Applied Stochastic Models in Business and Industry

Research Article

Received 26 August 2013, Revised 20 October 2014, Accepted 20 October 2014 Published online 21 November 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/asmb.2087

Bayesian analysis of herding behaviour: an application to Spanish equity mutual funds

Laura Andreu[a], Pilar Gargallo[b], Manuel Salvador[b] and José Luis Sarto[a]

This paper proposes a dynamic Bayesian rolling window estimation procedure applied to the three-factor model of Fama and French to analyse herding behaviour in the style exposures of mutual funds. This procedure allows a user to dynamically select the length of the estimation window by means of weighted likelihood functions that discount the loss of information because of time. This method is very flexible and allows us to consider different approaches of detecting herding behaviour by taking into account the uncertainty associated in the estimation of the style coefficients. In particular, the paper first determines the convergence behaviour following the traditional LSV herding measure and then refines this method by removing the influence exerted by market conditions, such as market volatility and returns, on this convergence. This process is empirically illustrated by an application to Spanish equity mutual funds. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: Bayesian inference; weighted likelihood; MCMC; herding behaviour; Fama and French three-factor model; style exposures; equity mutual funds

1. Introduction

Recent research in financial economics suggests that both rationality and emotion are important factors in the decision-making process. This idea, introduced by behavioural finance, challenges the efficient market paradigm, in which prices reflect all available information. In fact, the financial media often portrays institutional investors as driving prices from fundamental values and generating excess volatility, which can lead to mispricing and bubbles. In this context, certain psychological biases, such as investors' preference to avoid losses (see, e.g., [1,2]), may imply that significant price fluctuations are not necessarily related to the arrival of economic information but may instead correspond to collective phenomena, such as herding behaviour.

Herding is generally defined as a simple convergence of behaviour [3]. However, this convergence can arise from the interactions among financial agents when they decide to imitate or copy the decisions of other agents or from the existence of certain common external factors or information available for the group. The first behaviour would be considered 'true herding', whereas the second would be termed 'spurious herding', according to Bikhchandani and Sharma [4]. The empirical identification of the split between both types of herding is not an easy task because of the large number of factors that can influence an investment decision.

Several theoretical models of herding behaviour have been proposed in the literature. These models suggest different arguments to explain this phenomenon highlighting the following: underlying investors' flows [5], institutional feedback trading [6], reputation [7,8], informational cascades [9,10], and the existence of correlated signals [11]. Consequently, herding is usually explained within the context of the agency theory, in which compensation and reputation play an important role.

In spite of the theoretical arguments, the dominant role of institutional investors in financial markets worldwide has led researchers to empirically examine this phenomenon. To do so, researchers have used different methodologies, relying on varying data and markets. Concretely, Lakonishok *et al.* [12] proposed a metric (Lakonishok; Shleifer; Vishny (LSV) metric hereafter) that has been widely used when examining institutional herding.[‡]

[a] *Department of Accounting and Finance, University of Zaragoza, Zaragoza, Spain*
[b] *Department of Statistical Methods, School of Business and Economics, University of Zaragoza, Zaragoza, Spain*
**Correspondence to: Manuel Salvador, Facultad de Economía y Empresa Gran Vía 2, 50.005, Zaragoza, Spain.*
[†] *E-mail: salvador@unizar.es*
[‡] *See, e.g.,* [13–15], *among the first references.*

The LSV measure is based on the trades conducted by a group of market participants and compares the actual behaviour of managers with an ideal scenario that assumes independent and random trades. The aim of the LSV metric is to detect convergent patterns in the buying and selling decisions of portfolio managers. In spite of the growing empirical literature focusing on herding in individual stocks, the proposed reasons for herding at the level of individual assets hold at least equally well at the industry and style levels (see, e.g., [16, 17]).

As opposed to Choi and Sias [17], who focused on industry herding because of the industry/sector recommendations of financial analysts, we focus on convergent patterns of investment styles, considering mutual funds' exposures to the market, to market capitalisation and to book-to-market ratios, given the importance of the investment style to explain the performance achieved by portfolio managers. Consequently, we first determine the style exposures of Spanish mutual funds to the three factors included in the model of Fama and French [18] and then examine the evolution of investment styles through time to detect convergent behaviour among managers.

As previously stated, these convergent patterns can reflect the same reactions of managers to common information. This circumstance leads us to account for the market conditions when evaluating the herding level. Hence, the paper also aims at evaluating the estimated level of spurious herding linked to concrete market conditions in each period. From this analysis, we can determine the intentional herding depending on whether the observed behaviour of the manager differs significantly from the expected behaviour, depending on market conditions.

Additionally, it is important to note that the style coefficients have been estimated and therefore introduce uncertainty within the estimation process; this circumstance should be considered in capturing the herding level. To tackle this issue, we propose a Bayesian rolling window estimation procedure that dynamically selects the length of the estimation window by means of weighted likelihood functions that discount the loss of information because of time. Once the posterior distributions of the beta coefficients of the model are calculated, we introduce a general method to analyse the herding behaviour, comparing the changes in the style coefficients estimated to those expected under the hypothesis of no herding.

The first contribution of our paper refers to the extension of the style investing literature. Despite the importance of investment styles to portfolio management, only recent research has been devoted to analysing the style herding so far (see, e.g., [16, 17]). Second, this paper provides a methodological contribution because it is the first study, to the best of our knowledge, to take into account the uncertainty associated with the estimation of style exposures in capturing style herding. To this end, we use a Bayesian rolling window estimation procedure that allows us to estimate the style coefficients without any previous specification of the dynamic evolution equation.

Our results first provide evidence of herding behaviour among managers when using the traditional LSV herding measure. Second, splitting this herding level into the imitation and convergence explained by the commonly available information about market conditions (return and volatility) allows us to detect more accurate herding levels.

The paper is organised as follows. In Section 2, we set up the problem and describe the estimation procedure of the style exposures. In Section 3, we detail the methodology used to analyse the herding behaviour. Section 4 applies the proposed procedure to the analysis of Spanish equity mutual funds. Finally, Section 5 draws final conclusions. Some mathematical details of the paper are located in the Appendix.

2. Bayesian dynamic estimation of the style exposures

In this section, we set up the problem and describe the estimation procedure of monthly style exposures from the daily gross excess returns.

2.1. Problem set-up

We consider N mutual funds observed in period $\{1, \dots, T\}$. Let $R_{i,t}$ be the daily return of mutual fund i in day t where $t \in \mathbf{T}_i \subseteq \{1, \dots, T\}$ is the observation period, for $i = 1, \dots, N$[§]; let $y_{i,t} = R_{i,t} - RF_t$ be the excess return of mutual fund i in day t where $RF_t; t = 1, \dots T$ is the daily evolution of the risk-free rate. Finally, let $\{\mathbf{M}_k; k = 1, \dots, K\}$ be the months included in the period $\{1, \dots, T\}$.

To determine the monthly style coefficients of each mutual fund, we use the three-factor model of Fama and French [18] given by:

$$y_{i,t} = \beta_{0,i,k} + \beta_{1,i,k}\,\text{RMRF}_t + \beta_{2,i,k}\,\text{SMB}_t + \beta_{3,i,k}\text{HML}_t + \varepsilon_{i,t} \text{ with } \varepsilon_{i,t} \sim N\left(0, \sigma^2_{i,k}\right) \tag{2.1}$$

[§] *Note that our data are free of survivorship bias, and therefore, the life of each mutual fund may differ.*

where $t \in \mathbf{M}_k$; $k = 1, \ldots, K; i \in \{1, \ldots, N\}$ such that $\mathbf{M}_k \cap \mathbf{T}_i \neq \emptyset$, RMRF_t is the excess return of the market portfolio from the risk-free rate the day t, SMB_t is the return on factor mimicking portfolios for size the day t, HML_t is the return on factor mimicking portfolios for value the day t, $\beta_{0,i,k}$ is the value added by the manager of the i^{th} fund in the month $\mathbf{M}_k$, $\beta_{1,i,k}$ is the market style exposure of the i^{th} fund in the month $\mathbf{M}_k$, $\beta_{2,i,k}$ is the SMB style exposure of the i^{th} fund in the month $\mathbf{M}_k$ and $\beta_{3,i,k}$ is the HML style exposure of the i^{th} fund in the month $\mathbf{M}_k$.

2.2. *Estimation of the model*

To estimate the monthly beta coefficients $\{\beta_{i,k} = (\beta_{0,i,k}, \beta_{1,i,k}, \beta_{2,i,k}, \beta_{3,i,k})'; i = 1, \ldots, N; k = 1, \ldots, K\}$, we use, for each mutual fund and each month, a Bayesian rolling window estimation procedure based on the use of a geometrically weighted likelihood function that discounts the loss of information because of time. This procedure sequentially determines the importance of past months to the estimation of the previous coefficients, taking into account their position with respect to the current month. The use of a weighted likelihood function enables us to avoid an explicit specification of a concrete dynamic evolution for the beta coefficients $\beta_{i,k}$ by selecting for each month and mutual fund, the data that must be taken into account for the estimation and the weights to be assigned to them in the estimation process. Consequently, this method permits the data to 'speak for themselves', lending a semi-parametric flavour to our estimation procedures.

Note that time-based weighted likelihood methods have been traditionally used to calculate robust maximum likelihood estimators [19, 20], as they are particularly appealing for dealing with the quite nonstandard properties of assets returns, including nonlinear dynamics that change over time (i.e., volatility clustering), high excess kurtosis, and possible time-varying parameters [21, 22]. To the best of our knowledge, there are no Bayesian analyses that have used these methods, with the exceptions of Newton and Raftery [23], who proposed a Bayesian weighted likelihood bootstrap procedure to calculate the posterior distribution of the parameters in a different context, and of West [24], who proposed the use of power discounting of prior distributions as a general approach to build neutral alternatives in Bayesian model monitoring processes. Next, we explain how this weighted likelihood function is built and describe the prior distribution of the parameters of the model.

2.2.1. Likelihood function.

Let $\boldsymbol{\theta}_{i,k} = \left(\boldsymbol{\beta}'_{i,k}, \tau_{i,k}, \gamma_{i,k}\right)'$ where $\tau_{i,k} = \frac{1}{\sigma^2_{i,k}}$ and $\gamma_{i,k} \in (0, 1]$ is the discount factor applied to each observation. Let $\mathbf{M}^i_{\leqslant k} = \bigcup_{j=1}^{k} \left(\mathbf{M}_j \cap \mathbf{T}_i\right)$ be the previous observed time interval of the i^{th} mutual fund, including the current month $\mathbf{M}_k$; $\mathbf{Y}_{i,k} = \left(y_{i,t}; t \in \mathbf{M}^i_{\leqslant k}\right)'$ be the observations of the i^{th} mutual fund returns in $\mathbf{M}^i_{\leqslant k}$, and $\mathbf{X}_k = \begin{pmatrix} \mathbf{1}'_{\left|\mathbf{M}^i_{\leqslant k}\right|} \\ \mathrm{RMRF}_t; t \in \mathbf{M}^i_{\leqslant k} \\ \mathrm{SMB}_t; t \in \mathbf{M}^i_{\leqslant k} \\ \mathrm{HML}_t; t \in \mathbf{M}^i_{\leqslant k} \end{pmatrix}$ where $\mathbf{1}_m$ denotes the vector of m ones, the $4 \times \left|\mathbf{M}^i_{\leqslant k}\right|$ matrix of the independent covariates of (2.1).

In order to estimate $\boldsymbol{\theta}_{i,k}$ for each mutual fund i, and each month, $\mathbf{M}_k$ we set up a weighted likelihood function given by:

$$
\begin{aligned}
L(\boldsymbol{\theta}_{i,k}; \mathbf{Y}_{i,k}, \mathbf{X}_k) &\propto \prod_{j=1}^{k} \prod_{t \in \mathbf{M}_j \cap \mathbf{T}_i} \tau_{i,k}^{\frac{\gamma_{i,k}(j)}{2}} \exp\left[-\frac{\gamma_{i,k}(j)\tau_{i,k}}{2}\left(y_{i,t} - \beta_{0,i,k} - \beta_{1,i,k}\mathrm{RMRF}_t - \beta_{2,i,k}\mathrm{SMB}_t - \beta_{3,i,k}\mathrm{HML}_t\right)^2\right] \\
&= \tau_{i,k}^{\frac{\left|\mathbf{M}^i_{\leqslant k}\right|}{2}} \exp\left[-\frac{\tau_{i,k}}{2}\sum_{j=1}^{k}\gamma_{i,k}(j)\sum_{t \in \mathbf{M}_j \cap \mathbf{T}_i}\left(y_{i,t} - \beta_{0,i,k} - \beta_{1,i,k}\mathrm{RMRF}_t - \beta_{2,i,k}\mathrm{SMB}_t - \beta_{3,i,k}\mathrm{HML}_t\right)^2\right]
\end{aligned}
\tag{2.2}
$$

where $0 \leqslant \gamma_{i,k}(1) \leqslant \gamma_{i,k}(2) \leqslant \ldots \leqslant \gamma_{i,k}(k)$ being $\sum_{j=1}^{k}\gamma_{i,k}(j) = \left|\mathbf{M}^i_{\leqslant k}\right|$ a set of discount factors applied to observations $\mathbf{Y}_{i,j}; j = 1, \ldots, k$, determining their influence on the estimation of $\boldsymbol{\theta}_{i,k}$.

We assume that the further in the past an observation is, the lower its influence is on this estimation. Specifically, we take $\gamma_{i,k}(j) = \frac{\gamma_{i,k}^{k-j}}{\sum_{\ell=1}^{k}\left|\mathbf{M}_\ell \cap \mathbf{T}_i\right|\gamma_{i,k}^{k-\ell}}\left|\mathbf{M}^i_{\leqslant k}\right|$ where $0 \leqslant \gamma_{i,k} \leqslant 1$ is a parameter that determines the discount rate of the information provided by past observations on the estimation of $\boldsymbol{\theta}_{i,k}$ in such a way that the lower its value, the larger the discount of the past information.[¶] Note that if $\gamma_{i,k} = 1$, the observations $\{\mathbf{Y}_{i,j};\ j = 1, \dots, k\}$ are assigned the same weight (equal to 1), and if $\gamma_{i,k} = 0$, the past observations $\{\mathbf{Y}_{i,j}; j = 1, \dots, k\text{-}1\}$ are discarded to estimate $\boldsymbol{\theta}_{i,k}$ because they are assigned a weight equal to 0. Only current observations $\mathbf{Y}_{i,k}$ are assigned a weight equal to $\frac{\left|\mathbf{M}^i_{\leqslant k}\right|}{\left|\mathbf{M}_k \cap \mathbf{T}_i\right|}$.

This estimation procedure has several advantages. On the one hand, the use of a weighted likelihood function enables us to avoid the specification of any evolution equation of the coefficients $\boldsymbol{\theta}_{i,k}$, the values of which are determined by the observations $\{\mathbf{Y}_{i,j}; j = 1, \dots, k\}$ and the discount factor $\gamma_{i,k}$. Moreover, the assumption of an unknown discount factor, $\gamma_{i,k}$, which is estimated using the data, provides a non-parametric flavour to the procedure, increasing its flexibility. On the other hand, the application of a Bayesian approach allows us to make exact inferences about the parameters $\boldsymbol{\theta}_{i,k}$ and $\gamma_{i,k}$. To that end, we need to specify the prior distribution described next.

2.2.2. Prior distribution. The prior distribution is given by the following:

$$\beta_{i,k} \dots N_4\left(\mathbf{b}_0, \mathbf{S}_\beta\right) \tag{2.3}$$

$$\tau_{i,k} \sim \text{Gamma}\left(\frac{n_\tau}{2}, \frac{n_\tau s^2_\tau}{2}\right) \tag{2.4}$$

$$\gamma_{i,k} \sim \text{Uniform}(g_{min}, g_{max}) \tag{2.5}$$

where $\mathbf{b}_0(4\times1)$ is a known vector, $\mathbf{S}_\beta = \text{diag}\left(s^2_{\beta,j}; j = 0, 1, 2, 3\right)$, $n_\tau > 0$, s^2_τ, and $0 \leqslant g_{min} < g_{max} \leqslant 1$ are known constants. Additionally, we assume that (2.3)–(2.5) are mutually independent.

2.2.3. Posterior distribution. Inference about the component of $\boldsymbol{\theta}_{i,k}$ parameters are made from their posterior distribution that, by the Bayes theorem, is given by the following:

$$\begin{aligned}\pi(\boldsymbol{\theta}_{i,k} \left| Y_{i,k}, X_k\right.) \propto \tau_{i,k}^{\frac{\left|\mathbf{M}^i_{\leqslant k}\right|}{2}} \exp\left[-\frac{\tau_{i,k}}{2}\sum_{j=1}^{k}\gamma_{i,k}(j)\sum_{t\in \mathbf{M}_j\cap\mathbf{T}_i}\left(y_{i,t} - \beta_{0,i,k} - \beta_{1,i,k}RMRF_t - \beta_{2,i,k}SMB_t - \beta_{3,i,k}HML_t\right)^2\right] \\ \exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}_{i,k} - \mathbf{b}_0\right)'\mathbf{S}_\beta^{-1}\left(\boldsymbol{\beta}_{i,k} - \mathbf{b}_0\right)\right]\tau_{i,k}^{\frac{n_\tau}{2}-1}\exp\left[-\frac{n_\tau s^2_\tau}{2}\tau_{i,k}\right] I_{(0,\infty)}\left(\tau_{i,k}\right) I_{\left(g_{min},g_{max}\right)}\left(\gamma_{i,k}\right)\end{aligned} \tag{2.6}$$

where $I_A(.)$ denotes the indicator function.

Given that this distribution is not analytically tractable, we used MCMC methods. Specifically, we used Gibbs sampling (see, e.g., [25]) to make these calculations. Most of the full conditional distributions of (2.6) are standard and are given in the appendix. From this algorithm, for each mutual fund $i = 1, \dots, N$ and month $\mathbf{M}_k; k \in \{1, \dots., K\}$ such that $\mathbf{M}_k \cap T_i \neq \emptyset$, we obtain a sample:

$$\left\{\boldsymbol{\theta}^{(s)}_{i,k} = \left(\beta^{(s)}_{0,i,k}, \beta^{(s)}_{1,i,k}, \beta^{(s)}_{2,i,k}, \beta^{(s)}_{3,i,k}, \tau^{(s)}_{i,k}, \gamma^{(s)}_{i,k}\right); s = 1, \dots., S\right\} \tag{2.7}$$

of (2.6), from which we can make inferences about the style coefficients $\beta_{i,k}$ used to analyse the existence of herding.

3. Bayesian analysis of herding behaviour

In this section, we describe a general algorithm to analyse the existence of herding behaviour in the monthly evolution of the style coefficients $\{(\beta_{1,i,k}, \beta_{2,i,k}, \beta_{3,i,k}); k = 1,\dots,K; i = 1,\dots,N\}$ of each mutual fund. This algorithm takes into account

[¶]*Other methods of discounting the past information could have been adopted* [22]. *We have chosen this exponential approach because it is used in the exponential weighted moving average methods, given that the discount of information is usually very fast in financial data.*

the uncertainty associated with the estimation of the above coefficients, providing a more accurate quantification of the estimation error. We first apply this general algorithm considering the approach of the LSV measure, and then we take the market conditions into account.

3.1. General procedure

Our objective is to detect the potential existence of herding, analysing the convergence of movements in style coefficients. This analysis is carried out for each month, $\mathbf{M}_k$, and for each style coefficient, β_j, through the comparison of the 'observed' percentage of mutual funds that have increased the style exposure, $P^+_{j,k}$, with their expected value under the hypothesis of no herding in month $\mathbf{M}_k$, $EP^+_{j,k}$.

The 'observed' percentage is given by $P^+_{j,k} = \frac{N^+_{j,k}}{|I_k|}$ with $I_k = \{i \in \{1,\ldots,N\}: \mathbf{T}_i \cap \mathbf{M}_k \neq \emptyset\}$and $N^+_{j,k} = \left|\{i \in I_k : \beta_{j,i,k} > \beta_{j,i,k-1}\}\right|$. Therefore, if $P^+_{j,k} = EP^+_{j,k}$, there is not herding; otherwise, if $P^+_{j,k} > EP^+_{j,k}$ ($P^+_{j,k} < EP^+_{j,k}$), there is buying (selling) herding.

If there is no herding in month $\mathbf{M}_k$, the actual behaviour of managers would be given by an ideal scenario that assumes independent and random trades in such a way that:

$$M^+_{j,k} \mid \text{no herding in month } \mathbf{M}_k \sim \text{Binomial}\left(|I_k|, EP^+_{j,k}\right) \tag{3.1}$$

where $M^+_{j,k}$ is the number of funds that increase their risk exposure to the jth factor.

In order to decide if there is significant evidence of herding, we fix a threshold $0<\gamma<1$ and, taking into account (3.1), we calculate the posterior *p*-values [26]:

$$P\left[Q^+_{j,k} \leqslant P^+_{j,k} | \textit{no} \text{ herding in month } \mathbf{M}_k, \{\mathbf{Y}_{i,k}, \mathbf{X}_k; i \in I_k\}\right] = P\left[M^+_{j,k} \leqslant N^+_{j,k} | M^+_{j,k} \sim Bi\left(|I_k|, EP^+_{j,k}\right), \{\mathbf{Y}_{i,k}, \mathbf{X}_k; i \in I_k\}\right] \tag{3.2}$$

and

$$P\left[Q^+_{j,k} \geqslant P^+_{j,k} | \textit{no} \text{ herding in month } \mathbf{M}_k, \{\mathbf{Y}_{i,k}, \mathbf{X}_k; i \in I_k\}\right] = P\left[M^+_{j,k} \geqslant N^+_{j,k} | M^+_{j,k} \sim Bi\left(|I_k|, EP^+_{j,k}\right), \{\mathbf{Y}_{i,k}, \mathbf{X}_k; i \in I_k\}\right] \tag{3.3}$$

where $Q^+_{j,k} = \frac{M^+_{j,k}}{|I_k|}$.

These posterior *p*-values measure the evidence for no herding hypothesis by taking into account that $P^+_{j,k}$ is measured with error because it depends on the estimated values of the style coefficients $\{(\beta_{j,i,k-1}, \beta_{j,i,k}) : i \in I_k\}$. If (3.2) (resp. (3.3)) is less than or equal to γ, we will conclude that $P^+_{j,k} < EP^+_{j,k}$(resp. $P^+_{j,k} > EP^+_{j,k}$) and, therefore, there is significant evidence of selling (resp. buying) herding in month $\mathbf{M}_k$. On the contrary, if the minimum of (3.2) and (3.3) is larger than γ, we will conclude that there is not significant evidence of herding.

Using the Monte Carlo method, we calculate (3.2) and (3.3) by means of the following expressions:

$$\frac{1}{S}\sum_{s=1}^{S} P\left[M^+_{j,k} \leqslant N^{+,(s)}_{j,k} | M^+_{j,k} \sim Bi\left(|I_k|, EP^+_{j,k}\right)\right] \tag{3.4}$$

and

$$\frac{1}{S}\sum_{s=1}^{S} P\left[M^+_{j,k} \geqslant N^{+,(s)}_{j,k} | M^+_{j,k} \sim Bi\left(|I_k|, EP^+_{j,k}\right)\right] \tag{3.5}$$

where

$$N^{+,(s)}_{j,k} = \left|\left\{i \in I_k : \beta^{(s)}_{j,i,k} > \beta^{(s)}_{j,i,k-1}\right\}\right| \tag{3.6}$$

and $\beta^{(s)}_{j,i,k}$ is the sth element of the sample (2.7) for s=1,…,S.

3.2. Applications

After describing the general procedure of analysing the existence of herding, in this section, we consider two applications. First, we apply this methodology to the LSV measure where $EP^+_{j,k} = 0.5$, under the assumption that the probability of increasing the style exposure is the same as the probability of decreasing it. Second, given that the LSV herding metric does not consider that convergent patterns can be partially explained by similar reactions to common information, we propose an alternative approach to establish an expected herding level that takes into account the information provided by a set of Q covariates (C_1, …, C_Q) that describe the market conditions. Concretely, we set up the following binomial regression model to calculate the expected percentages $EP^+_{j,k}$ according to the market conditions:

$$N^+_{j,k} \sim \text{Binomial}\left(\left|I_k\right|, EP^+_{j,k}\right) \tag{3.7}$$

$$\text{logit}\left(EP^+_{j,k}\right) = \log\left(\frac{EP^+_{j,k}}{1 - EP^+_{j,k}}\right) = \sum_{m=1}^{Q} \psi_{j,m} C_{m,k} \tag{3.8}$$

where ($C_{1,k}$, …, $C_{Q,k}$) are the values of (C_1, …, C_Q) in the month $\mathbf{M}_k$ for k=1,…, K. We take a fairly non-informative prior on $\Psi_j = (\psi_{j,1}, \ldots, \psi_{j,Q})'$ given by $\Psi_j \sim N_Q(\mathbf{0},\mathbf{D}_0)$ with $\mathbf{D}_0 = 100\mathbf{I}_Q$. In this case, and given that $EP^+_{j,k}$ depends on Ψ_j, the calculation of (3.2) and (3.3) will be made using the following expressions:

$$\frac{1}{S}\sum_{s=1}^{S} P\left[M^+_{j,k} \leqslant N^{+,(s)}_{j,k} | M^+_{j,k} \sim Bi\left(\left|I_k\right|, EP^{+,(s)}_{j,k}\right)\right] \tag{3.9}$$

and

$$\frac{1}{S}\sum_{s=1}^{S} P\left[M^+_{j,k} \geqslant N^{+,(s)}_{j,k} | M^+_{j,k} \sim Bi\left(\left|I_k\right|, EP^{+,(s)}_{j,k}\right)\right] \tag{3.10}$$

respectively, where

$$EP^{+,(s)}_{j,k} = \frac{\exp\left[\sum_{m=1}^{Q} \psi^{(s)}_{j,m} C_{m,k}\right]}{1 + \exp\left[\sum_{m=1}^{Q} \psi^{(s)}_{j,m} C_{m,k}\right]} \tag{3.11}$$

and $\left\{\psi^{(s)}_{j,m}; s = 1, \ldots, S\right\}$ is a sample of the posterior distribution of Ψ_j. This sample is obtained by the composition sampling method where $N^{+,(s)}_{j,k}$ is given by (3.6) and $\psi^{(s)}_{j,m}$ is drawn from the posterior distribution of $\boldsymbol{\Psi}_j | N^{(s)}_{j,k}$, which can be calculated using MCMC procedures.[‖]

4. Empirical application to Spanish equity mutual funds

In this section, we apply the methodology described in Section 3 to analyse the herding behaviour of Spanish equity mutual funds.

4.1. Data

Our primary data source is the survivorship bias-free mutual fund dataset provided by the Spanish Securities Exchange Commission (CNMV). This database covers Spanish open-end mutual funds and provides information about net asset values, total net assets, investment objectives, and other fund characteristics. Additionally, we collect information on the

‖*In the empirical application to Spanish equity mutual funds, we used the slicesample procedure of MATLAB 7.9.0 (Mathworks, Inc. Natick, Massachusetts, USA) with a burning period of 1000 iterations and a number of random samples equals to 1.*

Table I. Descriptive statistics.

	Market excess return	Mutual fund excess return	Mutual fund excess return volatilities	Mutual fund excess returns skewness	Mutual fund excess returns kurtosis
Mean	0.56%	0.33%	5.24%	1.90	8.96
Standard deviation	5.96%	5.41%	3.34%	0.79	4.59
Median	0.94%	0.85%	4.34%	2.05	9.19
Minimum	−17.95%	−38.14%	0.00%	0.26	1.89
Maximum	15.84%	22.97%	41.04%	4.32	28.29

Spanish value-weighted index, on the Fama–French factors from Morgan Stanley Capital International (MSCI) , and on 1-day Treasury Bill Repos (the risk-free rate) from the Bank of Spain.

Our analysis of style herding is based on return data. Although this information is apparently less informative than portfolio holdings, return data are collected with a higher frequency than portfolio holdings. Furthermore, portfolio holdings information is problematic when fund managers are following window dressing practices to disguise the actual portfolios held. In this context, we use daily return data, whereas previous papers analysing mutual fund herding either were focused on quarterly portfolio holdings ([12, 17], among others) or used monthly returns [27].

We focus on the Spanish mutual fund industry because of its rapid growth in the last two decades. In 1989, the industry managed €5 billion, increasing to €162 billion by December 2009. Concretely, we focus on domestic equity funds because of their relevance in the equity market segment based on the number of funds and money managed. Fund classification is based on self-stated investment objectives declared by the fund and reported to the supervisory entity (CNMV).

Our final sample spans the nine-year period from January 2001 to December 2009 ($K = 108$ months and 2235 daily observations) and includes $N = 144$ funds. Consequently, it covers the most important years of the development of the Spanish fund industry and different market scenarios (bull and bear markets). Summary statistics of our sample are shown in Table I.

Table I shows that the median monthly excess return of the sample reaches the value of 0.85%, although the mean value is 0.33%. Additionally, a high standard deviation of the funds' excess returns (5.41%), a high asymmetry (mean value 1.90), and a high leptokurtosis (mean value 8.96) are reported. The last two features are generally observed in our mutual fund sample due to the extreme excess return values achieved by some portfolios (from -38.14% to 22.97%). Market premium also follows similar patterns, although taking less extreme values.

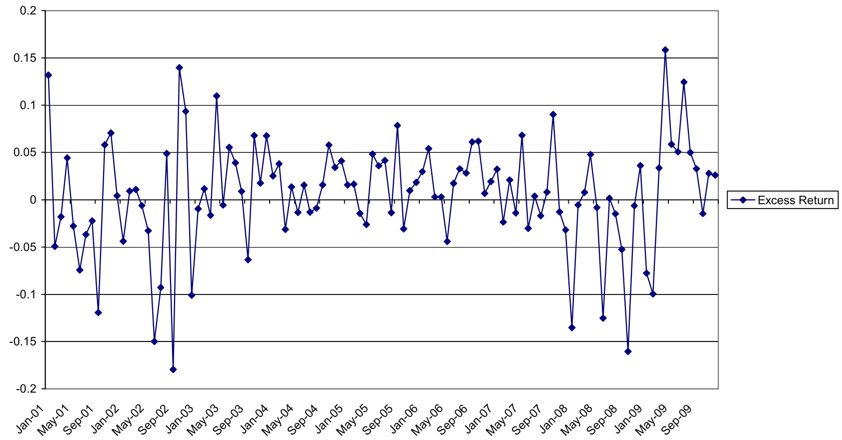
Figure 1 shows the monthly evolution of the market excess return (Figure 1a) and volatility (Figure 1b). Figure 2 shows the distribution of the funds' excess return (Figure 2a) and volatility (Figure 2b) over time. The evolution of these magnitudes allows us to distinguish four different time periods. The first period extends from the beginning of the time period analysed to May 2003 and is characterised by a high volatility and extreme excess return values. The second period spans from June 2003 to December 2007 and presents low volatility, especially around December 2004, and excess return values near to zero. The third period spans from January 2008 to March 2009 and shows similar patterns to those of the first one. Finally, the fourth period spans from April 2009 until the end of the period analysed and presents lower volatility and positive excess return values.

4.2. *Estimation of the style coefficients*

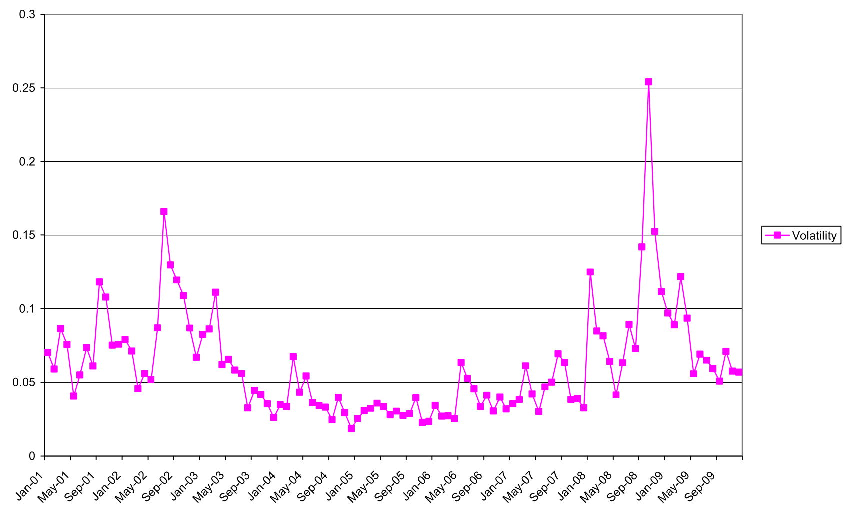
The use of daily return data allows us to collect sufficient observations to estimate the style exposures of each mutual fund in each month without losing the ability to capture reasonably rapid changes in the investment style [28]. The consideration of longer time intervals (usually quarterly in the case of institutional investors) undervalues herding whether this behaviour occurs within shorter periods. On the other hand, shorter periods provide less reliable estimations of the beta coefficients by increasing the probability of detecting spurious herding behaviour. For that reason, we analyse herding behaviour through the monthly evolution of style exposures.

The estimation of the style coefficients $\{(\beta_{1,i,k}, \beta_{2,i,k}, \beta_{3,i,k}); k = 1,\dots,K; i = 1,\dots,N\}$ is performed as described in Section 2, in which we take $\mathbf{b}_0 = (0,1,0,0)'$; $s^2_{\beta,j} = 100$ for j = 1, …, 4; $n_\tau = 0.01$ and $s^2_\tau = 1$ in the prior distribution (2.3)–(2.5) for each mutual fund, which is a fairly non-informative prior.** Furthermore, to avoid large uncertainty in the estimation of style coefficients due to a small window size, we use $g_{min} = 0.5$ and $g_{max} = 1$.

** *We have centred the prior distribution (2.3) on the values predicted by Capital Asset Pricing Model (CAPM). However, an alternative prior with* $\mathbf{b}_0$ *equals to the posterior mean of* $\beta_{i,k-1}$ *provided results very similar to those reported in the paper. These results are available upon request to the authors.*



a: Monthly evolution of the market excess return



b: Monthly evolution of the market excess return volatility

Figure 1. Monthly evolution of the (a) market excess return and (b) market excess return volatility.

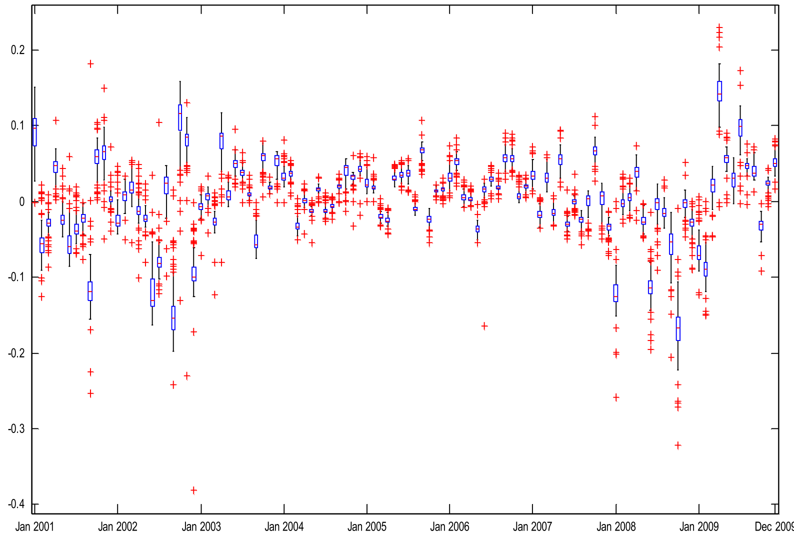
Gibbs sampling was run using 5000 iterations for each month and fund combination, and we discard the first 4000 iterations to achieve convergence. Figure 3 shows the monthly evolution of 2.5, 50, and 97.5 quantiles of the posterior median of the style exposures ($\beta_{1,i,k}$, $\beta_{2,i,k}$, $\beta_{3,i,k}$) of the mutual fund sample.

Figure 3a shows that the median exposure to the market was stable at a level of 0.9 across the time period analysed. Additionally, Figure 3b indicates that the median exposure to the SMB factor was slightly positive for the majority of the mutual funds studied and also increased in the last months of time period. Finally, the median exposure to the HML factor is mainly negative (see Figure 3c).

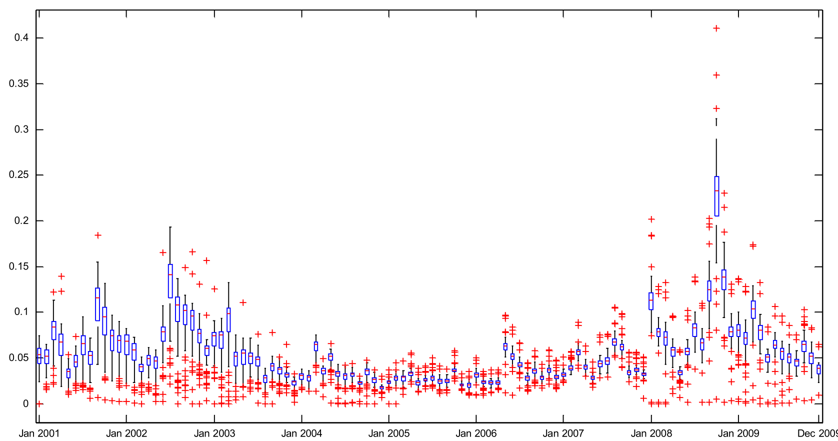
4.3. Herding behaviour considering the LSV measure

Table II shows the results obtained from the application of LSV metric to the coefficients $\{\beta_j;\ j = 1,2,3\}$.[††] In this analysis, we try to detect herd behaviour, considering $\gamma = 0.05$ and 0.01 and using expressions (3.4) and (3.5) to calculate *p*-values (3.2) and (3.3) with $EP^{+}_{j,k} = 0.5$. The table indicates those months with significant herding along

[††] *See Section 3.1 for a detailed description.*



a: Boxplot of the monthly excess return of Spanish mutual funds



b: Boxplot of the monthly excess return volatility of Spanish mutual funds

Figure 2. Boxplot of the (a) monthly excess return of Spanish mutual funds and (b) monthly excess return volatility of Spanish mutual funds.

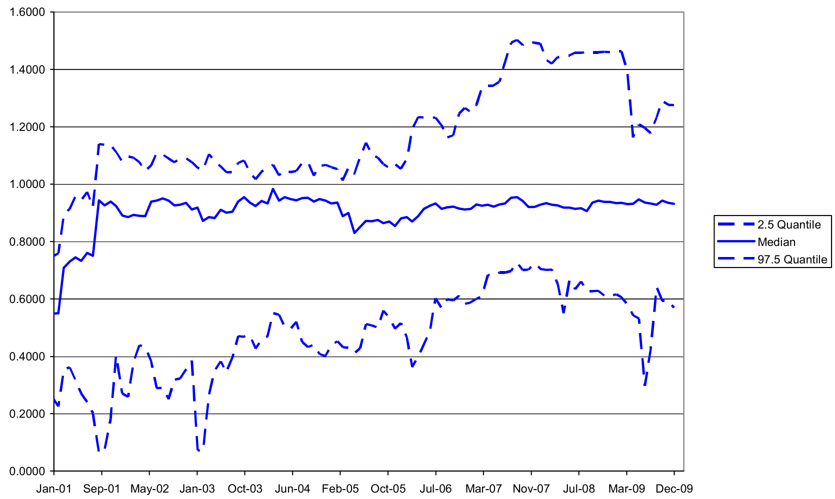
with the direction of that convergence (increase (+) or decrease (-) of the style exposures). Specifically, buying herding situations are detected when (3.3)/(3.2) are lower than 0.05 or 0.01 (signalled with + or ++, respectively), which means that the 'observed' percentage $P^+_{j,k}$ is significantly higher/lower than the 'expected' reference, $EP^+_{j,k}$, 50% in this case. Similarly, selling herding situations are detected when (3.3)/(3.2) are larger than 0.05 or 0.01 (signalled with - or - -, respectively), which means that the 'observed' percentage $P^+_{j,k}$ is significantly lower than $EP^+_{j,k}$.

Table II provides evidence of the number of months with significant herding for the different coefficients analysed. Specifically, 32 months for β_1 (17 months with increasing exposures to the equity market and 15 with decreasing exposures), 32 for β_2 (19 months with increasing exposures in small stocks and 13 with increasing exposures in large stocks), and 30 for β_3 (16 months with increasing exposures in value stocks and 14 in growth stocks).

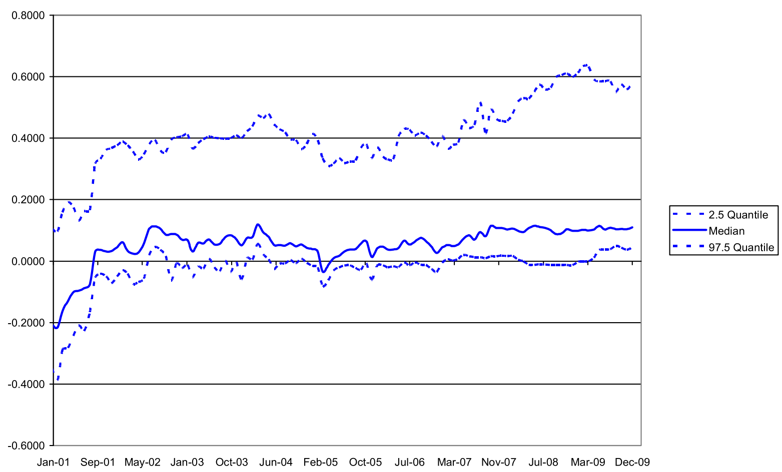
4.4. Analysis of the influence of market conditions on style parameters

In this section, we analyse the influence of the market conditions on the evolution of the estimated style coefficients by means of the regression model (3.7)-(3.8). Later, these results will be used to analyse the existence of herding, taking into account the market conditions (Section 3.2).

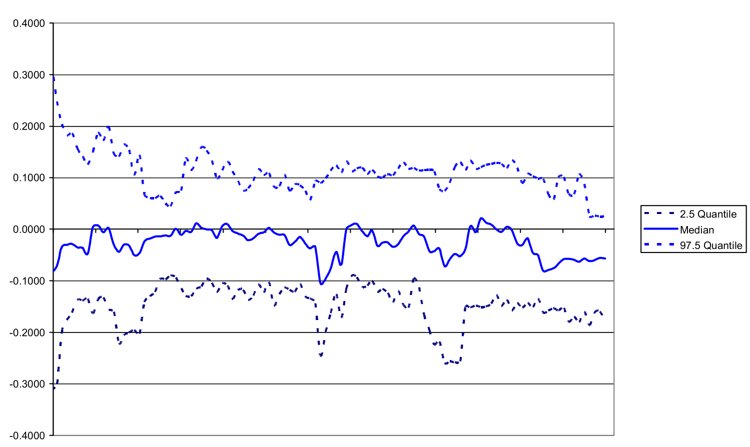
To measure the market conditions, we considered the market excess return and its volatility in upward and downward periods. These variables are available for all managers and can inform them about the risk and the profitability conditions



a: Dynamic evolution of the style exposure β$_1$



b: Dynamic evolution of the style exposure β$_2$



c: Dynamic evolution of the style exposure β$_3$

Figure 3. Dynamic evolution of the style exposure (a) β_1, (b) β_2, and (c) β_3.

Table II. Herding results considering the LSV metric.

Month	β_1(Market)	β_2(SMB)	β_3(HML)	Month	β_1(Market)	β_2(SMB)	β_3(HML)
Mar–01	++	0	0	May–06	++	0	0
Sep–01	++	++	+	Jun–06	++	++	0
Jan–02	−	0	−−	Aug–06	−−	0	0
May–02	0	0	−	Nov–06	0	−−	0
Jun–02	++	+	0	Dec–06	0	−−	−
Jul–02	0	++	0	Jan–07	0	++	0
Oct–02	0	−	0	Feb–07	0	+	−−
Jan–03	0	0	+	May–07	0	++	−−
Mar–03	−	−−	+	Jun–07	0	+	+
Apr–03	0	+	0	Jul–07	0	-	+
Jun–03	+	0	0	Aug–07	++	++	0
Sep–03	++	+	0	Sep–07	0	−	+
Oct–03	0	0	++	Oct–07	−−	++	++
Jan–04	0	+	0	Nov–07	−−	0	−
Mar–04	++	++	0	Jan–08	+	0	++
Apr–04	−−	−−	0	Mar–08	0	−	−−
Jun–04	0	−−	0	May–08	−−	++	0
Oct–04	−	0	0	Jul–08	0	0	−−
Nov–04	0	0	−	Aug–08	0	−	0
Mar–05	−−	−−	0	Sep–08	−	++	+
Apr–05	+	0	0	Oct–08	++	0	−−
May–05	−−	++	−−	Dec–08	−−	0	−−
Jun–05	+	0	0	Mar–09	0	0	+
Jul–05	++	0	+	Apr–09	0	0	++
Aug–05	0	0	++	Jun–09	++	0	0
Oct–05	0	0	++	Jul–09	−	−	0
Dec–05	−−	−−	0	Aug–09	0	0	+
Jan–06	++	++	0	Sep–09	0	0	−
Mar–06	−	0	0	Dec–09	0	+	0
Apr–06	+	0	−−				

+(++):5% (1%) significant buying herding, i.e., there exist a percentage of mutual funds higher than expected (50%) that increased their exposure with posterior p-value (3.3) $\leqslant 0.05$ ($\leqslant 0.01$)
−(−−):5% (1%) significant selling herding, i.e., there exist a percentage of mutual funds higher than expected (50%) that decreased their exposure with posterior p-value (3.2) $\leqslant 0.05$ ($\leqslant 0.01$)
0:herding is not detected, i.e., minimum of (3.2) and (3.3) >0.05

of the market. Concretely, we consider the following covariates:

$$C_{k,1} = I(\overline{RMRF}_k \leqslant 0)$$
$$C_{k,2} = I(\overline{RMRF}_k > 0)$$
$$C_{k,3} = I(\overline{RMRF}_k \leqslant 0)\overline{RMRF}_k$$
$$C_{k,4} = I(\overline{RMRF}_k > 0)\overline{RMRF}_k$$
$$C_{k,5} = I(\overline{RMRF}_k \leqslant 0)LVol_k$$
$$C_{k,6} = I(\overline{RMRF}_k > 0)LVol_k$$

where $\overline{RMRF}_k = \frac{1}{|\mathbf{M}_k|}\sum_{t\in\mathbf{M}_k} RMRF_t$ is the mean excess return of the market portfolio and LVol_k is the logarithm of the volatility of market excess return in month $\mathbf{M}_k$, where $\mathrm{Vol}_k = \frac{1}{|\mathbf{M}_k|-1}\sum_{t\in\mathbf{M}_k}\left(RMRF_t - \overline{RMRF}_k\right)^2$.

Table III shows the estimation results of the coefficients $\{\psi_{j,m};\ m = 1,\dots,6\}$of model (3.7)–(3.8) for each style coefficient $\{\beta_j;\ j = 1,2,3\}$. With respect to the exposure to the market portfolio (β_1), the only significant ψ coefficient corresponds to the covariate $I\left(\overline{RMRF}_k \leqslant 0\right)\overline{RMRF}_k$, which is significantly negative. This result implies that in bearish periods, a decrease in the excess return of the market causes a growth in the percentage of mutual funds, increasing their exposure to the market. This apparent incoherence is collected in the literature as 'perverse' market timing (see, e.g., [29], which demonstrates that the individual stock betas tend to increase/decrease in bear/bull markets).

With respect to the exposure to the factor SMB (β_2), the significant ψ coefficients correspond to the $I\left(\overline{RMRF}_k > 0\right)$, $I\left(\overline{RMRF}_k \leqslant 0\right)\overline{RMRF}_k$ and $I\left(\overline{RMRF}_k > 0\right)\overline{RMRF}_k$ covariates, with the first two being significantly negative and the third significantly positive. There is, therefore, a growth trend of the percentages of mutual fund managers that increases their exposure for extreme levels of the market excess return, both in positive and negative terms. Hence, it seems that portfolio managers tend to increase the proportion of small stocks in their portfolios in extreme market situations, most likely in search of investment opportunities. Finally, we also detect that portfolio managers tend to increase their exposure to factor HML in bull markets. Hence, the higher the excess market returns are, the higher is the proportion of value stocks they include in their portfolios.

According to these results, we can conclude that market returns exert a significant influence on the dynamic evolution of the investment styles of mutual funds. Therefore, the exposures tend to be influenced by the bearish or bullish character of

Table III. Influence of the market conditions on the monthly evolution of $\{P^{+}_{j,k};\ k = 2,..,K\}$under the no-herding hypothesis.

Style coefficient	Variable	Q0.5	Q2.5	Q5	Q50	Q95	Q97.5	Q99.5
β_1(Market)	$I\left(\overline{RMRF}_k \leqslant\right)0$	−0.1894	−0.1590	−0.1417	−0.0364	0.0707	0.0987	0.1398
	$I\left(\overline{RMRF}_k > 0\right)$	−0.1872	−0.1516	−0.1221	−0.0302	0.0597	0.0782	0.1344
	$I\left(\overline{RMRF}_k \leqslant 0\right)\overline{RMRF}_k$	**−0.0591**	**−0.0513**	**−0.0487**	**−0.0292**	**−0.0108**	**−0.0070**	**−0.0031**
	$I\left(\overline{RMRF}_k > 0\right)\overline{RMRF}_k$	−0.0376	−0.0286	−0.0267	−0.0076	0.0105	0.0144	0.0225
	$I\left(\overline{RMRF}_k \leqslant 0\right)LVol_k$	−0.1265	−0.0957	−0.0513	0.1051	0.2542	0.2834	0.3465
	$I\left(\overline{RMRF}_k > 0\right)\mathrm{LVol}_k$	−0.1642	−0.1430	−0.1055	0.0168	0.1384	0.1689	0.2009
β_2(SMB)	$I\left(\overline{RMRF}_k \leqslant 0\right)$	−0.0989	−0.0653	−0.0466	0.0522	0.1407	0.1648	0.2035
	$I\left(\overline{RMRF}_k > 0\right)$	−0.2624	−0.2296	**−0.2044**	**−0.1003**	**−0.0019**	0.0082	0.0504
	$I\left(\overline{RMRF}_k \leqslant 0\right)\overline{RMRF}_k$	−0.0434	−0.0383	**−0.0344**	**−0.0171**	**−0.0014**	0.0020	0.0102
	$I\left(\overline{RMRF}_k > 0\right)\overline{RMRF}_k$	−0.0017	**0.0041**	**0.0083**	**0.0258**	**0.0458**	**0.0502**	0.0579
	$I\left(\overline{RMRF}_k \leqslant 0\right)\mathrm{LVol}_k$	−0.2751	−0.1994	−0.1724	−0.0353	0.1404	0.1737	0.2608
	$I\left(\overline{RMRF}_k > 0)\right)\mathrm{LVol}_k$	−0.1556	−0.1228	−0.1018	0.0080	0.1361	0.1598	0.1999
β_3(HML)	$I\left(\overline{RMRF}_k \leqslant 0\right)$	−0.0370	**0.0026**	**0.0241**	**0.1191**	**0.2154**	**0.2338**	0.2718
	$I\left(\overline{RMRF}_k > 0\right)$	**−0.2996**	**−0.2515**	**−0.2341**	**−0.1413**	**−0.0428**	**−0.0215**	**−0.0032**
	$I\left(\overline{RMRF}_k \leqslant 0\right)\overline{RMRF}_k$	−0.0182	−0.0096	−0.0073	0.0097	0.0267	0.0315	0.0409
	$I\left(\overline{RMRF}_k > 0\right)\overline{RMRF}_k$	−0.0116	−0.0032	**0.0017**	**0.0202**	**0.0376**	0.0413	0.0475
	$I\left(\overline{RMRF}_k \leqslant 0\right)\mathrm{LVol}_k$	−0.3315	−0.2521	−0.2164	−0.0629	0.1013	0.1302	0.1911
	$I\left(\overline{RMRF}_k > 0\right)\mathrm{LVol}_k$	−0.1759	−0.1341	−0.1182	0.0004	0.1231	0.1423	0.1626

We show the posterior median (Q50) and the limits of the 90%, 95%, and 99% Bayesian credibility intervals built from the posterior quantiles 5% (Q5) and 95% (Q95); 2.5% (Q2.5) and 97.5% (Q97.5), and 0.5% (Q0.5) and 99.5% (Q99.5), respectively. Significant coefficients are in bold.

the periods and the magnitude of the market excess return. Meanwhile, the market volatility is not revealed as a significant factor to explain this evolution. In the next section, we go one step further and try to capture 'true' herding, considering the market conditions, which are common to all managers and can therefore lead to similar reactions.

4.5. *Herding behaviour considering market conditions*

Tables IV–VI show the monthly herding results for coefficients $\{\beta_j; j = 1, 2, 3\}$, respectively, using $\gamma = 0.05$ and 0.01. The probability of no herding was calculated as the minimum of (3.2) and (3.3), which were calculated using expressions (3.9) and (3.10), respectively. Additionally, we show the point estimation and the 95% Bayesian credibility intervals of $P^{+}_{j,k}$ calculated from the posterior median (Q50) and the posterior 2.5 (Q2.5) and 97.5 (Q97.5) quantiles of the posterior

Table IV. Periods and types of herding detected for the market style exposures $\{\beta_{1,t}; t = 1, \ldots, T\}$.

	Herding results		Observed percentages			Expected percentages		No herding probability	
Month	LSV	Regression	Q2.5	Q50	Q97.5	LSV	Regression	LSV	Regression
Mar–01	++	++	65.93%	73.63%	82.42%	50.00%	51.67%	0.0004	0.0009
Sep–01	++	++	78.57%	84.69%	89.80%	50.00%	59.80%	0.0000	0.0000
Jan–02	−	−−	26.80%	37.11%	46.39%	50.00%	53.33%	0.0329	0.0100
Jun–02	++	0	59.78%	67.39%	76.09%	50.00%	61.18%	0.0034	0.1695
Mar–03	−	−−	25.56%	34.44%	44.44%	50.00%	52.18%	0.0158	0.0060
Apr–03	0	+	52.81%	61.80%	70.79%	50.00%	47.31%	0.0603	0.0277
Jun–03	+	+	52.75%	62.64%	71.43%	50.00%	48.27%	0.0477	0.0280
Sep–03	++	+	58.43%	68.54%	77.53%	50.00%	53.31%	0.0058	0.0201
Mar–04	++	++	66.67%	74.19%	81.72%	50.00%	52.01%	0.0001	0.0004
Apr–04	−−	−−	13.98%	21.51%	29.03%	50.00%	48.92%	0.0000	0.0000
Oct–04	−	0	25.81%	36.56%	45.16%	50.00%	48.07%	0.0293	0.0536
Mar–05	−−	−−	14.58%	21.88%	30.21%	50.00%	48.90%	0.0000	0.0000
Apr–05	+	+	56.38%	65.96%	74.47%	50.00%	50.03%	0.0166	0.0158
May–05	−−	−−	12.50%	18.75%	25.00%	50.00%	48.17%	0.0000	0.0000
Jun–05	+	++	56.25%	66.67%	76.04%	50.00%	48.32%	0.0139	0.0075
Jul–05	++	++	61.46%	70.83%	77.08%	50.00%	48.25%	0.0016	0.0007
Dec–05	−−	−	23.47%	32.65%	41.84%	50.00%	48.56%	0.0064	0.0108
Jan–06	++	++	63.27%	71.43%	78.57%	50.00%	48.52%	0.0008	0.0004
Mar–06	−	−	27.45%	36.27%	45.10%	50.00%	48.91%	0.0263	0.0378
Apr–06	+	+	54.90%	63.73%	72.55%	50.00%	48.88%	0.0235	0.0155
May–06	++	++	64.71%	72.55%	80.39%	50.00%	52.79%	0.0001	0.0008
Jun–06	++	++	62.38%	70.30%	77.23%	50.00%	48.93%	0.0011	0.0007
Aug–06	−−	−−	19.00%	27.00%	36.00%	50.00%	48.54%	0.0003	0.0007
Apr–07	0	+	53.47%	61.39%	68.32%	50.00%	49.52%	0.0530	0.0467
Aug–07	++	++	66.00%	73.00%	81.00%	50.00%	51.02%	0.0002	0.0002
Oct–07	−−	−−	24.00%	31.00%	40.00%	50.00%	47.46%	0.0030	0.0084
Nov–07	−−	−−	19.00%	27.00%	34.00%	50.00%	49.25%	0.0001	0.0003
Jan–08	+	0	56.44%	63.37%	71.29%	50.00%	61.04%	0.0163	0.3515
May–08	−−	−−	24.75%	32.67%	41.58%	50.00%	49.06%	0.0054	0.0082
Sep–08	−	−−	29.90%	38.14%	46.39%	50.00%	55.45%	0.0415	0.0050
Oct–08	++	0	65.98%	73.20%	80.41%	50.00%	64.42%	0.0001	0.0954
Dec–08	−−	−	23.96%	32.29%	41.67%	50.00%	48.91%	0.0079	0.0103
Jun–09	++	++	58.24%	65.93%	74.73%	50.00%	48.44%	0.0087	0.0053
Jul–09	−	0	28.89%	37.78%	46.67%	50.00%	47.06%	0.0465	0.1072

LSV columns report the herding results obtained with the LSV measure, whereas Regression columns report the herding results considering the market conditions. No herding probabilities have been calculated as the minimum of (3.2) and (3.3). Additionally, we show the months with significant herding levels according to the following notation:
+ (++):5% (1%) significant buying herding, i.e., there exist a percentage of mutual funds higher than expected that increased their exposure with posterior p-value (3.3) $\leqslant 0.05$ ($\leqslant 0.01$) 5% (1%)
−(−−): 5% (1%) significant selling herding, i.e., there exist a percentage of mutual funds higher than expected that decreased their exposure with posterior p-value (3.2) $\leqslant 0.05$ ($\leqslant 0.01$)
0:herding is not detected, i.e., minimum of (3.2) and (3.3) > 0.05.

Table V. Periods and types of herding detected for the SMB style exposures $\{\beta_{2,t}; t = 1, \dots, T\}$.

	Herding procedures		Observed percentages			Expected percentages		No herding probability	
Month	LSV	Regression	Q2.5	Q50	Q97.5	LSV	Regression	LSV	Regression
Sep–01	++	++	71.43%	80.10%	87.76%	50.00%	55.70%	0.0000	0.0001
Jun–02	+	0	56.52%	65.22%	73.91%	50.00%	57.21%	0.0177	0.1445
Jul–02	++	++	68.13%	75.82%	83.52%	50.00%	54.31%	0.0000	0.0005
Sep–02	0	−	31.87%	41.76%	50.55%	50.00%	58.20%	0.1363	0.0116
Oct–02	−	−−	27.78%	36.67%	46.67%	50.00%	56.79%	0.0464	0.0053
Mar–03	−−	−−	20.00%	27.78%	36.67%	50.00%	51.35%	0.0008	0.0004
Apr–03	+	+	58.43%	67.42%	76.40%	50.00%	54.73%	0.0103	0.0472
Sep–03	+	0	56.18%	66.29%	74.16%	50.00%	54.08%	0.0182	0.0603
Dec–03	0	−	27.66%	38.30%	47.87%	50.00%	51.77%	0.0564	0.0369
Jan–04	+	+	54.84%	64.52%	72.04%	50.00%	49.06%	0.0247	0.0192
Mar–04	++	++	65.59%	74.19%	82.80%	50.00%	52.38%	0.0001	0.0005
Apr–04	−−	−−	21.51%	29.03%	37.63%	50.00%	48.33%	0.0020	0.0042
Jun–04	−−	−−	22.58%	31.18%	40.86%	50.00%	48.42%	0.0040	0.0076
Mar–05	−−	−−	12.50%	17.71%	23.96%	50.00%	52.21%	0.0000	0.0000
May–05	++	++	62.50%	70.83%	79.17%	50.00%	50.56%	0.0010	0.0011
Dec–05	−−	−−	13.27%	20.41%	28.57%	50.00%	48.52%	0.0000	0.0000
Jan–06	++	++	61.22%	70.41%	78.57%	50.00%	49.35%	0.0015	0.0012
Jun–06	++	++	65.35%	73.27%	81.19%	50.00%	48.62%	0.0001	0.0001
Nov–06	−−	−−	18.81%	26.73%	34.65%	50.00%	47.88%	0.0001	0.0005
Dec–06	−−	−−	19.80%	27.72%	36.63%	50.00%	48.65%	0.0007	0.0015
Jan–07	++	++	62.63%	71.72%	79.80%	50.00%	49.52%	0.0007	0.0006
Feb–07	+	0	53.54%	61.62%	70.71%	50.00%	52.48%	0.0472	0.0967
May–07	++	++	63.37%	71.29%	80.20%	50.00%	51.83%	0.0006	0.0017
Jun–07	+	0	53.54%	62.63%	71.72%	50.00%	52.62%	0.0367	0.0789
Jul–07	−	−	26.26%	34.34%	42.42%	50.00%	47.73%	0.0107	0.0242
Aug–07	++	++	64.00%	72.00%	79.00%	50.00%	51.73%	0.0003	0.0012
Sep–07	−	−	26.00%	35.00%	43.00%	50.00%	48.08%	0.0124	0.0260
Oct–07	++	++	70.00%	77.00%	84.00%	50.00%	53.34%	0.0000	0.0001
Nov–07	0	−	30.00%	40.00%	48.00%	50.00%	52.01%	0.0742	0.0422
Mar–08	−	−	29.41%	36.27%	44.12%	50.00%	48.12%	0.0228	0.0437
May–08	++	++	61.39%	70.30%	77.23%	50.00%	51.76%	0.0013	0.0030
Aug–08	−	−	28.28%	36.36%	45.45%	50.00%	51.61%	0.0273	0.0156
Dec–08	++	++	60.42%	67.71%	76.04%	50.00%	50.05%	0.0041	0.0043
Jul–09	−	−−	25.56%	34.44%	42.22%	50.00%	55.69%	0.0117	0.0010
Dec–09	+	+	53.01%	62.65%	73.49%	50.00%	49.23%	0.0462	0.0371

LSV columns report the herding results obtained with the LSV measure, whereas Regression columns report the herding results considering the market conditions. No herding probabilities have been calculated as the minimum of (3.2) and (3.3). Additionally, we show the months with significant herding levels according to the following notation:
+(++) :5% (1%) significant buying herding, i.e., there exist a percentage of mutual funds higher than expected that increased their exposure with posterior *p*-value (3.3) $\leqslant 0.05$ ($\leqslant 0.01$)
− (−−):5% (1%) significant selling herding, i.e., there exist a percentage of mutual funds higher than expected that decreased their exposure with posterior *p*-value (3.2) $\leqslant 0.05$ ($\leqslant 0.01$).
0: herding is not detected, i.e., minimum of (3.2) and (3.3) > 0.05.

sample $\left\{ P^{+,(s)}_{j,k} = \frac{\left| \left\{ i \in I_k : \beta^{(s)}_{j,i,k} > \beta^{(s)}_{j,i,k-1} \right\} \right|}{|I_k|}; s = 1, \dots, S \right\}$, and the expected percentage, $EP^{+}_{j,k}$, which equals to 0.5 when using LSV metric, and it is calculated from the binomial regression model (3.7)–(3.8), when considering the market conditions, by means of the expression $\frac{1}{S}\sum_{s=1}^{S} EP^{+,(s)}_{j,k}$ with $EP^{+,(s)}_{j,k}$ calculated using (3.11). These tables reveal that when significant evidence of buying/selling herding is detected, the expected percentage $EP^{+}_{j,k}$ takes higher/lower values than the limits of the 95% Bayesian credibility intervals of $P^{+}_{j,k}$, which confirms the validity of the proposed procedure.

This analysis shows similar results to those previously obtained using the LSV metric. Both methods tend to detect convergence over the same time periods and in the same direction (convergent movements of increasing or

Table VI. Periods and types of herding detected for the HML style exposures $\{\beta_{3,t}; t = 1, \dots, T\}$.

	Herding procedures		Observed percentages			Expected percentages		No herding probability	
Month	LSV	Regression	Q2.5	Q50	Q97.5	LSV	Regression	LSV	Regression
Sep–01	+	+	53.06%	62.24%	71.43%	50.00%	48.89%	0.0440	0.0310
Jan–02	−−	−−	22.68%	31.96%	40.21%	50.00%	51.34%	0.0040	0.0022
May–02	−	−	27.17%	36.96%	46.74%	50.00%	52.23%	0.0407	0.0213
Jan–3	+	0	53.41%	62.50%	71.59%	50.00%	52.11%	0.0403	0.0734
Mar–03	+	0	53.33%	62.22%	70.00%	50.00%	51.51%	0.0482	0.0756
Sep–03	0	−	28.09%	38.20%	47.19%	50.00%	51.70%	0.0565	0.0335
Oct–03	++	++	58.43%	67.42%	77.53%	50.00%	49.86%	0.0096	0.0090
Aug–04	0	−	30.11%	38.71%	48.39%	50.00%	53.46%	0.0698	0.0256
Nov–04	−	−	26.60%	35.11%	44.68%	50.00%	48.19%	0.0244	0.0436
May–05	−−	−−	13.54%	19.79%	26.04%	50.00%	48.89%	0.0000	0.0000
Jun–05	0	+	51.04%	61.46%	69.79%	50.00%	48.28%	0.0686	0.0420
Jul–05	+	+	55.21%	63.54%	71.88%	50.00%	48.55%	0.0260	0.0160
Aug–05	++	+	57.58%	65.66%	74.75%	50.00%	53.62%	0.0082	0.0350
Sep–05	0	−	28.57%	38.78%	47.96%	50.00%	50.39%	0.0521	0.0446
Oct–05	++	++	72.45%	80.61%	86.73%	50.00%	52.67%	0.0000	0.0000
Apr–06	−−	−−	21.57%	29.41%	37.25%	50.00%	46.64%	0.0008	0.0046
Nov–06	0	+	50.50%	60.40%	69.31%	50.00%	46.83%	0.0767	0.0316
Dec–06	−	0	27.72%	36.63%	45.54%	50.00%	47.45%	0.0272	0.0607
Feb–07	−−	−−	22.22%	30.30%	37.37%	50.00%	52.89%	0.0015	0.0003
May–07	−−	−−	18.81%	27.72%	34.65%	50.00%	49.87%	0.0002	0.0002
Jun–07	+	+	55.56%	64.65%	72.73%	50.00%	52.44%	0.0203	0.0489
Jul–07	+	+	53.54%	62.63%	70.71%	50.00%	46.68%	0.0416	0.0117
Sep–07	+	+	54.00%	63.00%	71.00%	50.00%	46.89%	0.0367	0.0120
Oct–07	++	++	60.00%	70.00%	77.00%	50.00%	50.97%	0.0027	0.0040
Nov–07	−	−	29.00%	37.00%	46.00%	50.00%	53.14%	0.0389	0.0142
Dec–07	++	++	64.36%	72.28%	79.21%	50.00%	52.93%	0.0003	0.0017
Mar–08	−−	−	27.45%	34.31%	43.14%	50.00%	46.88%	0.0099	0.0316
Jul–08	−−	−−	19.39%	27.55%	35.71%	50.00%	46.57%	0.0004	0.0025
Sep–08	+	+	56.70%	64.95%	74.23%	50.00%	50.26%	0.0136	0.0171
Oct–08	−−	−−	18.56%	26.80%	34.02%	50.00%	46.77%	0.0002	0.0013
Dec–08	−−	−−	12.50%	18.75%	25.00%	50.00%	48.28%	0.0000	0.0000
Mar–09	+	+	55.79%	64.21%	72.63%	50.00%	48.16%	0.0244	0.0120
Apr–09	++	+	58.51%	67.02%	75.53%	50.00%	54.34%	0.0089	0.0447
Aug–09	+	+	53.33%	62.22%	70.00%	50.00%	48.96%	0.0467	0.0342
Sep–09	−	0	27.06%	36.47%	47.06%	50.00%	48.12%	0.0458	0.0753

LSV columns report the herding results obtained with the LSV measure, whereas Regression columns report the herding results considering the market conditions. No herding probabilities have been calculated as the minimum of (3.2) and (3.3). Additionally, we show the months with significant herding levels according to the following notation:
+ (++):5% (1%) significant buying herding, i.e., there exist a percentage of mutual funds higher than expected that increased their exposure with posterior *p*-value (3.3) $\leqslant 0.05$ ($\leqslant 0.01$)
− (−−):5% (1%) significant selling herding, i.e., there exist a percentage of mutual funds higher than expected that decreased their exposure with posterior *p*-value (3.2) $\leqslant 0.05$ ($\leqslant 0.01$)
0:herding is not detected, i.e., minimum of (3.2) and (3.3) > 0.05.

decreasing the beta) because most of $EP^+_{j,k}$ values tend to have values not too different from 50% (the expected value of the LSV metric).

Specifically, we observe that our alternative approach confirms the existence of herding in β_1 in 27 out of the 32 months previously detected by the LSV herding measure. Meanwhile, in the remaining 5 months, the apparent herding is explained by the market conditions. On the other hand, our market condition approach finds herding in two additional months. Regarding β_2 (β_3), we find that in 28 (26) of the 32 (30) previously detected by the LSV herding measure, the alternative approach confirms the existence of herding.

These consistent findings show that the LSV measure detects convergence patterns that can be partially explained by the identical responses of managers to the same financial information. On the other hand, our alternative approach goes a step further, refining these convergence patterns and discarding those similar movements because of the concrete market conditions in each month.

5. Conclusions

This paper proposes a new statistical methodology to analyse the existence of herding behaviour in the investment style followed by mutual funds. The methodology is based on a dynamic calculation of the difference between the percentage of mutual funds that change the style in comparison with the expected percentage under the assumption of no herding. Additionally, we refine the procedure taking into account the market conditions to dynamically determine the herding level. This improvement, to the best of our knowledge, has not been previously proposed in the literature.

The starting point was the three-factor model of Fama and French [18] estimated by a Bayesian rolling window procedure that allows a user to dynamically select the length of the estimation window through weighted likelihood functions that discount the loss of information because of time. Using the estimations of the style coefficients, two posterior *p*-values were calculated to examine the existence of any significant difference between the observed and expected mutual fund behaviour under the hypothesis of no herding. Unlike other methodologies proposed in the literature, our procedure addresses the uncertainty associated with the estimation of the style coefficients and, therefore, increases the reliability of the results. In future research, it would be interesting to extend our analysis using other ways of building the weights in the Bayesian procedure, such as, for instance, hyperbolic [30] or more general schemes [21, 22].

Finally, an illustrative application is provided on the analysis of the dynamic evolution of Spanish mutual fund style exposures. Our results show that managerial decisions about investment styles are significantly influenced by market conditions and, more concretely, by the market excess return. Additionally, the results show the existence of significant herding behaviour over several months. With this study, we confirm the existence of herding, considering the univariate movements in the style coefficients of Spanish mutual funds. Future research could focus on multivariate movements.

Appendix A. Full conditional distributions of the posterior distribution (2.6)

The appendix shows the full conditional distributions of the posterior distribution (2.6), which are necessary to carry out the Gibbs sampling algorithm referenced in section 2. Most of them are standard and are given by the following:

(a) $\beta_{i,k}\ |\mathbf{Y}_{i,k}, \mathbf{X}_k, \gamma_{i,k}, \tau_{i,k} \sim N_4(\mathbf{m}_{i,k}, \mathbf{S}_{i,k})$ where

$$\mathbf{m}_{i,k} = \mathbf{S}_{i,k}\left(\tau_{i,k}\mathbf{X}_k'\mathbf{D}_{i,k}\mathbf{Y}_{i,k} + \mathbf{S}_\beta^{-1}\mathbf{b}_0\right) \text{ and } \mathbf{S}_{i,k} = \left(\mathbf{S}_\beta^{-1} + \tau_{i,k}\mathbf{X}_k'\mathbf{D}_{i,k}\mathbf{X}_k\right)^{-1}$$

and $\mathbf{D}_{i,k} = diag\left(\left(\gamma_{i,k}(1)\mathbf{1}'_{|\mathbf{M}_1\cap\mathbf{T}_i|}, \ldots, \gamma_{i,k}(k)\mathbf{1}'_{|\mathbf{M}_k\cap\mathbf{T}_i|}\right)\right)$

(b) $\tau_{i,k}\ |\mathbf{Y}_{i,k}, \mathbf{X}_k, \gamma_{i,k}, \beta_{i,k} \sim$

$$Gamma\left(\frac{n_\tau + \left|\mathbf{M}^i_{\leqslant k}\right|}{2}, \frac{n_\tau s_\tau^2 + \sum_{j=1}^{k}\gamma_{i,k}(j)\sum_{t\in\mathbf{M}_j\cap\mathbf{T}_i}\left(y_{i,t} - \beta_{0,i,k} - \beta_{1,i,k}RMRF_t - \beta_{2,i,k}SMB_t - \beta_{3,i,k}HML_t\right)^2}{2}\right)$$

(c) $\gamma_{i,k}\ |\mathbf{Y}_{i,k}, \mathbf{X}_k, \beta_{i,k}, \tau_{i,k}$ is a continuous distribution in (g_{min}, g_{max}) with density proportional to

$$\tau_{i,k}^{\frac{\left|\mathbf{M}^i_{\leqslant k}\right|}{2}}\exp\left[-\frac{\tau_{i,k}}{2}\sum_{j=1}^{k}\gamma_{i,k}(j)\sum_{t\in\mathbf{M}_j\cap\mathbf{T}_i}\left(y_{i,t} - \beta_{0,i,k} - \beta_{1,i,k}RMRF_t - \beta_{2,i,k}SMB_t - \beta_{3,i,k}HML_t\right)^2\right]$$

This distribution is not standard, and we use a Hastings–Metropolis step with the Uniform(g_{min}, g_{max}) as transition distribution.

Acknowledgements

Laura Andreu and José Luis Sarto acknowledge the financial support provided by the Spanish Ministry of Economy and Competitiveness Grant (ECO2013-45568-12). Pilar Gargallo and Manuel Salvador were partially supported by grants CSO2011-29943-CO3-02 and ECO2012-35029 of the Spanish Ministry of Science and Innovation.

References

1. Kahnemann D, Tversky A. Prospect theory: an analysis of decisions under risk. *Econometrica* 1979; **47**:262–291.
2. Tversky A, Kahnemann D. Rational choice and the framing of decisions. *Journal of Business* 1986; **59**:251–278.
3. Hirshleifer D, Teoh SH. Herd behaviour and cascading in capital markets: a review and synthesis. *European Financial Management* 2003; **9**: 25–66.
4. Bikhchandani S, Sharma S. Herd behaviour in financial markets. *IMF Staff Papers* 2000; **47**:279–310.
5. Frazzini A, Lamont O. Dumb money: mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics* 2008; **88**: 299–322.
6. Barberis N, Shleifer A. Style investing. *Journal of Financial Economics* 2003; **68**:161–199.
7. Scharfstein DS, Stein JC. Herd behavior and investment. *American Economic Review* 1990; **80**:465–479.
8. Trueman B. Analyst forecasts and herding behaviour. *Review Financial Studies* 1994; **7**:97–124.
9. Banerjee A. A simple model of herd behavior. *Quarterly Journal of Economics* 1992; **107**:797–817.
10. Bikhchandani S, Hirshleifer D, Welch I. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* 1992; **100**:992–1026.
11. Hirshleifer D, Subrahmanyam A, Titman S. Security analysis and trading patterns when some investors receive information before others. *Journal of Finance* 1994; **49**:1665–1698.
12. Lakonishok J, Shleifer A, Vishny RW. The impact of institutional trading on stock prices. *Journal of Financial Economics* 1992; **32**:23–43.
13. Grinblatt M, Titman S, Wermers R. Momentum investment strategies, portfolio performance, and herding: a study of mutual fund behaviour. *American Economic Review* 1995; **85**:1088–1105.
14. Choe H, Kho BC, Stulz RM. Do foreign investors destabilize stock markets? The Korean experience in 1997. *Journal of Financial Economics* 1999; **54**:227–264.
15. Wermers R. Mutual fund herding and the impact on stock prices. *Journal of Finance* 1999; **54**:581–622.
16. Teo M, Woo S. Style effects in the cross-section of stock returns. *Journal of Financial Economics* 2004; **74**:367–398.
17. Choi N, Sias RW. Institutional industry herding. *Journal of Financial Economics* 2009; **94**:469–491.
18. Fama EF, French KR. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 1993; **33**:3–56.
19. Hadi AS, Luceño A. Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. *Computational Statistics and Data Analysis* 1997; **25**:251–272.
20. Cheng T-C. Robust regression diagnostics with data transformations. *Computational Statistics and Data Analysis* 2005; **49**(3):875–891.
21. Paolella MS, Steude SC. Risk prediction: a DWARF-like approach. *The Journal of Risk Model Validation* 2008; **2**(1):25–43.
22. Steude SC. *Weighted Maximum Likelihood for Risk Prediction*. University of Zurich: NCCR FINRISK 689, 2011.
23. Newton MA, Raftery AE. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society Series B* 1994; **1**:3–48.
24. West M. Bayesian model monitoring. *Journal of the Royal Statistical Society Series B* 1986; **48**(1):70–78.
25. Robert CP, Casella G. *Monte Carlo Statistical Methods* Second Edition. Springer: New York, 2004.
26. Meng X-L. Posterior predictive p-values. *The Annals of Statistics* 1994; **22**(3):1142–1160.
27. Ferruz L, Sarto JL, Vicente L. Herding behaviour in Spanish equity funds. *Applied Economic Letters* 2008; **15**:573–576.
28. Hwang S, Salmon M. Market stress and herding. *Journal of Empirical Finance* 2004; **11**:585–616.
29. Matallín JC, Moreno D, Rodríguez R. Why is timing perverse? *The European Journal of Finance* 2014. published online July 2014.
30. Mittnik S, Paolella MS. Conditional density and value-at-risk prediction of Asian currency exchange rates. *Journal of Forecasting* 2000; **19**: 313–333.