weighted average of industrial production around month t is 19 months.

Table 7
Predictive regressions of market returns and economic activity by industry returns (International Evidence)

This table presents results of forecasting the market (RM) and industrial production growth (IPG) in month t using industry portfolio returns at month $t-1$ and lagged RM as a control for each of the eight largest markets outside of the US In Panel A, we report the results of forecasting the market and IPG using individual industry portfolios separately with the lagged market as a control: the number of industry portfolios that can significantly predict the market at the 10% level of significance. In Panel B, we report the F-test of forecasting the market and IPG using all industry portfolios simultaneously with the lagged market as a control. The F-tests and p-values are reported below. The degrees of freedom of the F-tests are determined as in Hayashi (2000, p. 43). In all tests, the standard errors are adjusted for cross-industry correlation in the error terms using an estimate of the covariance matrix computed with estimated residuals from the industry regressions at a point in time. The standard errors also include a Newey–West serial correlation and heteroskedasticity correction with three monthly lags. In Panel C, we report the parameter of proportionality obtained from estimating the predictive regressions of RM and IPG using all industry returns jointly and testing whether the predictive coefficients from these two regressions are proportional. The sample is from 1973 to 2002. *Significant at 10% level. **Significant at 5% level.

	Panel A: Individual regressions		*Panel B: Joint regressions*	
Country	RM At 10 percent	IPG At 10 percent	RM F-test	IPG F-test
United Kingdom	8 of 34	6 of 34	4.584**	5.753**
Australia	18 of 31	11 of 31	5.458**	6.764**
Canada	11 of 31	12 of 31	5.049**	6.077**
France	14 of 32	12 of 32	4.923**	6.983**
Germany	12 of 31	10 of 31	5.167**	5.403**
Japan	8 of 34	10 of 34	4.735**	5.790**
Netherlands	16 of 29	11 of 29	5.274**	6.248**
Switzerland	14 of 27	10 of 27	4.753**	6.452**

	Panel C: Parameter of proportionality (RM and IPG)	
Country	Parameters	T-statistic
United Kingdom	2.491	(3.795)**
Australia	1.036	(2.159)**
Canada	1.343	(3.353)**
France	1.634	(3.140)**
Germany	1.517	(3.037)**
Japan	−0.780	(−1.190)
Netherlands	1.004	(2.450)**
Switzerland	3.248	(5.453)**

specification (4), we include as controls one lag of the monthly market return and three monthly lags of IPG.

In Panel A of Table 7, rather than listing all the coefficients for all countries, we report the number of industries that can significantly predict the market and IPG at the 10% level of significance for each country. Across the eight countries, the UK and Japan have the smallest proportion of significant industries (8 out of 34). Australia and the Netherlands have the most (18 out of 31 and 16 out of 29, respectively). The numbers for IPG are comparable. These findings are very similar to those obtained for the US stock market and they far exceed the threshold figures that one would expect just from chance (see discussion in Section 4.3). Hence, we can safely conclude that the basic findings that industry returns can predict both aggregate market returns and IPG extend beyond the US.

Moreover, the remarkable consistency of these findings across the eight countries outside of the US should calm any lingering doubts that our results are due to chance, which our numerical simulations did not alleviate or that our US findings are purely a by-product of data mining. Indeed, the consistency of these patterns strongly indicates that we are capturing genuine economic phenomena.

In addition to estimating regression specifications (3) and (4) for each country, we have also attempted to forecast the market and IPG using all industries simultaneously just as we did for the US (see Section 4.3). The results are quite similar to those presented in Table 6 for the US. Across all eight countries, the p-values from F-tests under the null hypothesis that all the coefficients of all industry returns are jointly equal to zero are less than 0.01. The F-tests are reported in Panel B of Table 7.

While it is comforting to know that industries do indeed have information about future aggregate stock market movements and IPG, the crux of our paper lies in whether the propensity of an industry to predict the stock market is related to its propensity to predict economic activity. So we take the estimates from regression specifications (3) and (4) for each country and estimate the relation between these two sets of coefficients for each country, just as we did for the US—we present these findings in Fig. 1. Notice that for every country except Japan there is a positive and statistically significant relation between the estimated λ_i's and γ_i's.

Finally, in Panel C of Table 7 we present the results of the restricted regressions given in (5) for each of the countries. Not surprisingly, we find that the results are consistent with those in Fig. 1. For every country except Japan, the coefficient of proportionality κ is positive and statistically significant, thereby strongly indicating that the propensity of industries to predict returns is correlated with their informativeness about fundamentals. In Panel C of Table 7, the parameter of proportionality from a regression of estimated λ_i's on estimated γ_i's for Japan is -0.78 with a t-statistic of -1.190. In other words, the relation in Japan between the propensity of an industry to predict the market and its propensity to predict economic activity is not statistically significant. While an alternative model specification might yield significant results for Japan, we want to stay with the simple specification developed for the US and see to what extent it holds in the rest of the world. We view the fact that it works for seven of the eight countries as confirming the robustness of our findings in the US.

At the suggestion of the referee, we investigated whether our results in Table 7 are driven by outliers. First, we drop industries with fewer than four years of data or fewer than five firms and check to see if we get the same results. This screen takes out a number of the outlier industries in Japan and Germany. So it appears that the large means and standard deviations of some of the Datastream industry portfolio returns are due in part to the fact that there are only a few stocks in some of the portfolios. More importantly, we replicate Table 7 using these outlier-robust data and the results are similar to those obtained without this additional data screen. Indeed, our results are stronger with the cleaned data. The interesting thing to note is that the coefficient on Japan which had a wrong sign is significantly smaller in absolute value (basically zero) using the cleaned data. Second, we also winsorize the industry portfolio returns at the 1% and 99% levels and the message is the same—dealing with outliers makes our results, if anything, more consistent with the theory.

6. Conclusion

We find that the returns of industry portfolios are able to predict the movements of stock markets. An industry's predictive ability is strongly correlated with its propensity to forecast indicators of economic activity. The results are similar for the eight largest non-US stock markets. These findings indicate that markets incorporate information contained in industry returns about their fundamentals only with a lag because information diffuses gradually across asset markets.

The logic of our hypothesis suggests that we also find cross-asset return predictability in many contexts beyond industry portfolios and the broad market index. The key is to first identify sets of assets whose payoffs are likely correlated. As such, other contexts for interesting empirical work include looking at whether returns of stocks from one industry predict those in another or looking at stocks and the options listed on them.

Indeed, a number of papers following ours have taken up this task and found confirming results. For instance, Menzly and Ozbas (2004) find that industry returns lead and lag each other according to their place in the supply chain. And Pan and Poteshman (2004) find that information could diffuse slowly from option markets to stock markets as option volume seems to be able to predict stock price movements. But much more work remains to be done on this topic.

References

Allen, F., Gale, D., 1994. Limited market participation and volatility of asset prices. American Economic Review 84, 933–955.

Badrinath, S., Kale, J., Noe, T., 1995. Of shepherds, sheep and the cross-autocorrelations in equity returns. Review of Financial Studies 8, 401–430.

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307–343.

Baxter, M., King, R., 1999. Measuring business cycles: Approximate band-pass filters for economic time series. Review of Economics and Statistics 81, 575–593.

Blume, M., Friend, I., 1978. The Changing Role of the Individual Investors: A Twentieth Century Fund Report. Wiley, New York.

Boudoukh, J., Richardson, M., Whitelaw, R., 1994. A tale of three schools: Insights on autocorrelations of short-horizon stock returns. Review of Financial Studies 7, 539–573.

Brennan, M., 1975. The optimal number of securities in a risky asset portfolio when there are fixed costs of transacting: Theory and some empirical results. Journal of Financial Quantitative Analysis 10, 483–496.

Brennan, M., Jegadeesh, N., Swaminathan, B., 1993. Investment analysis and the adjustment of stock prices to common information. Review of Financial Studies 6, 799–824.

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

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and over-reactions. Journal of Finance 53, 1839–1886.

Eleswarapu, V., Tiwari, A., 1996. Business cycles and stock market returns: Evidence using industry-based portfolios. Journal of Financial Research 19, 121–134.

Fama, E., French, K., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23–49.

Fama, E., French, K., 1997. Industry costs of equity. Journal of Financial Economics 43, 153–193.

Fama, E., Schwert, G.W., 1977. Asset returns and inflation. Journal of Financial Economics 5, 115–146.

French, K., Schwert, G.W., Stambaugh, R., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Grinblatt, M., Moskowitz, T., 1999. Do industries explain momentum? Journal of Finance 54, 1249–1290.

Hayashi, F., 2000. Econometrics. Princeton University Press, New Jersey.

Hirshleifer, D., Lim, S., Teoh, S.H., 2002. Disclosure to a Credulous Audience: The Role of Limited Attention. Ohio State University Working Paper.
Hong, H., Lim, T., Stein, J., 2000. Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. Journal of Finance 55, 265–295.
Hong, H., Stein, J., 1999. A unified theory of underreaction, momentum trading and overreaction in asset markets. Journal of Finance 54, 2143–2184.
Hong, H., Torous, W., Valkanov, R., 2002. Do industries lead the stock market? Gradual diffusion of information and cross-asset return predictability. FEN Working Paper 〈http://papers.ssrn.com/sol3/papers.cfm?abstract_id=326422〉.
Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. Journal of Finance 48, 93–130.
Jegadeesh, N., Titman, S., 1995. Overreaction, delayed reaction and contrarian profits. Review of Financial Studies 8, 973–993.
Kahneman, D., 1973. Attention and Effort. Prenctice-Hall, Englewood Cliffs, New Jersey.
King, M., Leape, J., 1984. Wealth and portfolio composition: Theory and evidence. NBER Working Paper No. 1468.
Lamont, O., 2001. Economic tracking portfolios. Journal of Econometrics 105, 161–184.
Lo, A., MacKinlay, C., 1990. When are contrarian profits due to stock market overreaction? Review of Financial Studies 3, 175–206.
Menzly, L., Ozbas, O., 2004. Cross-industry momentum. USC Working Paper.
Merton, R., 1987. A simple model of capital market equilibrium with incomplete information. Journal of Finance 42, 483–510.
Newey, W., West, K., 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Nisbett, R., Ross, L., 1980. Human Inference: Strategies and Shortcomings of Social Judgment. Prentice-Hall, New Jersey.
Pan, J., Poteshman, A., 2004. The information in option volume for future stock prices, MIT Working Paper.
Peng, L., Xiong, W., 2002. Capacity constrained learning and asset price comovements. Princeton University Working Paper.
Pollet, J., 2002. Predicting Asset Returns with Expected Oil Price Changes. Harvard University Working Paper.
Rogers, W., 1993. Regression standard errors in clustered samples. Stata Technical Bulletin Reprints STB-13–STB-18, 88–94.
Shiller, R., 2000. Irrational Exuberance. Broadway Books, New York.
Shleifer, A., Vishny, R., 1997. The limits of arbitrage. Journal of Finance 52, 35–55.
Sims, C., 2001. Rational Inattention. Princeton University Working Paper.
Stock, J., Watson, M., 1989. New indexes of coincident and leading economic indicators. NBER Macroeconomics Annual 351–394.
Torous, W., Valkanov, R., Yan, S., 2004. On predicting stock returns with nearly integrated explanatory variables. Journal of Business 77, 937–966.
Valkanov, R., 2003. Long-horizon regressions: Theoretical results and applications. Journal of Financial Economics 68, 201–232.
Vuolteenaho, T., 2002. What drives firm-level stock returns? Journal of Finance 57 (1), 233–264.