Review of Finance, 2017, 237–267
doi: 10.1093/rof/rfv065
Advance Access Publication Date: 10 February 2016

Investing in Disappearing Anomalies*

Christopher S. Jones[1] and Lukasz Pomorski[2]

[1]Marshall School of Business, University of Southern California and [2]AQR Capital Management, LLC

Abstract

We argue that anomalies may experience prolonged decay after discovery and propose a Bayesian framework to study how that impacts portfolio decisions. Using the January effect and short-term index autocorrelations as examples of disappearing anomalies, we find that prolonged decay is empirically important, particularly for small-cap anomalies. Papers that document new anomalies without accounting for such decay may actually underestimate the original strength of the anomaly and imply an overstated level of the anomaly out of sample. We show that allowing for potential decay in the context of portfolio choice leads to out-of-sample outperformance relative to other approaches.

JEL classification: G11 (primary), G12, C11

1. Introduction

Documenting, explaining, and debunking anomalies is a prime fodder for the empirical asset pricing literature. Anomalies may improve our understanding of financial markets by posing a challenge to the joint hypothesis of market efficiency and an asset pricing model, perhaps leading to a new priced risk factor or helping uncover new market frictions or barriers to arbitrage activity. For that to happen, however, we need to understand the likely drivers of the anomaly and assure that it is not merely a statistical artifact. To do so, it is often helpful to investigate how the anomaly evolves over time.

Prior literature suggests that many anomalies are not stable outside of the sample in which they were discovered. Schwert (2003) suggests that some anomalies, notably the size and value effects, are not robust across sample periods and attributes at least part of the attenuation in abnormal returns to the dissemination of academic research findings. Hand, Green, and Soliman (2011) document the demise of the accruals anomaly and show that it

* We are grateful to Burton Hollifield (the editor), an anonymous referee, Narayan Bulusu, Eugene Fama, Jianjian Jin, Tom McCurdy, Tobias Moskowitz, Lubos Pastor, Pietro Veronesi, and the participants of seminars at the University of Chicago, University of Southern California, University of Toronto, York University, Pontificia Universidad Catolica de Chile, and the Bank of Canada Workshop on Advances in Portfolio Management for many helpful comments. All errors remain our own. The views expressed in this article are those of the authors.

likely persisted for close to 10 years following Sloan's (1996) discovery of it. However, other anomalies do not seem to disappear. Schwert (2003) shows that the January effect, which measures the tendency of small firms to outperform large firms in the month of January, has endured through the 1990s. Jegadeesh and Titman (2001) find that the magnitude of the momentum effect has been relatively unchanged since their 1993 study.

These examples show that anomalies may develop in very different ways following their discovery or publication. In particular, they may not experience a sudden decline. This may happen because academics and practitioners struggle to determine whether the observation is an artifact of data, compensation for a new risk factor, the result of market frictions, or a truly attractive investment opportunity that can be taken advantage of. Furthermore, even if money managers are convinced of the trading opportunity, they still face frictions in implementing the strategy on a large scale as they gather data, build models, satisfy prudence requirements, and possibly market the new strategy to their clients.

We argue that such gradual decay is an important feature of many anomalies and propose a modeling framework that allows researchers to test for it. Specifically, we propose that the magnitude of the anomaly is constant until it is discovered at some time τ, after which it declines geometrically toward zero at a rate determined by δ. While traditional descriptions of anomalies are typically limited to a single parameter (e.g., its Jensen's alpha), we argue that anomalies may be better described by a triple of parameters: the initial strength of the anomaly (e.g., α), the time the data suggest the anomaly was discovered (τ), and the speed of the disappearance (δ). For example, if the anomalous behavior generated a nonzero Jensen's alpha, its time variation would be captured as

$$\alpha_t = \alpha\, \delta^{(t-\tau)^+}, \tag{1}$$

where $x^+ = \max\{x, 0\}$. As we show in our article, this framework is flexible enough to accommodate patterns not only in the mean return, but also in covariances, which makes it suitable also for phenomena such as return predictability.

This framework allows us to answer several types of questions. First, what does the evolution of the anomaly suggest about its underlying causes? Although our model is simple, it nests several economically motivated special cases that can be used to test for the anomaly's likely drivers. For example, an abrupt disappearance ($\delta \approx 0$) immediately after or prior to the end of the sample considered by the original study makes a data mining explanation more likely. An anomaly that does not decline ($\delta = 1$) may be more likely to be explained by a systematic risk factor currently outside the model. An anomaly that declines gradually suggests a market inefficiency, and the speed of its decline may indicate the severity of limits to arbitrage. The time of discovery, τ, may also be informative. An estimate of τ close to the publication of the first study documenting the anomaly suggests that the dissemination of academic research leads to improved market efficiency, as hypothesized by McLean and Pontiff (2016); τ close to an institutional change (e.g., the opening of the futures market) may indicate lessening of a trading friction or a limit to arbitrage.

We estimate this model within the Bayesian framework for several reasons. Bayesian analysis allows us to perform exact finite sample inference, which is particularly important when the likelihood function is multi-modal, as is the case in some of the settings we consider. Moreover, the Bayesian framework allows us to impose economically motivated priors on model parameters, which previous literature (e.g., Pastor and Stambaugh, 1999) has found useful in reducing extreme portfolio weights implied by a purely data-based approach. Finally, the Bayesian setting easily allows us to incorporate parameter uncertainty into the problem of solving for optimal portfolios, which is vital given our focus on out-of-sample asset allocation.

We apply our framework to two of the most puzzling anomalies in empirical finance: the January effect, identified by Keim (1983) and Reinganum (1983), and short-term index autocorrelation, usually associated with Lo and MacKinlay (1988). Both have been the focus of substantial academic debate that is to some extent unresolved. They are, to varying degrees, difficult to explain from pure risk arguments, and it is possible that at one point they represented attractive investment opportunities.

We estimate that in the 30 years since its discovery, the January effect has gradually declined from about 8% at its peak to a statistically insignificant 2.3% at the end of our sample in December 2011. These findings suggest that the January effect was neither a risk factor nor data mining, but rather a market inefficiency that investors have gradually learned to exploit. Interestingly, our estimates show that it is unlikely that the decline started near the publication dates of Keim (1983) and Reinganum (1983). Instead, most of the posterior probability mass for τ lies in the second half of the 1970s, with the mode at 1976, substantially predating those papers. This estimate coincides with Rozeff and Kinney (1976), the first study we are aware of that discusses any form of January seasonality. This does not prove a causal link between academic research and prices but can perhaps be interpreted as circumstantial evidence.

Stock index autocorrelation, in contrast to the January effect, appears to have disappeared completely as of the end of our sample. Autocorrelations began their decline much sooner for the value-weighted (VW) index than the equally weighted (EW) index, but in both cases they seem to have vanished by the mid-1990s.[1] Their complete elimination suggests that the underlying cause was not risk premia, as argued by Conrad and Kaul (1988), or a similarly deep behavioral bias. The timing of the disappearance further suggests that publication was also not the primary driver of decreasing autocorrelation. We estimate that autocorrelations started to decay around 1970, whereas the first study documenting the anomaly was Hawawini (1980).

Next, we ask whether accounting for a potential decline makes a difference for the investor. We first evaluate the question in a controlled environment using simulated data. We show that accounting for decay has a large effect on portfolio weights. Importantly, it also improves out-of-sample portfolio performance, beating both an approach that does not allow for decay ($\delta = 1$) and an approach that rules out anomalies to begin with ($\alpha = 0$).

We then discuss the out-of-sample portfolio performance for the two anomalies we study. For both the January effect and market return autocorrelations, portfolios that account for disappearance dominate portfolios that do not allow for it, both in terms of Sharpe ratios and realized utilities. The investor who allows for decay also outperforms, especially in terms of Sharpe ratios, an investor who is unaware of or who rules out the existence of the anomaly.

This superior performance is mainly due to a reduction in the weight of the anomalous asset in the investor's portfolio, particularly once the decay of the anomaly is evident. This reduction is the result of a type of shrinkage introduced by our framework that we believe

1 The differences in the decay rates across anomalies suggest the importance of transaction costs. Our estimates imply that the January effect, largely limited to small-cap stocks, has a half-life of 21 years, compared with 4.4 years for EW index autocorrelations and 10 months for VW index autocorrelations.

is not present in the existing literature on Bayesian analysis and portfolio choice. In existing work, priors shrink mean asset returns to a fixed value (e.g., Jorion, 1986; Kandel and Stambaugh, 1996), to values implied by economic theory (e.g., Pastor and Stambaugh, 1999; Pastor, 2000; Jones and Shanken, 2005), to the means of related benchmark assets (Pastor and Stambaugh, 2002), or to values consistent with reasonable portfolio weights (Tu and Zhou, 2010). In almost all cases, however, those values are assumed to be constant over time.[2] In addition to allowing for time variation that is gradual, our work also adds to this literature in its consideration of out-of-sample performance. Our results complement studies such as Avramov (2004), Busse and Irvine (2006), Tu and Zhou (2010), and Johannes, Korteweg, and Polson (2014), which also present impressive results highlighting different strengths of the Bayesian approach in other contexts.

An additional contribution is to demonstrate that economic theory-based priors can have counterintuitive effects in the presence of a declining anomaly. Such priors have been proposed in part to alleviate the problem of extreme weights often generated by portfolio optimizers and to tilt allocations toward reasonable theory-based benchmarks, such as the market portfolio. We show, for the January effect, that when an anomaly declines, CAPM-based priors may actually lead to more aggressive allocations to the anomaly. Intuitively, priors that shrink the initial magnitude of the January alpha toward zero cause us to infer that its decline started later and that the anomaly persists for longer. At the end of the sample, the investor with such a prior turns out to have a higher predictive mean for future January returns and actually invests more in the January spread portfolio.

Our work is related to papers documenting the disappearance of various anomalous return patterns. Watts (1978) conjectures that the shrinking of abnormal reactions to earnings announcements might have been due to learning. Mittoo and Thompson (1990), McQueen and Thorley (1997), Schwert (2003), and McLean and Pontiff (2016) examine more directly whether published articles provide a source for learning by the market. McLean and Pontiff (2016), in the most comprehensive study among the group, conclude that publication reduces abnormal returns by about a third. Our work differs in that we propose a methodology that explicitly allows for gradual decay of the anomaly and show that it empirically works better than previous approaches (e.g., adopting a discrete break in the level of the anomaly) both in terms of model fit and portfolio choice. As argued above, our framework can also offer insight into potential drivers of the anomaly and the mechanism that causes it to disappear.

2. Examples of Implementation

2.1 The January Effect

> The January effect persisted for a long time because no one was paying attention to it. Then it became just the talk of everybody.
>
> Robert Shiller

> Still, you can't say that anything has changed. The plot just shows that there's been more variability in the last five years or so.
>
> Donald Keim

2 An exception is Pastor and Stambaugh (2001), who allow structural breaks in the market return while imposing a prior that puts more weight on breaks that do not radically change the market risk premium or Sharpe ratio.

I think it was all chance to begin with. There are strange things in any body of data.

Eugene Fama[3]

The January effect refers to the tendency of small capitalization firms to outperform large capitalization firms in the month of January. Observations related to this anomaly were first made by Rozeff and Kinney (1976), who noted higher returns on an EW stock index in the month of January, and by Banz (1981), who identified a relation between size and risk-adjusted equity returns. These results were refined by Keim (1983) and Reinganum (1983), who showed that January effect and size effect were highly interrelated.[4]

More recent evidence on the effect is mixed. For instance, Haugen and Jorion (1996) maintain that the January effect has shown no evidence of dissipating and that no significant trend portends its eventual disappearance. In contrast, Schwert (2003) documents that the January effect has lessened, but it has not disappeared completely. In light of the potential instability of the January effect, the anomaly is an interesting phenomenon to study in the framework we propose.

Following most papers on the topic, we work with monthly returns on the spread portfolio that goes long in a portfolio of small stocks and shorts a portfolio of large stocks. As in Reinganum (1983), we choose the lowest capitalization decile of the NYSE and AMEX exchanges as the former portfolio and the highest capitalization decile as the latter, and start our sample in July 1962. The sample ends in December 2011. We model the excess market return, $R_{\mathrm{em},t}$, and the January spread portfolio return, $R_{\mathrm{spr},t}$, as follows:

$$\begin{aligned} R_{\mathrm{em},t} &\sim N\left(\mu_m, \sigma_m^2\right) \\ R_{\mathrm{spr},t} &= \alpha_0 + \alpha_1 \, I_J(t) \, \delta^{([t-\tau]/12)^+} + \beta R_{\mathrm{em},t} + \epsilon_t, \end{aligned} \tag{2}$$

where $x^+ = \max\{x, 0\}$, $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, and $I_J(t)$ is an indicator that takes value 1 in Januaries and 0 in all other months.

A January effect is present if $\alpha_1 \neq 0$. If the effect exists, then τ is the last period in which it existed in full force and δ determines the speed of its decay. The exponent of δ is the number of years elapsed since the anomaly started to disappear.[5] When $\alpha_1 = 0$ the spread portfolio does not exhibit any January seasonality but can exhibit a size effect, and when $\alpha_0 = \alpha_1 = 0$ the expected return on the spread portfolio conforms with the CAPM.[6]

3 All quotes were taken from "Early January: The Storied Effect on Small-Cap Stocks," by James H. Smalhout, *Barrons*, December 11, 2000.

4 Arguably, the size effect is limited to the January effect in the most recent decades. In our 1962–2011 sample, the Fama–French SMB factor returned 2% per month (t-stat of 4) in Januaries, but only 8 basis points (t-stat of 0.6) outside of Januaries.

5 We assume that the anomaly, once discovered, will decay geometrically. This assumption is a parsimonious way to capture the economic intuition that anomalies should eventually be diversified completely and that disappearance proceeds at a decreasing rate (the most obvious mispricing may be eliminated more quickly, but frictions may delay further decay). Extending these assumptions is straightforward in the framework we propose here.

6 A more complete model would allow seasonality in the market's expected return as well, and possibly introduce seasonal effects in all other parameters of the model. We make these implicit simplifying assumptions because it has been suggested (e.g., Reinganum, 1983) that the January effect is limited to small stocks and is not noticeable in the VW market portfolio. Moreover, our sample contains relatively few Januaries, so the estimation of a more complex model is probably unrealistic.

A possible extension of our specification would be to include a price jump that occurs when the anomaly is discovered. As market participants realize that small cap stocks tend to be underpriced relative to large cap stocks, they will drive up their prices before the January of year τ. However, it is not clear how to accommodate this effect. While the CAPM alpha of the spread portfolio should increase prior to January, that increase may occur at any time during the previous year. Moreover, in line with the idea that anomalies may dissipate gradually, further price increases may also occur after the January effect is discovered. This effect will likely lead to relatively higher estimates of α_0.

2.1.a. Framework for Bayesian estimation

We estimate our model in the Bayesian framework.[7] We consider several prior distributions for the model parameters of the form

$$p(\mu_m, \sigma_m, \alpha_0, \alpha_1, \beta, \delta, \tau, \sigma_\epsilon) \propto p(\alpha_0, \alpha_1)p(\delta, \tau)/(\sigma_\epsilon \sigma_m). \tag{3}$$

That is, the priors on μ_m, β, σ_m, and σ_ϵ are "flat" and independent of all remaining parameters. For the no-decay ($\delta = 1$) and no-anomaly ($\alpha_1 = 0$) specifications, $p(\delta, \tau)$ is eliminated. For the no-anomaly specification, $p(\alpha_0)$ replaces $p(\alpha_0, \alpha_1)$.

For the full model, the prior on δ and τ, $p(\delta, \tau)$, incorporates uncertainty about whether the anomaly has begun to decline as of the end of the sample. Somewhat arbitrarily, we use a prior that reflects a 50% probability that the anomaly has not decayed at all, which we represent as a point mass on $\delta = 1$.[8] Conditional on $\delta \neq 1$, the prior on δ is uniform on [0,1) and the prior on τ is uniform over the set of all years in the sample.

For each model, we consider three different priors for α_0 and α_1. The first is the "diffuse" prior, under which $p(\alpha_0, \alpha_1) \propto 1$. The others are informative CAPM-based priors proposed by Pastor and Stambaugh (1999) and Pastor (2000). Namely, the priors on α_0 and α_1 are independent normal with mean zero and standard deviations of either 0.01 or 0.02. These priors shrink α_0 and α_1 toward zero, so that the process governing $R_{\text{spr},t}$ should be closer to that implied by the CAPM.

In all cases, the posterior distribution is computed using the Gibbs sampler, a Markov chain Monte Carlo approach developed in Geman and Geman (1984).[9] Given δ, τ, and σ_ϵ, the "regression" parameters α_0, α_1, and β have a multivariate normal distribution. The draw of σ_ϵ is from the inverse gamma distribution. The parameters δ and τ are individually drawn, conditional on all other parameters, using the griddy Gibbs sampler.[10] Finally,

7 As a preliminary step, we estimated our proposed model using maximum likelihood and compared it with a number of other specifications. We found strong evidence that the January anomaly slowly disappears, with the likelihood ratio test rejecting plausible alternatives with the p-value of 0.4% or lower. Estimation details are available on request.

8 Equivalently, we can represent the lack of decay with a τ that is later than end of the sample rather than $\delta = 1$. It is also straightforward to incorporate any other prior probability of disappearance. The value we chose here (50%) seems reasonable in the absence of a natural economically motivated prior, particularly since the data seem relatively informative: We have estimated the models with 100% prior weight on disappearance and obtained similar results.

9 We discard the first 5,000 iterations of the Gibbs chain and retain every 100th draw afterward until we have a sample of 10,000 draws. Sample moments computed from these draws estimate the corresponding posterior moments.

10 For δ, we use a 1000-point grid on the interval [0,1]. The grid of τ consists of all Januaries in the sample.

because of prior independence, the posterior distributions of μ_m and σ_m are Student-*t* and inverted gamma, respectively.

2.1.b. Posterior summary under diffuse priors

Table I presents the estimation results, in the form of posterior modes and 95% highest posterior density (HPD) intervals, for the full model and for the special cases of no-decay ($\delta = 1$) and no-anomaly ($\alpha_1 = 0$), all under diffuse (non-informative) priors.[11]

For the full model, we find the posterior probability that decay has begun to be 78%. That is, there is a 22% posterior probability that the January effect remains at full strength as of the end of our sample in 2011. The estimated decay parameter, δ, and decline start date, τ, are 0.97 and 1976, respectively. The initial level of the anomaly, α_1, is estimated at 7.3%. Overall, the mean return for the first out-of-sample January (2012) is estimated at 2.3%, a decline of about two-thirds from the pre-discovery level.

As discussed in Section 1, one advantage of the Bayesian approach is the exact finite sample inference it offers, even for discrete parameters such as τ. This is illustrated in Figure 1, which plots the posterior distribution of τ under diffuse priors (conditional on decay having begun). The posterior is clearly bimodal, with the primary mode in 1976 and a secondary mode in the early 2000s, driven by the high January returns of 2000 and 2001.

It is possible that bimodality of τ reflects "multiple discoveries" of the anomaly.[12] When the anomaly is first incorporated into investors' portfolios, it gradually declines and becomes less attractive. Over time, investors chasing returns may focus on other strategies, consistent with work showing that portfolio managers respond to fashions (e.g., Cooper, Dimitrov, and Rau, 2001). As long as the underlying reasons for the anomaly (e.g., tax loss selling) persist, this may lead to a rebound in the anomaly's strength, and the increasing returns may eventually attract investors back. While the evidence we present here is consistent with such behavior, without further data to support it this explanation is only speculative.

The second panel of Figure 1 presents the evolution of the January mean return over the sample period and the next few out-of-sample years based on the full-sample posterior. While there is a substantial amount of estimation uncertainty, a downward path is clearly evident. The estimates are fairly constant at about 8% until the mid-1970s, when the decline likely started, and then decrease steadily until about 2000. At that stage, past the second estimated peak of τ, the decline speeds up and the mean return drops to 3.4% in the last in-sample January. The first out-of-sample mean estimate is 3.1% for January 2012 (the mode of the posterior distribution is 2.3%, as reported in Table I). Extrapolating the trend forward, the effect would appear likely to survive for some time further.

Table I also presents the estimation results for the two restricted specifications, no-decay and no-anomaly. The former provides a more optimistic view of the anomaly, with the first out-of-sample estimate of 7.1%, more than twice above the corresponding estimate from the full model.

Interestingly, the no-decay estimate of the January alpha, α_1, is lower than that of the full model (6.7% versus 7.3%). This happens because the high January returns early on in the sample are averaged with the lower returns in the second half of the sample. This point

11 A 95% HPD is the shortest interval containing 95% of the mass of the posterior distribution.

12 We note that the modeling framework we propose here is easily extended to multiple "discovery points."

Table I. Bayesian estimates of the January anomaly

The table presents modes and 95% confidence intervals (shortest intervals that cover 95% of the estimated posterior density) for different versions of the model. Posteriors for δ and τ are conditional on decay having started. The full version of the model is presented under both the diffuse and CAPM-based priors; for the two restricted versions, only diffuse prior estimates are exhibited. The last line is the posterior probability that the anomaly has declined (where the prior probability is 0.5).

	Full model	No-decay	No-anomaly	CAPM 2%	CAPM 1%
α_0	−0.004	−0.005	0.001	−0.005	−0.001
	(−0.008, −0.000)	(−0.008, −0.000)	(−0.003, 0.006)	(−0.007, 0.001)	(−0.006, 0.002)
α_1	0.073	0.067		0.061	0.043
	(0.058, 0.115)	(0.054, 0.082)		(0.047, 0.077)	(0.034, 0.057)
δ	0.970			0.977	0.995
	(0.253, 0.998)			(0.083, 0.999)	(0.075, 0.999)
τ	1976			2010	2010
	(1967, 2011)			(1966, 2011)	(1965, 2011)
β	0.244	0.256	0.262	0.269	0.263
	(0.159, 0.328)	(0.165, 0.331)	(0.184, 0.363)	(0.168, 0.338)	(0.173, 0.338)
σ_ϵ	0.047	0.047	0.050	0.047	0.048
	(0.044, 0.050)	(0.044, 0.050)	(0.048, 0.054)	(0.045, 0.050)	(0.045, 0.050)
$E[R_{spr,2012}]$	0.023	0.071	−0.002	0.058	0.044
	(−0.004, 0.071)	(0.055, 0.112)	(−0.007, 0.001)	(−0.004, 0.067)	(−0.001, 0.054)
μ_m			0.005		
			(0.001, 0.008)		
σ_m			0.046		
			(0.043, 0.048)		
P(decline)	0.780			0.411	0.214

is worth stressing. Papers that describe a new anomaly may actually underestimate its level unless they account for its potential decline prior to their sample end. Ironically, this also means that since this lower estimate is not allowed to decay, such papers may at the same time overestimate out-of-sample predictions.

Finally, under the no-anomaly prior the spread portfolio is allowed to exhibit a size premium but not a January effect. This prior leads to economically and statistically small estimated alphas.

2.1.c. The effect of economic theory-based priors

One potential drawback of the analysis above is that it may lead (and, as we show below, indeed leads) to very large portfolio weights. Such extreme weights present an implementation challenge and are perhaps economically unrealistic. Consequently, a number of ways have been suggested to alleviate this issue. Perhaps the most straightforward is to impose *ex ante* constraints. A more elegant approach is to shrink optimization inputs toward values implied by economic theory and consequently tilt the implied portfolio toward a well-accepted benchmark. In the Bayesian context, one way to do so has been proposed in Pastor and Stambaugh (1999) and Pastor (2000). They recommend that the prior for alphas (mispricing) be centered at zero, in line with a preferred asset pricing model, and that the strength of the belief in the model be reflected in the prior standard deviation of alpha.

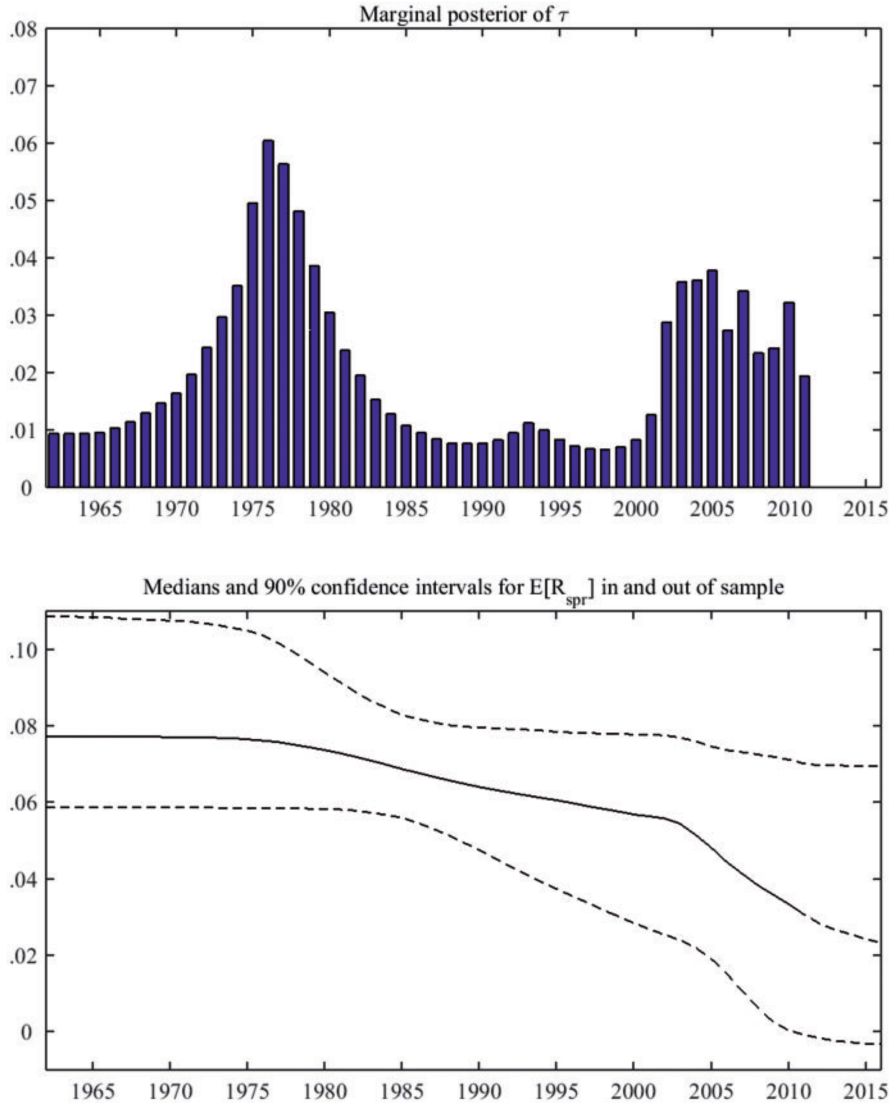

Figure 1. Posterior summary for the January anomaly. This figure presents the posterior for τ (the time of the discovery of the anomaly) conditional on decay having begun as well as the mean January returns estimated for each sample year. Results were obtained under diffuse priors using the whole sample period.

Thus, we now repeat our analysis with the "2% CAPM" prior, in which $\alpha_0, \alpha_1 \sim N(0, 0.02^2)$ and the "1% CAPM" prior, where $\alpha_0, \alpha_1 \sim N(0, 0.01^2)$.

The two rightmost columns of Table I present the full model estimates under these CAPM priors. As the prior belief about α_1 becomes tighter around zero, the posterior naturally shrinks toward zero as well. This is not the only effect of imposing this view, however. Tighter priors for α_0 and α_1 also lead to much later estimates for τ and higher estimates for the persistence of the anomaly, δ. Consequently, the mean January return predicted for

2012, 5.8% for the 2% CAPM prior and 4.4% for the more conservative 1% CAPM prior, are both substantially larger than the 2.3% implied by the diffuse prior.

This effect arises because the CAPM-based priors depress the initial level of the anomaly, which must then persist for longer to be consistent with the data. Suppose, for instance, that under diffuse priors we estimate the initial level of the anomaly at $\alpha_1 = 8\%$ and the end of sample level at $\alpha_1 = 3\%$. If the CAPM prior shrinks the initial level from 8% to 3%, the data will suggest that there was no decline in the sample. This may reduce estimates of δ so substantially as to more than offset the initial shrinkage of α_1 and lead to higher out-of-sample predictions. An implication of this result is that a prior that displays skepticism toward the initial level of anomalies and that generates more in-sample shrinkage may actually lead to higher return forecasts at the end of the sample.

2.2 Short-horizon Autocorrelations in Equity Index Returns

> … we learned that over the past decade several investment firms—most notably, Morgan Stanley and D.E. Shaw—have been engaged in high-frequency equity trading strategies specifically designed to take advantage of the kind of patterns we uncovered in 1988.
>
> Lo and MacKinlay (1999)

While short-horizon autocorrelations in individual equity returns had been noticed as far back as Fama (1965), the extremely high short-term autocorrelations of diversified equity indices were generally unknown until Lo and MacKinlay (1988).[13] Their results imply weekly return autocorrelations as high as 15% for the VW and 30% for the EW CRSP index. While not a violation of market efficiency *per se*, autocorrelations of this magnitude are viewed as anomalous in light of the dominant view in the finance literature, summarized by Ahn *et al*. (2002), that "time variation in expected returns is not a high-frequency phenomenon."

While the analysis within Lo and MacKinlay (1988) suggests that index autocorrelations may have begun to decline by the end of their sample, it is difficult to find more recent comparable evidence. For example, Ahn *et al*. (2002) examine various US index and futures returns since 1982 and find daily autocorrelations ranging from −9% to +22%, albeit with large standard errors.

2.2.a. Modeling disappearance

As with the January effect, we model the potential disappearance of index autocorrelations using geometric decay. Since the anomaly affects covariances rather than means, we need to rewrite our model (2). Specifically, the market return R_t is described by

$$R_t - \mu = \rho\, \delta^{(t-\tau)^+} (R_{t-1} - \mu) + \epsilon_t, \tag{4}$$

where $x^+ = \max\{x, 0\}$.

The model implies that returns display a first-order autocorrelation of ρ up until and including date τ. After that time, the autocorrelation disappears at a rate determined by δ. The extreme values $\delta = 0$ and $\delta = 1$ again correspond to the cases of immediate and no disappearance.

13 Interestingly, a similar finding was obtained by Hawawini (1980), but it was not the primary focus of his paper and appears to have been generally overlooked in the academic literature.

Initially, we assume $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. However, in light of the well-known heteroskedasticity of weekly market returns, we also pursue a generalization in which volatility is stochastic and returns, conditional on volatility, are distributed as Student-*t*. As discussed in Tu and Zhou (2004), the normality approximation is substantially worse for weekly returns than it is for daily returns, and realistic modeling of the distribution's tails is necessary for obtaining accurate portfolio weights. The combination of stochastic volatility and Student-*t* errors allows for a variety of departures from normality that are well documented in the literature. In this case, we replace the assumption that $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ with $\epsilon_t \sim \mathrm{t}(0, h_t, \nu)$, where the log variance process is modeled as

$$\ln h_t = a + b \ln h_{t-1} + cR_{t-1} + \eta_t \tag{5}$$

and $\eta_t \sim N\left(0, \sigma_\eta^2\right)$. The parameter b measures the persistence of log variance, while c allows for the possibility of a leverage effect, or a negative correlation between returns and volatilities. ν, the Student-*t* degrees of freedom parameter, measures the degree of leptokurtosis in returns conditional on volatility.

As before, we also consider restricted versions of this specification. The no-anomaly models (one with constant, one with stochastic volatility) impose the restriction that $\rho = 0$ (making δ and τ irrelevant). The no-decay models set $\delta = 1$ (making τ irrelevant). Finally, we consider an investor who believes that the date of discovery was April 1, 1988, approximately when Lo and MacKinlay (1988) was published. This investor is said to follow a "Lo and MacKinlay" model.

We consider relatively diffuse priors on the parameters in an attempt to represent prior ignorance. As before, our priors allow for a 50% probability that the anomaly has not yet begun to disappear as of sample end (time T). Conditional on decay having begun, the prior on δ is flat on [0,1), and the prior on τ is uniform over the set $\{1, 2, ..., T\}$. Priors on remaining parameters are independent of δ and τ and are given by

$$p(\mu, \rho, \sigma_\epsilon) \propto 1/\sigma_\epsilon \tag{6}$$

for the constant volatility model and by

$$p(\mu, \rho, a, b, c, \sigma_\eta, \nu) \propto \lambda \exp(-\lambda\nu)/\sigma_\eta \tag{7}$$

for the stochastic volatility model. The parameter λ is set to 0.05, a value that implies a relatively diffuse prior distribution for ν, with a mean of about 18 and a standard deviation of about 12.

All models are estimated using the Gibbs sampler, as before. For constant volatility specifications, conditional distributions are obtained for each parameter conditional on all the rest. The conditional distributions of μ and ρ are each Gaussian, while σ_ϵ is inverted gamma. The remaining parameters, δ and τ, are drawn by discretizing the parameter space, as in Section 2.1.

Our approach to estimation of the Student-*t* stochastic volatility model combines the algorithms of Jacquier, Polson, and Rossi (1994) and Geweke (1993). In essence, we augment observed price data with unobserved volatility data. Conditional on the augmented data set, estimation proceeds similarly to the above method, relying on the properties of the Gaussian augmented likelihood. We describe this procedure in more detail in Appendix A.

2.2.b. Data

Following Lo and MacKinlay (1988), we work with weekly returns on value- and EW CRSP indexes. To minimize biases induced by nontrading and other microstructure effects, we examine indexes based only on stocks in the S&P 500.[14] Our sample starts in January of 1953, shortly after the end of Saturday trading on the NYSE, and ends in December of 2011. Weekly returns are computed by geometrically compounding daily returns from one Wednesday to the following Wednesday. If a Wednesday return is missing, Thursday's return is used instead; if the Thursday return is also missing, then the Tuesday return is used. The only missing week is from September 11, 2001 to September 16, 2001, when trading was suspended due to the events of September 11. Our sample therefore consists of 3,077 weeks.

Table II reports sample autocorrelations for our entire sample and three subsamples. For both EW and VW indexes, autocorrelations are strong in the first third of the sample period, with values of 0.21 and 0.15, respectively. In the middle third of the sample, the autocorrelation of the EW index is about half of its original level, while autocorrelation in the VW index has disappeared. In the final third, both indexes display slightly negative but insignificant autocorrelations.

2.2.c. Posterior summaries

The results in Table II suggest that weekly autocorrelations have diminished over time. We refine this result by analyzing the model in Equation (4) using the Bayesian methods described above. The posterior modes and 95% HPD intervals for the model parameters are shown in Table III. The μ and σ parameters are annualized for easier interpretation. As before, we report results for the δ and τ parameters for the full model conditional on decay having begun. For both indexes and for both volatility specifications, the probability that decay has started by the end of the sample is essentially 100%.

Panel A of Table III was obtained under the assumption of constant volatility. For the EW index, the mode of ρ differs significantly across specifications. It ranges from about 30% under the Lo and MacKinlay (LM) model to 5% under the no-decay model. The full model produces an intermediate but still sizable value of 15%. For both the full and LM models, the posterior mode of δ is close to one, suggesting extreme persistence or near-zero decay. However, this parameter is inaccurately estimated for the full model, and values close to zero are within the 95% HPD interval.

For the VW index, results are somewhat different. For the full model, the posterior of δ now has a mode of 0.486 rather than 0.997, though that parameter remains very uncertain. The initial autocorrelation ρ is, surprisingly, about the same for the VW index as it is for the EW index, though under the LM model it is somewhat lower. For both models, the posterior of ρ is much less precise for the VW index as it was for the EW index.

Panel B of Table III extends the model to the fat-tailed stochastic volatility process proposed in Equation (5). Posteriors of the stochastic volatility parameters a, b, c, and σ_η are fairly typical of those found in the literature. Values of b near 1 indicate that volatility is highly persistent, while negative values of c are consistent with a leverage effect. Average levels of volatility, implied by a and b, are consistent with the unconditional estimates of Panel A. The parameter ν, which represents the degrees of freedom in the Student-t-distributed residual, is centered between 27 and 36, indicating that stochastic volatility is responsible for almost all of the leptokurtosis in weekly returns.

14 Following Fisher (1966), Ahn *et al.* (2002) argue that positive index autocorrelations may be at least partly spurious as the result of nontrading.

Table II. Equity index sample autocorrelations

This table reports sample autocorrelations of weekly returns on EW and VW indexes over four different sample periods. Indexes are constructed from all stocks in the S&P 500. OLS standard errors are in parentheses. Newey–West standard errors are in brackets and are computed using five lags.

	EW index	VW index
1953–2011	0.055	−0.004
	(0.018)	(0.018)
	[0.028]	[0.027]
1953–1972	0.210	0.148
	(0.030)	(0.031)
	[0.040]	[0.036]
1973–1992	0.102	0.001
	(0.031)	(0.031)
	[0.043]	[0.043]
1993–2011	−0.047	−0.079
	(0.032)	(0.032)
	[0.048]	[0.047]

Under stochastic volatility, the autocorrelation results for the EW index are similar to the constant volatility case. Initial autocorrelations remain large for the full and LM models, even for the VW index. In addition, there continues to be much uncertainty about the decay parameter δ, particularly under for the full model and for the VW index.

Because of their asymmetry and multimodality, the posteriors for τ are in some cases inadequately described by modes and 95% HPD intervals. Therefore, in Figure 2, we display histograms of the posterior distribution of τ for the full model. Under constant volatility (top left panel), autocorrelation decay most likely began during the 1980s, but τ is estimated quite imprecisely and values as low as 1975 and as high as 1995 receive some support. When stochastic volatility is introduced in the top right panel, the posterior adds a new mode centered on the mid-1970s.

For the VW index, shown in the bottom two panels, the posterior is relatively consistent between constant and stochastic volatility specifications. In both cases, the posterior of τ has a single mode in the early 1970s. The shape of the posteriors of τ is helpful for understanding the posteriors of other parameters reported in Table III. For instance, Table III reports that the initial autocorrelation parameter ρ estimated for the full model was similar for the EW and VW indexes. This contrasts with the observation from Table II that autocorrelations were historically higher for the EW index. The explanation from Figure 2 is that autocorrelation appears to have begun disappearing much earlier for the VW index, so the high pre-decay level for the VW index describes the data in a much earlier period, when both markets displayed more serial dependence. Similarly, we observed that the addition of stochastic volatility caused the posterior of τ to shift to the left for the EW index. Ending the pre-decay period earlier again increases the magnitude of the original autocorrelation ρ.

Finally, it is notable that in all cases, but especially for the VW index, there is a very high posterior probability that τ substantially predates the Lo and MacKinlay (1988)

Table III. Bayesian estimates of the autocorrelation model under constant volatility and stochastic volatility

In Panel A, the table reports posterior modes and 95% HPD intervals (in parentheses) for the model $R_t - \mu = \rho\delta^{(t-\tau)^+}(R_{t-1} - \mu) + \epsilon_t$, where R_t denotes the market return, $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, and $x^+ = \max\{x, 0\}$. Results are presented for the full model and three restricted versions. In the Lo and MacKinlay specification, τ is fixed at April 1, 1988. In the no-decay specification, $\delta = 1$ and τ is undefined. In the no-anomaly specification, $\rho = 0$ and both δ and τ are undefined. The last line is the posterior probability that the anomaly has declined (where the prior probability is 0.5). Reported posteriors for δ and τ are conditional on decay having started. All models are estimated using weekly data over the 1953–2011 sample period. In Panel B, the table reports posterior modes and 95% HPD intervals (in parentheses) for the model $R_t - \mu = \rho\delta^{(t-\tau)^+}(R_{t-1} - \mu) + \epsilon_t$, where R_t denotes the market return, $\epsilon_t \sim t(0, h_t, \nu)$, and $x^+ = \max\{x, 0\}$. The stochastic volatility process is $\ln h_t = a + b \ln h_{t-1} + cR_{t-1} + \eta_t$, where $\eta_t \sim N\left(0, \sigma_\eta^2\right)$. Results are presented for the full model and three restricted versions. In the Lo and MacKinlay specification, τ is fixed at April 1, 1988. In the no-decay specification, $\delta = 1$ and τ is undefined. In the no-anomaly specification, $\rho = 0$ and both δ and τ are undefined. The last line is the posterior probability that the anomaly has declined (where the prior probability is 0.5). Reported posteriors for δ and τ are conditional on decay having started. All models are estimated using weekly data over the 1953–2011 sample period.

Panel A. Bayesian estimates of the autocorrelation model under constant volatility

	Full model	Lo and MacKinlay model	No-decay model	No-anomaly model
		EW index		
$\mu \times 52$	0.131	0.137	0.135	0.135
	(0.089, 0.179)	(0.091, 0.181)	(0.091, 0.178)	(0.093, 0.176)
ρ	0.153	0.306	0.053	0
	(0.096, 0.219)	(0.159, 0.479)	(0.020, 0.090)	
δ	0.997	0.999	1	
	(0.050, 0.999)	(0.998, 0.999)		
τ	March 13, 1991	April 1, 1988		
	(June 5, 1974, November 24, 1993)			
$\sigma \times \sqrt{52}$	0.161	0.162	0.161	0.162
	(0.157, 0.165)	(0.158, 0.166)	(0.158, 0.166)	(0.158, 0.166)
P(decline)	1.000			
		VW index		
$\mu \times 52$	0.104	0.110	0.108	0.110
	(0.069, 0.149)	(0.072, 0.148)	(0.071, 0.148)	(0.071, 0.148)
ρ	0.157	0.264	0.003	0
	(0.035, 0.237)	(0.010, 0.952)	(0.000, 0.033)	
δ	0.486	0.998	1	
	(0.046, 0.994)	(0.048, 0.999)		
τ	September 8, 1971	April 1, 1988		
	(May 3, 1961, March 30, 1988)			
$\sigma \times \sqrt{52}$	0.150	0.149	0.150	0.150
	(0.146, 0.153)	(0.146, 0.154)	(0.146, 0.154)	(0.146, 0.154)
P(decline)	0.997			

(continued)

Table III. Continued

Panel B. Bayesian estimates of the autocorrelation model under stochastic volatility

	Diffuse model	Lo and MacKinlay model	No-decay model	No-anomaly model
		EW index		
$\mu \times 52$	0.147	0.149	0.149	0.149
	(0.111, 0.182)	(0.109, 0.180)	(0.112, 0.181)	(0.121, 0.185)
ρ	0.205	0.303	0.091	0
	(0.150, 0.285)	(0.207, 0.423)	(0.053, 0.127)	
δ	0.997	0.999	1	
	(0.051, 0.999)	(0.999, 0.999)		
τ	February 26, 1975	April 1, 1988		
	(November 21, 1973, April 10, 1991)			
a	−0.168	−0.155	−0.161	−0.155
	(−0.248, −0.091)	(−0.247, −0.091)	(−0.248, −0.088)	(−0.243, −0.088)
b	0.978	0.979	0.978	0.979
	(0.968, 0.987)	(0.968, 0.987)	(0.968, 0.988)	(0.968, 0.988)
c	−4.919	−4.925	−4.866	−4.690
	(−5.963, −4.029)	(−5.907, −4.003)	(−5.848, −3.859)	(−5.782, −3.748)
σ_η	0.134	0.131	0.127	0.120
	(0.102, 0.166)	(0.101, 0.169)	(0.100, 0.167)	(0.096, 0.163)
ν	28	34	36	27
	(19, 50)	(20, 50)	(21, 50)	(19, 50)
P(decline)	1.000			
		VW index		
$\mu \times 52$	0.119	0.128	0.128	0.130
	(0.089, 0.155)	(0.093, 0.157)	(0.096, 0.158)	(0.097, 0.158)
ρ	0.174	0.228	0.017	0
	(0.102, 0.236)	(0.003, 0.873)	(0.000, 0.049)	
δ	0.985	0.998	1	
	(0.043, 0.988)	(0.095, 1.000)		
τ	September 13, 1972	April 1, 1988		
	(January 29, 1969, September 24, 1975)			
a	−0.235	−0.236	−0.254	−0.238
	(−0.362, −0.152)	(−0.352, −0.147)	(−0.373, −0.155)	(−0.358, −0.147)
b	0.967	0.970	0.968	0.969
	(0.954, 0.980)	(0.955, 0.981)	(0.952, 0.980)	(0.954, 0.981)
c	−5.195	−5.136	−5.429	−5.248
	(−6.750, −4.088)	(−6.531, −3.980)	(−6.782, −4.176)	(−6.613, −3.908)
σ_η	0.153	0.146	0.146	0.152
	(0.120, 0.188)	(0.114, 0.183)	(0.121, 0.190)	(0.116, 0.185)
ν	34	34	36	27
	(21, 50)	(20, 50)	(21, 50)	(19, 50)
P(*decline*)	1.000			

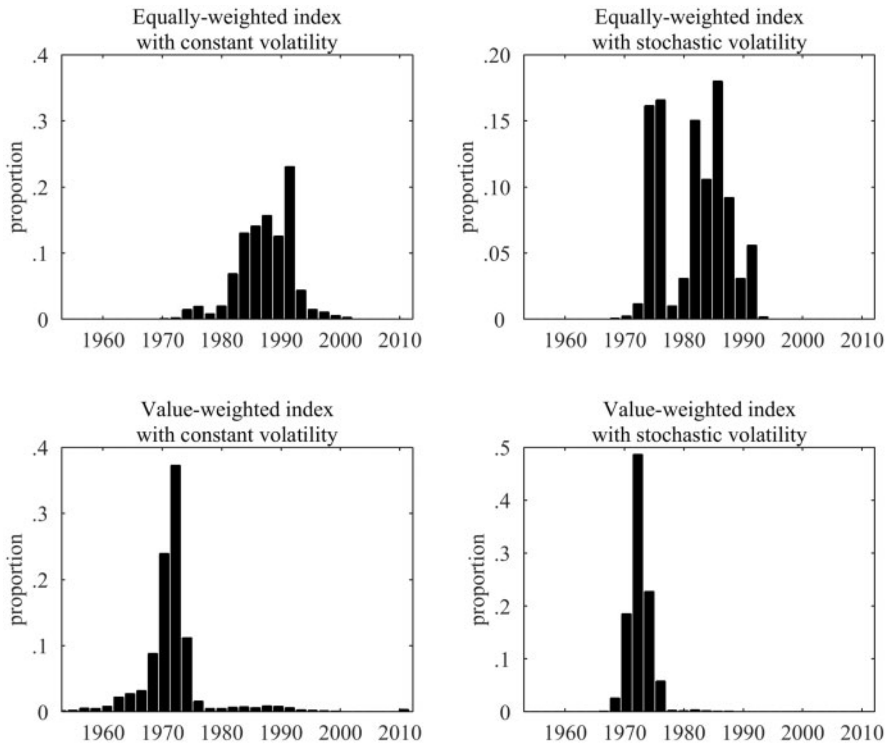

Figure 2. Posterior distributions of τ for disappearing autocorrelations under diffuse priors. This figure presents the posterior for τ (the time of the discovery of the anomaly) conditional on decay having begun for the constant and stochastic volatility specifications for both the EW and VW indexes.

article, which drew the greatest attention to this anomaly. Thus, if we are to attribute the decline in autocorrelations to the high-frequency trading strategies that Lo and MacKinlay (1999) described in the quote at the beginning of the section, then these strategies most likely were well-underway prior to the publication of their original study.

An alternative explanation that is the reduction of autocorrelation in the early 1970s was due to faster information flow, perhaps resulting from the advent of computerization on the floor of the NYSE and the rise of fully computerized brokers such as Instinet (see Blume, Siegel, and Rottenberg, 1993). Arguably, this explanation is more in line with our estimation results. They suggest the autocorrelations declined rather abruptly, which is perhaps more consistent with a "discrete" event such as the introduction of a new computerized broker.

3. Implications for Portfolio Choice

In the preceding section, we have documented that some anomalies decline after discovery, and that the speed of the decline can vary across anomalies. We now evaluate the implications of such dynamics for portfolio choice and the model's out-of-sample performance.

3.1 Evidence from Simulated Data

3.1.a. Simulation setup

Before investigating portfolio choice for the two anomalies we discuss above, we test the model's impact in a controlled setting with simulated data. To keep the simulation as straightforward as possible we consider a generic example of an asset-pricing anomaly, a portfolio that generates an alpha with respect to an asset pricing model:

$$R_t = \alpha\,\delta^{(t-\tau)^+} + \beta\,R_{\mathrm{mkt},t} + \epsilon_t, \tag{8}$$

where R_t is the excess return on the anomaly portfolio, $R_{\mathrm{mkt},t} \sim \mathrm{N}(\mu_{\mathrm{mkt}}, \sigma^2_{\mathrm{mkt}})$ is the excess return on the market portfolio, $\epsilon_t \sim N(0, \sigma^2_\epsilon)$, and $x^+ = \max\{x, 0\}$. As before, the evolution of the anomaly is described by the triple (α, τ, δ), which captures the initial level of the anomaly, the time that it starts to decline, and the speed of decay, respectively.

To evaluate the impact our approach has on portfolio choice, we need to translate estimation results into forward-looking predictive estimates of the mean and the variance of anomaly returns. The anomaly portfolio's predictive mean is computed as

$$\hat{\mu} = \widetilde{\mathrm{E}}[\mathrm{E}[R_{t+1}|\theta]] = \widetilde{\mathrm{E}}\left[\alpha\,\delta^{(t-\tau+1)^+} + \beta\,\mu_{\mathrm{mkt}}\right], \tag{9}$$

where a "tilde" denotes a moment estimated by averaging across all Gibbs draws and where θ denotes all parameters of the model. Predictive variances are calculated using the variance decomposition

$$\begin{aligned}\hat{\sigma}^2 &= \widetilde{\mathrm{E}}[\mathrm{Var}(R_{t+1}|\theta)] + \widetilde{\mathrm{Var}}(\mathrm{E}[R_{t+1}|\theta]) \\ &= \widetilde{\mathrm{E}}\left[\beta^2\sigma^2_{\mathrm{mkt}} + \sigma^2_\epsilon\right] + \widetilde{\mathrm{Var}}\left(\alpha\,\delta^{(t-\tau+1)^+} + \beta\,\mu_{\mathrm{mkt}}\right).\end{aligned}$$

Similar calculations produce the predictive moments of the market portfolio and the predictive covariance between the two assets.

Armed with these estimates we consider a myopic Bayesian investor with quadratic utility,

$$\hat{\mu}_p - \frac{A}{2}\hat{\sigma}^2_p, \tag{10}$$

where $\hat{\mu}_p$ and $\hat{\sigma}_p$ denote the predictive mean and standard deviation of the investor's portfolio return. To keep the model simple and tractable, here and for the January effect we consider two risky assets only, the anomaly portfolio and the market portfolio, as well as the one-month Treasury bill. For return autocorrelations, we simplify the setup to the choice between the market portfolio (either VW or EW) and the risk-free asset. In all our examples, the investor's risk aversion parameter A is set to 10. This choice implies that the investor who observed the entire sample of market returns would allocate approximately 100% of his wealth to the market portfolio.[15]

Finally, based on the above inputs, we compute optimal portfolios implied by our full model, which allows for anomalies' gradual disappearance. We assess the model's performance by comparing this portfolio to those implied by two nested versions of the model: the no-decay specification, which acknowledges the anomaly but does not allow it to

15 Note that A is just a scaling factor for the portfolio weights: as A becomes larger, the weight on the risk-free asset increases, but the composition of the risky assets portfolio remains the same.

disappear, and the no-anomaly specification, which rules out the existence of the anomaly. For each of these portfolios we compute the out-of-sample performance using a gradually extending estimation window. First, we estimate the model using an initial subset of our data. We then form portfolios that are held over the first out-of-sample period and record their returns. We next increase the estimation window by one period and repeat the process, until we reach the end of our sample.

3.1.b. Simulation results

To parametrize Equation (8) we choose $\alpha = 0.25\%$, $\beta = 0$, and $\sigma_\epsilon = 1\%$, which translates into an information ratio of 0.25. In other words, prior to its disappearance, the anomaly is an attractive investment opportunity. We allow the investor to observe the initial 200 periods of the anomaly's evolution (periods -199 to 0) before making any investment. We assume that the anomaly starts to disappear 400 periods after the start of the simulation ($\tau = 200$) and that it decays relatively slowly, at the rate of $\delta = 0.975$. We simulate the anomaly for 200 periods after the start of the decay at time τ, but already 100 periods after τ the anomaly is at just 8% of its original level. The investor may also allocate funds to the market portfolio, which has a mean return μ_{mkt} of 0.5% and a volatility σ_{mkt} of 4.5%.

We first present a summary of the estimation and the quality of out-of-sample results in Figure 3. The figure depicts the true alpha along with the out-of-sample predictions generated by two models. The first, in the left panel, is the full specification that allows for decay. The second, in the right panel, is the no-decay specification that restricts $\delta = 1$.

Not surprisingly, since we simulate the data under the null of decay, the full model does noticeably better in capturing the evolution of true alphas. There is, however, interesting nuance to this observation. Over the first two hundred out-of-sample periods (1–200) the anomaly does not decline. Within that subperiod the restricted model fits the data almost exactly on average. The full model, in contrast, builds in some conservatism. Since we start with a prior that assigns 50% probability to decay and since the data do not completely dominate the prior, the posterior incorporates a significant possibility of decay and leads to the expected returns prediction consistently below the true value. After the data generating model triggers decay, starting in period $\tau = 200$, the situation reverses. The full model recognizes the start of the decay and reduces the predictive mean return essentially to zero over the subsequent 200 periods. In contrast, the no-decay approach keeps the predictive mean relatively unchanged throughout the whole period.

We translate the estimation output into portfolio weights and record the out-of-sample performance of the resulting portfolios. For completeness, we also present the performance of the no-anomaly model, which rules out the anomaly to begin with by imposing $\alpha = 0$. Table IV summarizes the evidence for the full out-of-sample period, as well for the subperiods before and after the start of decay. To compare the performance of the various models we use two measures: the Sharpe ratio, capturing the risk-to-variability tradeoff for each portfolio, and the realized utility measure of Fleming, Kirby, and Ostdiek (2001), which is simply the investor's utility (10) evaluated using the sample moments of the realized out-of-sample portfolio returns.

Over the complete out-of-sample period, the full model dominates the two restricted specifications. The differences are in all cases statistically significant and are particularly pronounced for realized utility. The no-decline model comes close to the full model in terms of Sharpe ratio because of its aggressive allocations to the anomaly before it starts to decay. However, the tendency to be aggressive also causes the no-decay investor to take

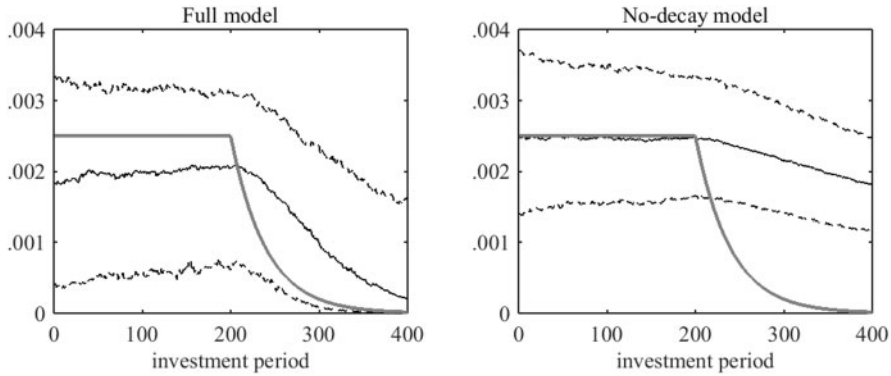

Figure 3. Simulation evidence: true and estimated expected returns. We estimate the disappearing anomalies model using simulated data on a rolling sample basis and compute the mean of the predictive distribution of returns for the next out-of-sample period. The figure presents the true (the thick gray line) and estimated expected returns (the solid black line) along with 90% confidence intervals (the dashed lines).

Table IV. Out-of-sample performance in simulated data

The table presents the performance of the full model (1) and two restricted specifications: one that allows no disappearance (2), and one that allows no anomaly at all (3). As performance measures, the table reports the Sharpe ratio and the realized utility ($U = E(R) - A\sigma^2(R)/2$, with $A = 10$). The results are based on simulations described in Section 3.1, where the anomaly operates at full strength in periods 1–200 and declines throughout periods 201–400. T-statistics are in parentheses.

	(1) Full model	(2) No decay	(3) No anomaly	(1)−(2) Difference	(1)−(3) Difference
		Realized utility (monthly, ×100)			
1–400	0.165	0.120	0.047	0.045	0.117
	(36.24)	(23.96)	(18.17)	(17.37)	(30.89)
1–200	0.300	0.325	0.041	−0.026	0.259
	(34.89)	(35.45)	(10.88)	(−17.48)	(33.15)
201–400	0.035	−0.078	0.054	0.113	−0.019
	(7.40)	(−10.34)	(14.95)	(23.82)	(−6.11)
		Sharpe ratio (monthly)			
1–400	0.180	0.177	0.092	0.003	0.088
	(75.72)	(81.43)	(35.93)	(3.56)	(34.91)
1–200	0.240	0.255	0.082	−0.015	0.158
	(65.20)	(75.12)	(22.01)	(−25.74)	(37.33)
201–400	0.105	0.092	0.100	0.012	0.005
	(33.92)	(29.28)	(30.60)	(8.22)	(1.90)

substantially more risk than an investor who uses the full model, especially after decay has begun, and this leads to lower realized utility.

There are interesting patterns in the two subsamples we consider. As expected, the no-decay model does particularly well before the anomaly starts to decline, slightly outperforming the full model but dominating the no-anomaly specification. Its performance, however, markedly decreases after decay is underway. In this period, judging by the realized utility, *ex post* the investor would have preferred not investing at all and earning no return rather than investing using the no-decay model. The best performer in the latter period depends on the performance metric, with the full model having a slightly higher Sharpe ratio but a slightly lower realized utility than the no-anomaly portfolio.

These results are intuitive: if investors know that the anomaly is in full strength, they should allocate to it aggressively. When the anomaly starts to decay to zero, the most conservative approach eventually becomes the most attractive. Of course, the problem is that investors would not know in which of these two situations they find themselves, so by examining pre- and post-decay periods separately we are effectively conditioning on an unknown. It is therefore more realistic and more relevant to include both the pre-decay and post-decay regimes in the evaluation period. As we see in Table IV, in this case the full model leads to better performance than the two alternatives.

3.2 Investing in January Anomaly

We initiate our out-of-sample analysis of the January anomaly in December 1976, which roughly represents the data available to an investor who read the Rozeff and Kinney (1976) article. Using this sample, we compute the implied allocation for 1977, then redo the estimation at the end of December 1977, etc. The shortest sample therefore contains fifteen Januaries from which the model parameters are estimated.

Figure 4 presents the model-implied predictive mean returns for each out-of-sample January and the corresponding January portfolio weights. As expected, the full model's weight is between the no-decay and no-anomaly allocations. It tends to be closer to the former early on and over time gradually approaches the latter.[16] Interestingly, in the early 1990s and in the early 2000s the full model dramatically increases its allocations, almost to the no-decay level. This rebound results from additional modes for τ, as we noted in Section 2.1.2, and from higher posterior mean returns in the samples ending during those periods. In the full sample, the early 1990s turn out to have been less important, leaving the posterior for τ with two clear modes (Figure 1).

A downward trend in the anomaly weights is visible for the two restricted models as well. The no-decay investor's January weight drops from 371% in 1977 to 280% in 2012. There is also a decline in the no-anomaly investor's allocation, because lower January returns translate into lower α_0 estimates when α_1 is restricted to be zero. The no-anomaly investor's weights decline from about 20% at the beginning of the sample to 5% in 2012. Note that the no-anomaly investor's allocation can be interpreted as an allocation that accounts for the size effect but not for the January effect.

Table V describes the out-of-sample performance of these portfolios. Under diffuse priors on α_0 and α_1, the no-decay portfolio generated the highest average excess returns, 1.56% per month, when compared with 0.91% for the full model and 0.15% for the no-anomaly allocation. However, this performance comes at the cost of very high risk: the

16 When one invokes trading costs, obviously high in the case of the spread portfolio, the allocation may well remain substantial. Assuming the round trip transaction costs of 2%, the full model allocates 42% to the January portfolio at the end of the sample, economically large but perhaps more realistic. The no-decay allocation declines to 198% and thus is still economically rather implausible.

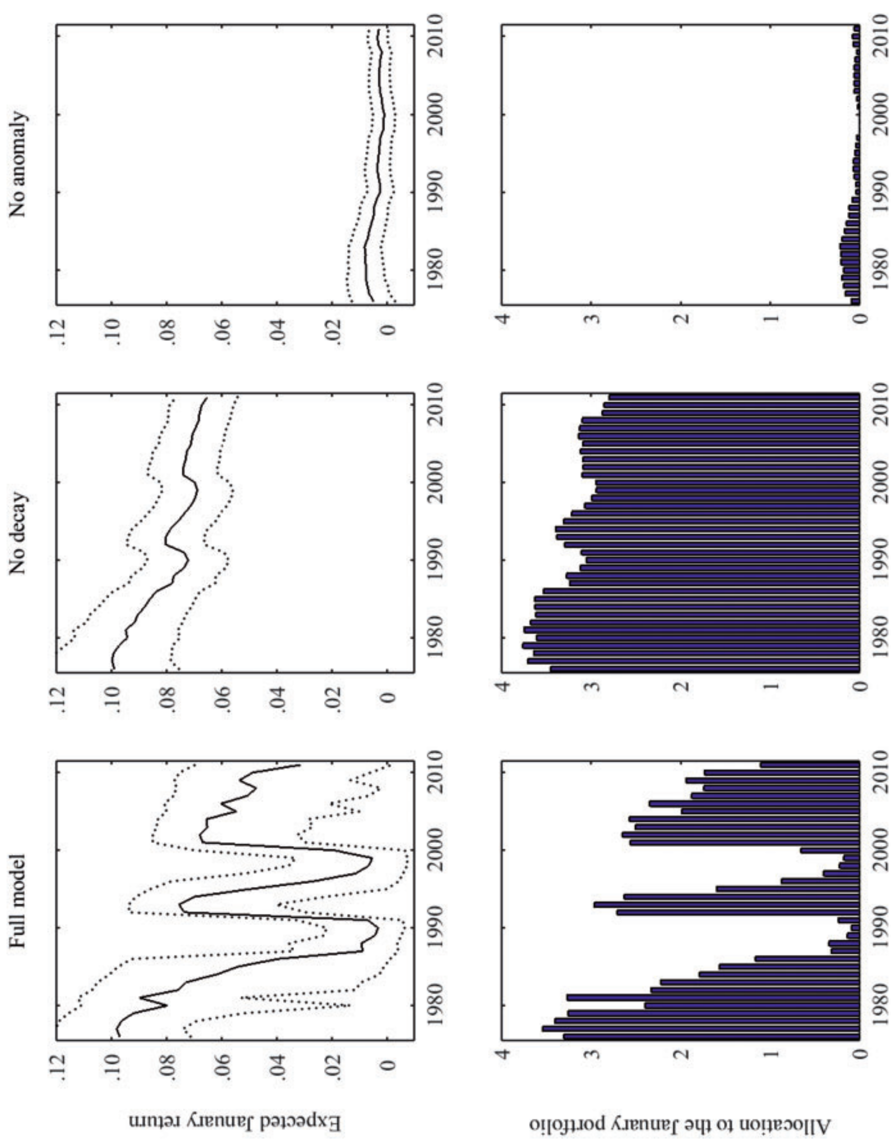

Figure 4. Rolling sample estimates of the January anomaly. The graph presents the predicted out-of-sample January mean returns (top panel) and the corresponding weights on the January spread portfolio (bottom panel). Results were obtained under diffuse priors for the rolling sample ending in the year identified on the *x*-axis.

no-decay standard deviation was 7.45%, almost twice the standard deviation of the full model, 3.94%. Consequently, it is the full model that earns the highest out-of-sample Sharpe ratio, 0.80, and the highest realized utility, 0.13% (1.56% using annualized average returns and standard deviation). The Sharpe ratio of the no-decay model is lower, at 0.73, and the excessive risk of that strategy results in a negative utility, indicating a portfolio that

Table V. Out-of-sample performance of the January effect

This table evaluates out-of-sample performance of allocations implied by the disappearing anomaly model, a model that does not allow for disappearance, and a model that does not allow for the anomaly. For each of these models, the table presents results for three priors (diffuse, 2% CAPM, and 1% CAPM). The first out-of-sample portfolio uses the data up to 1976 to estimate the model. For each sample used to estimate the model, the implied portfolio is held over the subsequent period; the table presents the performance of this portfolio. The table reports the mean and standard deviation of monthly excess returns, the annualized Sharpe ratio, and the annualized realized utility ($U = E(R) - A\sigma^2(R)/2$, with $A = 10$). All except Sharpe ratios are expressed in percentages.

		Average excess return	Standard deviation	Sharpe ratio	Realized utility
		Diffuse prior			
1976–2011	Full model	0.91	3.94	0.80	1.56
	No-decay	1.56	7.45	0.73	−14.64
	No-anomaly	0.15	1.08	0.48	1.08
		CAPM 2% prior			
1976–2011	Full model	0.88	3.67	0.83	2.52
	No-decay	1.28	6.05	0.73	−6.60
	No-anomaly	0.15	1.07	0.48	1.08
		CAPM 1% prior			
1976–2011	Full model	0.72	3.17	0.78	2.64
	No-decay	0.84	3.90	0.75	0.96
	No-anomaly	0.15	1.06	0.48	1.08

is worse (in utility terms) than Treasury bills. The no-anomaly portfolio offers both a lower Sharpe ratio and a lower (but positive) realized utility.[17]

Table V also documents the performance of investors guided by our two CAPM-based priors on α_0 and α_1. As in the diffuse prior case, we observe better out-of-sample performance of the full model over the no-decay and no-anomaly specifications under these theory-based priors. The 2% CAPM (1% CAPM) Sharpe ratios of the full model and no-decay specifications are 0.83 and 0.73 (0.78 and 0.75), respectively. For both CAPM priors, the realized utility measure strongly favors the full model. Comparing full model results under CAPM to those under diffuse priors, we see that CAPM priors generally improve performance. This is consistent with the notion that shrinkage has the biggest potential benefits when sample sizes are small, and that the shrinkage introduced by CAPM-based priors is complementary to that induced by our modeling of decay.

3.3 Investing in Index Autocorrelations

For the market index autocorrelations, we re-estimate each model using samples starting in January 1953 and ending in each week between January 1963 and December 2011. We start by plotting, in Figure 5, the rolling sample posterior means of the terminal autocorrelations $\rho\delta^{(T-\tau)^+}$, where T is the last observation of each rolling sample period.

17 In untabulated tests, available by request, we divide the 36 years of our out-of-sample evidence into three equal-length subperiods and show that the full model outperforms also in subsamples.

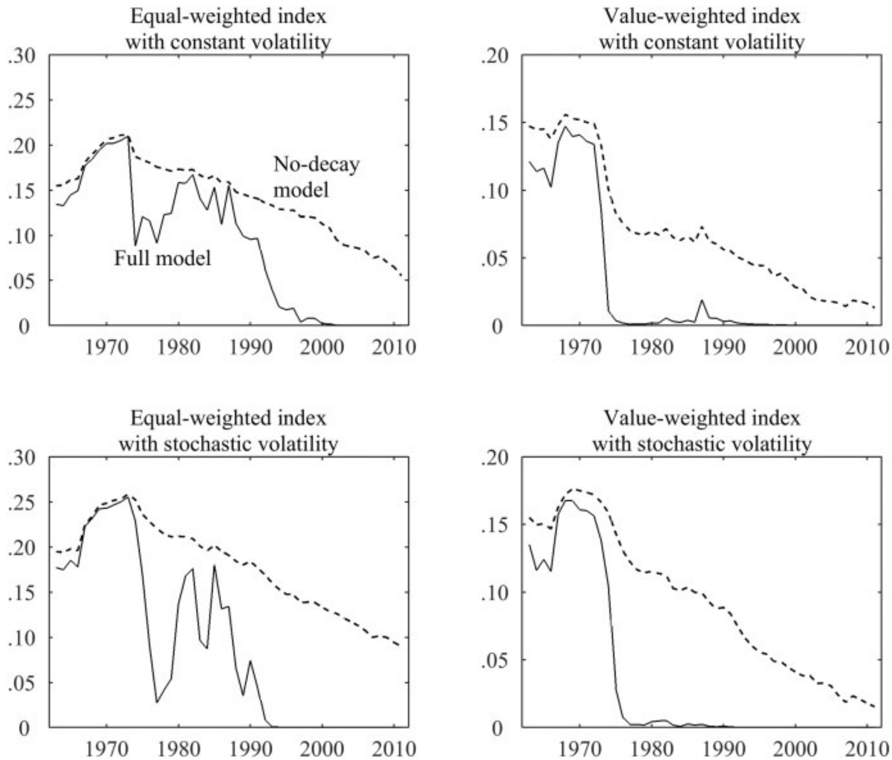

Figure 5. Rolling sample estimates of return autocorrelations. This figure shows the return autocorrelations estimated using rolling samples starting in January 1953 and ending each week between January 1963 and December 2011.

These are the autocorrelations that an investor might have predicted, at least in the very short term, under each model. For visual clarity, we only plot end-of-year values. Since the no-anomaly model sets autocorrelations equal to zero, there are only two lines per panel in Figure 5.

Through the mid-1970s, the full and no-decay models are in rough agreement on the level of return autocorrelation, both for the EW and VW indexes, and both for the constant and stochastic volatility models. Afterward, autocorrelations drop quickly under the full model (the solid line). The mid-1970s therefore appears to provide the first substantial evidence that the anomaly is disappearing. For the EW index, the mean autocorrelation under the full model recovers somewhat during the early to mid-1980s before trending to zero over the subsequent 5–10 years. For the VW index, the mean autocorrelation drops close to zero well before 1980.

In contrast, under the no-decay model (the dashed line), autocorrelations are only gradually trending downward after their peak in the 1970s. The decline becomes more apparent as the post-discovery period becomes a more important part of the rolling sample. Nonetheless, the no-decay model still implies a substantial level of autocorrelation in the EW index by the end of our sample in 2011. For the VW index, the autocorrelation in 2011 is smaller but clearly positive.

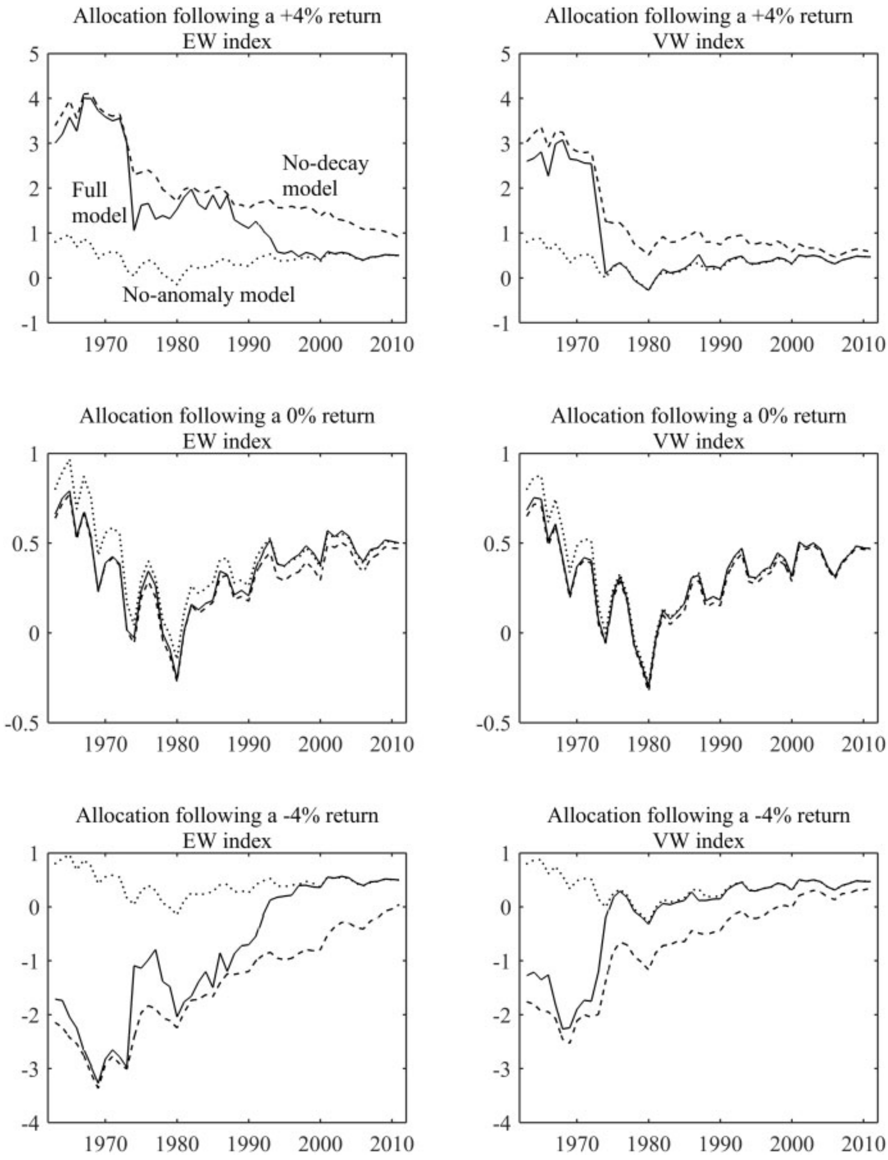

Figure 6. Rolling allocations under constant volatility. This figure shows the fraction of wealth invested in the market portfolio for portfolios formed on an out-of-sample basis using the constant volatility model of return autocorrelation. Rolling samples start in January 1953 and end each week between January 1963 and December 2011.

The portfolio allocations implied by the different models are shown in Figure 6. We plot these results only for the constant volatility model. Stochastic volatility introduces substantial variation in allocations that is unrelated to the conditional mean we are focusing on, making it more difficult to interpret the allocations. They are consistent with the conclusions we draw from the constant volatility case, however.

The panels in Figure 6 report hypothetical year-end allocations under three models for both the EW and VW indexes. Allocations are based on the 1-week ahead predictive distribution of returns computed on a rolling sample basis. They are hypothetical in that the conditional mean forecast is based on one of three hypothetical values of R_T rather than the actual value. The distribution of the parameters, however, is derived from the investor's posterior distribution given actual data.

In short, the results in Figure 6 are straightforward given the autocorrelations from Figure 5. In the early part of the sample, the investor who follows the full model chooses portfolio allocations that are fairly consistent with those of the no-decay investor. By the end of the sample, however, the investor who uses the full model is mirroring the investor with a fixed autocorrelation of zero and investing approximately 50% in the EW or VW market index regardless of the past return. The transition of the full model investor is relatively prolonged for the EW index, with a period of about 20 years during which the allocations are between those of the no-decline and no-anomaly investors. For the VW index, the transition is much quicker and follows the patterns in Figure 5.

Table VI summarizes the out-of-sample performance of these strategies. Under constant volatility, the strategy based on the full model fares best in terms of its Sharpe ratio. Realized utility, however, is lower than for the no-anomaly investor. The reason for this discrepancy is the misspecification represented by constant volatility. An investor who infers a substantial amount of autocorrelation will take a very significant position following a large positive or negative return. When a significant position is taken during a period of high volatility, an investor who treats volatility as constant runs the risk of earning an extremely negative return. For investors allocating to the EW index, these extreme returns were as low as −60% in a single week. Even a small number of these returns has a large negative impact on the realized utility measure.

Results under stochastic volatility, presented in the bottom panel, are much better. Aside from the last subsample, stochastic volatility-based strategies perform much better than constant volatility strategies, both for the EW and VW indexes. In addition, realized utilities for the full and no-decay models are much higher. This can be attributed to a dramatic reduction in extreme negative returns, the worst of which were −17% for the full model and −18% for the no-decay model.

Using stochastic volatility, the full model trounces the other specifications in terms of both Sharpe ratios and realized utilities. In the early subsample, its performance matches or slightly outperforms the no-decay strategy, as both take advantage of the strong autocorrelations observed during that period. In the late subsample, its performance matches the conservative no-anomaly strategy, while the no-decay portfolio performs badly as it trades on nonexistent autocorrelation.[18] As in Johannes, Korteweg, and Polson (2014), who consider the out-of-sample performance of a Bayesian investor in a much different setting, it appears that the full benefits of return predictability can only be gained under an appropriate model of conditional variance.

Overall, results are very positive for the full model, supporting the conclusion that modeling disappearance of the anomaly is crucial.

18 As an additional robustness check, we also combine the expected return forecasts of the constant volatility model with the out-of-sample volatility forecast of a GARCH(1,1) fitted to the same sequence of subsamples. These results are very similar to those of the model with stochastic volatility.

Table VI. Out-of-sample performance of autocorrelation-based strategies

This table evaluates out-of-sample performance of allocations implied by the disappearing anomaly model, a model that does not allow for disappearance, and a model that does not allow for the anomaly at all. For each of these models, the table presents results for the EW and VW market indexes. Models are estimated from samples starting in January 1953 and ending each week between January 1963 and December 2011. We include results for the full sample (1963–2011) and three approximately equal-length subsamples. The table reports the mean and standard deviation of weekly excess returns, the Sharpe ratio, and the realized utility ($U = E(R) - A\sigma^2(R)/2$, with $A = 10$). All are annualized and except for Sharpe ratios are expressed in percentages.

		Average excess return	Standard deviation	Sharpe ratio	Realized utility	Average excess return	Standard deviation	Sharpe ratio	Realized utility
		Constant volatility							
		EW index				VW index			
1963–2011	Full	13.81	22.17	0.62	13.56	3.22	8.71	0.37	3.18
	No-decay	15.39	26.41	0.58	15.04	3.26	11.98	0.27	3.19
	No-anomaly	3.86	7.54	0.51	3.83	2.15	5.92	0.36	2.13
1963–79	Full	27.17	24.15	1.13	26.88	5.59	11.79	0.47	5.52
	No-decay	31.84	26.91	1.18	31.48	6.12	15.84	0.39	6.00
	No-anomaly	3.62	6.75	0.54	3.59	1.73	5.31	0.33	1.72
1980–95	Full	9.77	27.93	0.35	9.38	2.00	4.60	0.43	1.99
	No-decay	13.05	32.04	0.41	12.54	3.32	10.60	0.31	3.26
	No-anomaly	4.23	5.28	0.80	4.21	2.81	3.86	0.73	2.80
1996–2011	Full	3.64	10.01	0.36	3.59	1.93	7.96	0.24	1.90
	No-decay	0.23	18.21	0.01	0.06	0.17	7.78	0.02	0.14
	No-anomaly	3.75	9.90	0.38	3.70	1.93	7.91	0.24	1.90
		Stochastic volatility							
		EW index				VW index			
1963–2011	Full	17.11	14.91	1.15	17.00	4.28	8.05	0.53	4.25
	No-decay	18.60	18.11	1.03	18.43	4.12	10.33	0.40	4.06
	No-anomaly	5.99	9.37	0.64	5.94	2.68	6.98	0.38	2.66
1963–79	Full	36.58	20.56	1.78	36.37	8.61	10.92	0.79	8.55
	No-decay	39.96	22.74	1.76	39.70	8.47	12.82	0.66	8.39
	No-anomaly	10.04	11.21	0.90	9.98	3.36	7.94	0.42	3.33
1980–95	Full	11.45	12.27	0.93	11.38	3.23	4.69	0.69	3.21
	No-decay	14.62	17.38	0.84	14.47	4.22	9.61	0.44	4.18
	No-anomaly	5.51	7.48	0.74	5.49	3.71	5.23	0.71	3.70
1996–2011	Full	2.08	8.33	0.25	2.04	0.75	7.03	0.11	0.72
	No-decay	−0.14	11.73	−0.01	−0.21	−0.62	7.71	−0.08	−0.65
	No-anomaly	2.14	8.88	0.24	2.11	0.93	7.41	0.13	0.90

4. Conclusions

We argue that when studying anomalies or their impact on portfolio choice, it is important to allow for the possibility of a prolonged decay. Such a decay may arise because investors need time to build models and capabilities or to market the idea to raise funds. There could also be implementation frictions, for example, transaction costs that only slowly decrease over time. Moreover, the market's efficiency may itself vary over time, as suggested by Lo (2004) and Akbas *et al*. (2016), which will make some anomalies longer lived than others.

We propose a framework for modeling anomalies that specifically allows for a gradual disappearance. Rather than describe an anomaly with a single parameter (e.g., its alpha), this framework calls for a triple of parameters: the initial strength (α), the time of discovery (τ), and the speed of disappearance (δ). Our approach nests and thus allows for direct tests of various economically motivated special cases: immediate disappearance, no decline, etc. We argue that estimating this model, or an equivalent model that allows for decay, would be a valuable part of any paper that identifies a new anomaly or that documents an existing anomaly in a different market or context. The model would allow the authors to better estimate the original full strength of the anomaly (which would be understated if the anomaly already started declining) and would improve out-of-sample predictions (which would be overstated if the decline already began).

To illustrate the value of our approach, we specialize the model to two well-known anomalies: the January effect and short-term autocorrelations of market returns. We show that the January effect is slowly dissipating, with the current magnitude of about a third of its original level. The decay likely started during the 1970s, with 1976 being the most probable first year of decline. We also find strong evidence for disappearance for the autocorrelations anomaly. Specifically, autocorrelations in the VW index have almost certainly declined to zero. Although there is some uncertainty about when this decline began, the early 1970s appear most probable, and point estimates suggest the disappearance was quite abrupt. The autocorrelations in the EW index have decayed more slowly, but they too have disappeared by about 1990s.

Our results shed light on how likely various explanations of the two anomalies are. The fact that both show gradual decline makes it unlikely that they were simply a result of data mining. Their decline also suggests that they are unlikely to be risk-driven—at least, such an explanation would need to explain why the compensation for risk decreased slowly over time. We believe that the most probable explanation is that both the January effect and autocorrelated market returns represented genuine mispricing that was initially unnoticed by investors.

Moreover, our approach sheds light on what may have triggered the decay of the anomalies. The most likely starting year for the decline of the January anomaly is 1976, coinciding with the publication of Rozeff and Kinney's 1976 study. To the best of our knowledge, Rozeff and Kinney (1976) is the first paper to document January seasonalities in the stock market, which suggests that academic research may play a role in eliminating market inefficiencies. In contrast, it appears that the decay of the stock index autocorrelations began well ahead of the publication of Lo and MacKinlay (1988), the seminal academic article on the subject. Instead, the timing coincides with the introduction of computerization to the NYSE in the early 1970s, suggesting that the anomaly was sustained by the frictions to trading, lessened in the early 1970s. This has important implications for research on the anomaly's underlying causes. Most obviously, studies using shorter, more recent sample periods may provide little insight on the historical factors that contributed to the anomalous behavior. For example, Ahn *et al.* (2002) attempt to determine the source of index autocorrelation by examining a variety of VW indexes and their corresponding futures contracts. However, equity index futures were not traded until the 1980s or later, by which time VW (though not EW) autocorrelations remained at just a tiny fraction of their original levels.

An additional contribution is to document a surprising result that shrinking an anomaly's returns toward a benchmark (in our case, the CAPM) may lead to increased allocations to the anomalous asset. Such shrinkage is often proposed to alleviate the issue of large weights implied by optimizers. However, a prior that reduces the initial level of an anomaly can cause that anomaly to persist for longer, and the higher estimated persistence may lead to higher end- and out-of-sample levels of the anomaly. Hence, skepticism about the existence of tradeable inefficiencies might be better represented as a belief both about the levels of anomalies and their rates of disappearance. We believe that this point, not yet made in the literature, will be of interest to both academics and practitioners.

Finally, we show that accounting for decay has a substantial impact on portfolio choice and, most importantly, that using a model that incorporates decay results in superior out-of-sample portfolio performance. This is notable, as the return predictability literature has recently been called into question by Goyal and Welch (2008), albeit in a different setting, for its inability to provide useful forecasts on an out-of-sample basis. The implicit assumption in Goyal and Welch (2008) is that the return generating process is stationary, so that the predictive relationships are unchanged over time. This "no-decay" approach often proves problematic in the settings that we analyze as well, in many cases underperforming strategies that ignore the abnormal return opportunities completely. However, by allowing anomalous returns to shrink over time, we are able to take advantage of return predictability in a way that is both conservative and adaptive. Our results suggest that predictive return models may be an important component of active investment strategies once those models allow for the possibility of decay.

Appendix A

Student-*t*/stochastic volatility sampling algorithm

In this appendix, we adapt the methods of Geweke (1993) and Jacquier, Polson, and Rossi (1994, hereafter JPR) to estimate the stochastic volatility model

$$R_t - \mu = \rho\delta^{(t-\tau)^+}(R_{t-1} - \mu) + \epsilon_t \tag{A.1}$$

$$\ln h_t = a + b \ln h_{t-1} + cR_{t-1} + \eta_t, \tag{A.2}$$

where $\epsilon_t \sim t(0, h_t, \nu)$, $\eta_t \sim \mathrm{N}\left(0, \sigma_\eta^2\right)$, and $t \in \{1, 2, ..., T\}$.

Following Geweke (1993), assuming that $\epsilon_t \sim t(0, h_t, \nu)$ is equivalent to assuming that $\epsilon_t = \sqrt{h_t}\sqrt{\omega_t}\epsilon_t^*$, where ϵ_t^* is a standard normal and $\nu/\omega_t \sim \chi^2(\nu)$. We adopt this latter representation.

Introducing stochastic volatility to the framework outlined in Section 2.2 requires adding an additional component to the Gibbs sampling algorithm. First, conditioning on the time series of h_t and of ω_t, we draw the parameters σ_η, a, b, c, and ν. Second, conditional on all the parameters as well as the asset returns, we draw values of h_t and ω_t.

Conditional on the h_t, drawing the parameters σ_η, a, b, and c uses standard regression results such as those found in Zellner (1971). In particular, the distribution of σ_η is an inverted gamma, and the vector $[a, b, c]$ given σ_η is multivariate normal. Conditional on the ω_t, drawing ν is also fairly straightforward. Following Geweke (1993), the conditional density of ν_t is

$$(\nu/2)^{T\nu/2}\Gamma(\nu/2)\exp(-\xi\nu), \quad \text{where} \quad \xi = \frac{1}{2}\sum_{t=1}^{T}\left[\ln(\omega_t) + \omega_t^{-1}\right] + \lambda. \tag{A.3}$$

We sample from this univariate density using the griddy Gibbs sampler. To draw the latent variable ω_t given all the parameter values and the time series of h_t and R_t, we use the result from Geweke that

$$(\epsilon_t^2 + \nu)/\omega_t \sim \chi^2(\nu + 1). \tag{A.4}$$

The last step is to draw the h_t conditional on the model parameters and the ω_t. As in JPR, we draw the entire time series of h_t by cycling through each element one at a time. In effect, this step actually consists of T separate draws from the densities

$$p(h_t|h_1, h_2, ..., h_{t-1}, h_{t+1}, ..., h_T, R_1, R_2, ..., R_T, \omega_1, \omega_2, ..., \omega_T, \theta), \tag{A.5}$$

where θ represents the vector of all model parameters. Similar to JPR, the Markovian nature of the h_t process and Bayes rule together imply that this density is proportional to

$$p(h_t|h_{t-1}, \theta)p(h_{t+1}|h_t)p(R_t|h_t, \omega_t). \tag{A.6}$$

Furthermore, this product of densities is proportional to

$$f(h_t) \equiv \frac{1}{h_t}\exp\left(-\frac{(\ln h_t - m_t)^2}{2s_t^2}\right)\frac{1}{\sqrt{h_t}}\exp\left(-\frac{\epsilon_t^2}{2\omega_t h_t}\right), \tag{A.7}$$

where

$$\epsilon_t = (R_t - \mu) - \rho\delta^{(t-\tau)^+}(R_{t-1} - \mu), \tag{A.8}$$

$$m_t = [a(1 - b) + b(\ln h_{t-1} + \ln h_{t+1}) + cR_{t-1} - bcR_t]/(1 + b^2), \tag{A.9}$$

$$s^2 = \sigma_\eta^2/(1 + b^2). \tag{A.10}$$

When $c = 0$ and $\omega_t = 1$, these match the formulas found in JPR. Similarly to JPR, we use the Metropolis Hastings algorithm with the candidate generating density

$$q(h_t) \propto h_t^{-(\phi+1)}\exp(-\psi_t/h_t), \tag{A.11}$$

where $\phi = -1.5 + (1 - 2\exp(s^2))/(1 - \exp(s^2))$ and $\psi_t = .5\epsilon^2/\omega_t + (\phi + 1)\exp(m_t + .5s^2)$. This produces an inverse gamma candidate generating density that approximates the target density in Equation (A.7). JPR show that this candidate generator has relatively thick tails and demonstrate good convergence properties.

A candidate draw h_t^* from this inverse gamma is then accepted, replacing the current draw h_t, with probability

$$\min\left\{\frac{f(h_t^*)/q(h_t^*)}{f(h_t)/q(h_t)}, 1\right\}. \tag{A.12}$$

If the draw is rejected, the current draw h_t is kept.

By drawing each h_t in turn, from $t=1$ to $t=T$, a new draw of the time series of h_t is obtained. While the algorithm must be modified for $t=1$ and $t=T$, this is straightforward following the procedure of Jacquier, Polson, and Rossi (2001).

References

Ahn, D., Boudoukh, J., Richardson, M., and Whitelaw, R. F. (2002) Partial adjustment or stale prices? Implications from stock index and futures return autocorrelations, *Review of Financial Studies* **15**, 655–689.

Akbas, F., Armstrong, W. J., Sorescu, S., and Subrahmanyam, A. (2016) Capital market efficiency and arbitrage efficacy, *Journal of Financial and Quantitative Analysis* **51**, 387–413.

Avramov, D. (2004) Stock return predictability and asset pricing models, *Review of Financial Studies* **17**, 699–738.

Banz, R. W. (1981) The relationship between return and market value of common stocks, *Journal of Financial Economics* **9**, 3–18.

Blume, M. E., Siegel, J. J., and Rottenberg, D. (1993) *Revolution on Wall Street: The Rise and Decline of the New York Stock Exchange*, W.W. Norton, New York, NY.

Busse, J. and Irvine, P. (2006) Bayesian alphas and mutual fund persistence, *Journal of Finance* **61**, 2251–2288.

Conrad, J. and Kaul, G. (1988) Time-variation in expected returns, *Journal of Business* **61**, 409–425.

Cooper, M. J., Dimitrov, O., and Rau, P. R. (2001) A Rose.com by any other name, *Journal of Finance* **56**, 2371–2388.

Fama, E. (1965) The behavior of stock prices, *Journal of Business* **38**, 34–105.

Fisher, L. (1966) Some new stock market indexes, *Journal of Business* **39**, 191–225.

Fleming, J., Kirby, C., and Ostdiek, B. (2001) The economic value of volatility timing, *Journal of Finance* **56**, 329–352.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Geweke, J. (1993) Bayesian treatment of the independent Student-*t* linear model, *Journal of Applied Econometrics* **8**, S19–S40.

Goyal, A. and Welch, I. (2008) A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* **21**, 1455–1508.

Hand, J., Green, G., and Soliman, M. (2011) Going, going, gone? The demise of the accruals anomaly, *Management Science* **57**, 797–816.

Haugen, R. A. and Jorion, P. (1996) The January effect: Still there after all these years, *Financial Analysts Journal* **52**, 27–31.

Hawawini, G. (1980) Intertemporal cross dependence in securities daily returns and the short-term intervailing effect on systematic risk, *Journal of Financial and Quantitative Analysis* **15**, 139–149.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994) Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics* **12**, 371–389.

Jegadeesh, N. and Titman, S. (2001) Profitability of momentum strategies: An evaluation of alternative explanations, *Journal of Finance* **56**, 699–720.

Johannes, M., Korteweg, A., and Polson, N. G. (2014) Sequential learning, predictive regressions and optimal portfolios, *Journal of Finance* **69**, 611–644.

Jones, C. and Shanken, J. (2005) Mutual fund performance with learning across funds, *Journal of Financial Economics* **78**, 507–552.

Jorion, P. (1986) Bayes–Stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* **21**, 279–292.

Kandel, S. and Stambaugh, R. F. (1996) On the predictability of stock returns: An Asset Allocation Perspective, *Journal of Finance* **51**, 385–424.

Keim, D. B. (1983) Size-related anomalies and stock return seasonality: Further empirical evidence, *Journal of Financial Economics* **12**, 13–32.

Lo, A. W. (2004) The Adaptive market hypothesis: Market efficiency from an evolutionary perspective, *Journal of Portfolio Management* **30**, 15–29.

Lo, A. W. and MacKinlay, A. C. (1988) Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of Financial Studies* **1**, 41–66.

Lo, A. W. and MacKinlay, A. C. (1999) *A Non-Random Walk Down Wall Street*, Princeton University Press, Princeton, NJ.

McLean, R. D. and Pontiff, J. (2016) Does academic research destroy stock market predictability?, *Journal of Finance* **71**, 5–32.

McQueen, G. and Thorley, S. (1997) Do investors learn? Evidence from a gold market anomaly, *Financial Review* **32**, 501–525.

Mittoo, U. and Thompson, R. (1990) Do the capital markets learn from financial economists? Working paper, University of Manitoba.

Pastor, L. (2000) Portfolio selection and asset pricing models, *Journal of Finance* **40**, 179–223.

Pastor, L. and Stambaugh, R. F. (1999) Cost of equity capital and model mispricing, *Journal of Finance* **54**, 67–121.

Pastor, L. and Stambaugh, R. F. (2001) The equity premium and structural breaks, *Journal of Finance* **56**, 1207–1239.

Pastor, L. and Stambaugh, R. F. (2002) Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics* **63**, 315–349.

Reinganum, M. R. (1983) The anomalous stock market behavior of small firms in January: Empirical tests for tax-loss selling effects, *Journal of Financial Economics* **12**, 89–104.

Rozeff, M. and Kinney, W. (1976) Capital market seasonality: The case of stock returns, *Journal of Financial Economics* **3**, 379–402.

Schwert, G. W. (2003) Anomalies and market efficiency, in: G. Constantinides, M. Harris, and R. M. Stulz (eds.), *Handbook of the Economics of Finance*, North-Holland, Amsterdam, pp. 937–972.

Sloan, R. (1996) Do stock prices fully reflect information in accruals and cash flows about future earnings, *The Accounting Review* **71**, 289–316.

Tu, J. and Zhou, G. (2004) Data-generating process uncertainty: What difference does it make in portfolio decisions? *Journal of Financial Economics* **72**, 385–421.

Tu, J. and Zhou, G. (2010) Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty, *Journal of Financial and Quantitative Analysis* **45**, 959–986.

Watts, R. (1978) Systematic 'abnormal' returns after quarterly earnings announcements, *Journal of Financial Economics* **6**, 127–150.

Zellner, A. (1971) *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.