

A Model-Free Measure of Aggregate Idiosyncratic Volatility and the Prediction of Market Returns

René Garcia, Daniel Mantilla-García, and Lionel Martellini*

Abstract

In this paper, we formally show that the cross-sectional variance of stock returns is a consistent and asymptotically efficient estimator for aggregate idiosyncratic volatility. This measure has two key advantages: It is model free and observable at any frequency. Previous approaches have used monthly model-based measures constructed from time series of daily returns. The newly proposed cross-sectional volatility measure is a strong predictor for future returns on the aggregate stock market at the daily frequency. Using the cross section of size and book-to-market portfolios, we show that the portfolios' exposures to the aggregate idiosyncratic volatility risk predict the cross section of expected returns.

I. Introduction

The recent financial literature has paid considerable attention to idiosyncratic volatility. Campbell, Lettau, Malkiel, and Xu (2001) and Malkiel and Xu (2002) document that idiosyncratic volatility increased over time, while Brandt, Brav, Graham, and Kumar (2010) show that this trend completely reversed itself by 2007, falling below pre-1990s levels, and suggest that the increase in idiosyncratic volatility through the 1990s was not a time trend but rather an “episodic phenomenon.” Bekaert, Hodrick, and Zhang (2012) confirm that there is no trend

*Garcia, rene.garcia@edhec.edu, EDHEC Business School, 393 Promenade des Anglais, BP 3116, Nice 06202, France and CIREQ and CIRANO; Mantilla-García, daniel@optimalam.com, Research Department, Optimal Asset Management, 171 Main St # 298, Los Altos, CA 94022, and EDHEC-Risk Institute; and Martellini, lionel.martellini@edhec.edu, EDHEC Business School, 393 Promenade des Anglais, BP 3116, Nice 06202, France. A previous version of this paper circulated under the title: “Idiosyncratic Risk and the Cross-Section of Stock Returns.” We are grateful to Turan Bali (associate editor and referee) for useful and insightful comments. We also thank Hendrik Bessembinder (the editor), Michel Crouhy, Robert Engle, Stéphane Gregoir, Abraham Lioui, Harrison Hong, Frederic Smadja, Stoyan Stoyanov, George Tauchen, Stéphane Tyc, Vijay Vaidyanathan, and Volker Ziemann, as well as seminar and conference participants at EDHEC Business School, the 2010 Financial Econometrics Conference (Toulouse), the 2010 Financial Econometrics Workshop (Singapore Management University), and the 2011 Conference in Asset and Risk Management in the Aftermath of the Financial Crisis (HEC Lausanne), for useful discussions and comments. Much of this paper was written while Mantilla-García was at Koris International. The views expressed in this paper are those of the authors and not of Koris International or Optimal Asset Management. This research has benefited from the support of the chair, “Produits Structurés et Produits Dérivés,” Fédération Bancaire Française.

both for the United States and other developed countries. A second fact about idiosyncratic volatility is also a source of contention. Goyal and Santa-Clara (GS) (2003) put forward that idiosyncratic volatility has forecasting power for future excess returns, while Bali, Cakici, Yan, and Zhang (2005) and Wei and Zhang (2005) find that the positive relation is not robust to the sample chosen. Finally, while some economic theories suggest that idiosyncratic volatility should be positively related to expected returns, Ang, Hodrick, Xing, and Zhang (2006) find that stocks with high idiosyncratic volatility have low average returns.

An underlying issue in all these studies is the measurement of idiosyncratic volatility. Campbell et al. (2001) use a value-weighted (VW) sum of individual firm idiosyncratic variances, computed as the variances of residuals of differences between individual firm returns and the return of an industry portfolio to which the firm belongs.¹ In addition to this measure, Bekaert et al. (2012) use the individual firm residuals of a standard Fama and French (FF) (1993) 3-factor model to compute a VW aggregate idiosyncratic volatility.²

We revisit the issues regarding the dynamics and forecasting power of idiosyncratic variance by using instead the cross-sectional dispersion of stock returns. Through central-limit arguments, we provide the formal conditions under which the cross-sectional variance (CSV) of stock returns asymptotically converges toward the average idiosyncratic variance.³ One key advantage of this measure is obviously its observability at any frequency, while the previous approaches have used monthly measures based on time series of daily returns. A second important feature is that this measure is model free, since we do not need to obtain residuals from a particular model to compute it.

We confirm empirically that the CSV is an excellent proxy for the idiosyncratic variance obtained from the capital asset pricing model (CAPM) or the FF (1993) models, as was done in the previous literature. Correlations between the CSV measure and the model-based measures estimated monthly are always above 99%, whether we consider equal-weighted (EW) or market capitalization-weighted (CW) measures of idiosyncratic variance.

On the debate about predictability of aggregate returns by the idiosyncratic variance, we first verify empirically that the CSV measure leads to the same conclusions that other studies (in particular, GS (2003) and Bali et al. (2005)) have reported at the monthly frequency. Then, we report new results at the daily frequency. Specifically, we show that the predictive power of idiosyncratic volatility is much stronger both quantitatively and statistically at the daily frequency than at the monthly frequency. This relation is robust to the inclusion of index return variance, option-implied variance, and index-realized variance as additional variables in the predictive regressions.

We find that the relation is much stronger and stable across periods between the equal-weighted measure of aggregate idiosyncratic volatility and the

¹This amounts to imposing unit beta restrictions in an industry-market model.

²This is also the approach followed in Ang et al. (2006).

³GS (2003) argue informally that their measure can be interpreted as a measure of cross-sectional dispersion of stock returns, but they do not establish a formal link between the two measures. Allen and Bali (2007) use cross-sectional moments of equity returns for financial institutions to measure catastrophic and operational risks.

returns on the equal-weighted index than for the market-cap-weighted equivalents. Economic sources of heterogeneity between firms, as diverse as they can be, are better reflected in an equal-weighted measure, all other things being equal. This argument is consistent with previous findings of Bali et al. (2005), who argue that the relation between equal-weighted average idiosyncratic risk and the market-cap-weighted index on the sample ending in 1999:12 is mostly driven by small stocks traded in the NASDAQ.

Finally, we unveil an asymmetry in the relation between idiosyncratic variance and returns and show that the predictive power of specific risk is substantially increased when a cross-sectional measure of idiosyncratic skewness is added as an explanatory variable.

Using a Fama-MacBeth (1973) procedure with several sets of portfolios, we find support for a positive and significant price of risk for the exposure to the idiosyncratic variance risk. This is consistent with the asset pricing literature, since most theories support a positive relation between idiosyncratic risk and expected returns (see, in particular, Levy (1978), Merton (1987), Malkiel and Xu (2002), Campbell et al. (2001), and Guo and Savickas (2008)).

The remainder of the paper is organized as follows: In Section II, we provide a formal argument for choosing the CSV of returns as a measure of average idiosyncratic volatility, explore its asymptotic and finite-distance properties, and compare it to other measures formerly selected in the literature. Section III provides an empirical implementation of the concept, again in comparison with other measures, and studies its relation with macroeconomic variables. In Section IV, we provide new results on the predictability of returns by idiosyncratic volatility, as well as idiosyncratic skewness. Section V focuses on the analysis of the cross-sectional relation between idiosyncratic risk and expected returns. Section VI concludes, and Appendixes A, B, and C collect proofs.

II. The CSV as a Measure of Idiosyncratic Variance

Let N_t be the total number of stocks in a given universe at day t , and assume with no loss of generality a conditional single-factor model for excess stock returns.⁴ That is, we assume that for all $i = 1, \dots, N_t$, the return on stock i in excess of the risk-free rate can be written as

$$(1) \quad r_{it} = \beta_{it}F_t + \varepsilon_{it},$$

where F_t is the factor excess return at time t , β_{it} is the beta of stock i at time t , and ε_{it} is the residual, with $E(\varepsilon_{it}) = 0$ and $\text{cov}(F_t, \varepsilon_{it}) = 0$. We assume that the factor model under consideration is a strict factor model, that is, $\text{cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0$ for $i \neq j$.⁵

⁴Assuming a single-factor structure is done for simplicity of exposure only and the results below can easily be extended to a multiple-factor setting.

⁵This assumption is made in the single index or diagonal model of Sharpe (1963) and in the derivation of the arbitrage pricing theory (APT) in Ross (1976). The very definition of idiosyncratic risk relies on the assumption of orthogonal residuals: Assuming that the model is the "true"

Given T observations of the stock returns and the factor returns, one can use the residuals of the regression to obtain a measure of the idiosyncratic variance of asset i by $\sigma_i^2 = (1/T) \sum_{t=1}^T \varepsilon_{it}^2$. An aggregate measure of idiosyncratic variance over the T observations (say, a month) can be obtained by averaging across assets such individual idiosyncratic variance estimates. This is the approach that has been followed by most related papers with observations of the returns at a daily frequency to compute monthly idiosyncratic variances.

We propose instead to measure at each time t the CSV of observed stock returns. Using formal central-limit arguments, we show that, under mild simplifying assumptions, this cross-sectional measure provides a very good approximation of the average idiosyncratic variance. Moreover, it offers two main advantages: It can be computed directly from observed returns, with no need to estimate other parameters such as betas, and it is readily available at any frequency and for any universe of stocks.

A. Measuring the CSV

Let $(w_t)_{t \geq 0}$ be a given weight vector process. The return on the portfolio defined by the weight vector process (w_t) is denoted as $r_t^{(w_t)}$ and given by

$$(2) \quad r_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} r_{it}.$$

We restrict our attention to nontrivial weighting schemes, ruling out situations such that the portfolio is composed of a single stock. We also restrict the weights to be positive at every point in time. Hence, a weighting scheme (w_t) is a vector process that satisfies $0 < w_{it} < 1 \forall i, t$. The CSV measure is defined as follows:

Definition: CSV. The *cross-sectional variance* measure under the weighting scheme (w_t) , denoted as $\text{CSV}_t^{(w_t)}$, is given by

$$(3) \quad \text{CSV}_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} \left(r_{it} - r_t^{(w_t)} \right)^2.$$

A particular case of interest is the equal-weighted CSV, denoted by CSV_t^{EW} and corresponding to the weighting scheme $w_{it} = 1/N_t \forall i, t$:

$$(4) \quad \text{CSV}_t^{\text{EW}} = \frac{1}{N_t} \sum_{i=1}^{N_t} \left(r_{it} - r_t^{\text{EW}} \right)^2,$$

where r_t^{EW} is the return on the equal-weighted portfolio.

Another important weighting scheme is the capitalization weighting scheme. If we denote by c_{it} the market capitalization of stock i at the beginning of the month

factor model implies that the “true” idiosyncratic risk is the one measured with respect to that model, which in turn implies that no commonalities should be left after controlling for the common factor exposure.

corresponding to day t , $C_t = \sum_{i=1}^{N_t} c_{it}$ the total market capitalization and r_t^{CW} the return on the market capitalization-weighted portfolio, the cap-weighted CSV, denoted by CSV^{CW} , is defined as:

$$(5) \quad CSV_t^{CW} = \sum_{i=1}^{N_t} w_{it}^{CW} (r_{it} - r_t^{CW})^2,$$

where $w_{it}^{CW} = c_{it}/C_t$.

For any given weighting scheme (in particular, EW or CW), the corresponding cross-sectional measure is readily computable at any frequency from observed returns. This stands in contrast with the previous approaches that have used monthly measures based on time-series regressions on daily returns. The second important feature of the CSV is its model-free nature, since we do not need to specify a particular factor model to compute it.⁶

B. A Formal Relation between CSV and Idiosyncratic Variance

The following proposition establishes a formal link between CSV and idiosyncratic variance. It is an asymptotic result ($N_t \rightarrow \infty$) obtained under the assumptions of homogeneous betas and residual variances across stocks. These assumptions will be relaxed later on.

Proposition 1: CSV as a Proxy for Idiosyncratic Variance (asymptotic results). Assume $\beta_{it} = \beta_t = 1 \forall i$ (homogeneous beta assumption) and $E(\varepsilon_{it}^2) = \sigma_\varepsilon^2(t) \forall i$ (homogeneous residual variance assumption); then, for any strictly positive weighting scheme satisfying for all dates t the condition $\sum_{i=1}^{N_t} w_{it}^2 = O(1/N_t)$, we have that

$$(6) \quad CSV_t^{(w_i)} = \sum_{i=1}^{N_t} w_{it} (r_{it} - r_t^{(w_i)})^2 \xrightarrow{N_t \rightarrow \infty} \sigma_\varepsilon^2(t) \text{ almost surely.}$$

Proof. See Appendix A. \square

This result draws a formal relation between the dynamics of the cross-sectional dispersion of realized returns and the dynamics of aggregate idiosyncratic variance. Note that the asymptotic result $CSV_t^{(w_i)} \rightarrow \sigma_\varepsilon^2(t)$ holds for any weighting scheme that satisfies the condition $\sum_{i=1}^{N_t} w_{it}^2 = O(1/N_t)$.⁷ This condition, which translates into the existence of a constant K_t such that $\sum_{i=1}^{N_t} w_{it}^2 \leq K_t/N_t$, is trivially satisfied for the EW scheme by taking $K_t = 1$ for all t .⁸ Intuitively, the condition rules out cases where portfolios show an extreme level of concentration. Therefore, the case of the CW scheme is more complex, since in

⁶While GS (2003) and Wei and Zhang (2005) consider the equal-weighted CSV in conjunction with other measures, they do not provide a thorough discussion about the conditions under which it can be interpreted as a proxy for idiosyncratic variance nor its empirical validity in the data, as we provide in this paper.

⁷Of course, at finite distance, different weighting schemes will generate different proxies for idiosyncratic variance.

⁸See Cuzick (1995) or Sung (2002) for slightly weaker conditions.

principle the condition $\sum_{i=1}^{N_t} w_{it}^2 = O(1/N_t)$ may not hold, depending on how market capitalization weights evolve as the number of stocks increases. In the empirical analysis that follows, we actually focus on the EW scheme, and we consider only the CW scheme for comparison purposes. Formal justification for our focus on the EW scheme is provided in the next section, where we show that the CSV^{EW} is the best estimator for idiosyncratic variance within the class of CSV obtained under a strictly positive weighting scheme.

C. Properties of CSV as an Estimator for Idiosyncratic Variance

First, we derive in Proposition 2 the bias and the variance of the CSV as an estimator of idiosyncratic variance. Then we study asymptotic limits as the number of firms grows large and conclude that the equal-weighted CSV is the best among all positively weighted estimators.

Proposition 2: Bias and Variance of CSV. Maintaining the homogeneous beta assumption ($\beta_{it} = \beta_t = 1 \forall i, t$) and the homogeneous residual variance assumption ($E(\varepsilon_{it}^2) = \sigma_\varepsilon^2(t) \forall i$), for any strictly positive weighting scheme, we have that

$$(7) \quad E \left[\text{CSV}_t^{(w_t)} \right] = \sigma_\varepsilon^2(t) \left(1 - \sum_{i=1}^{N_t} w_{it}^2 \right).$$

To analyze the variance of the CSV estimator, we further make the assumption of multivariate normal residuals $\varepsilon \sim N(0, \Sigma^\varepsilon)$, where Σ^ε denotes the variance-covariance matrix of the residuals. Under this additional assumption, we obtain

$$(8) \quad \text{var} \left[\text{CSV}_t^{(w_t)} \right] = 2\sigma_\varepsilon^2(t) \left(\left(\sum_{i=1}^{N_t} w_{it}^2 \right)^2 + \sum_{i=1}^{N_t} w_{it}^2 - 2 \sum_{i=1}^{N_t} w_{it}^3 \right).$$

Proof. See Appendix B for a proof in the slightly more general case when the homogeneous specific variance assumption has been relaxed. \square

Hence, the CSV is a biased estimator for idiosyncratic variance, with a bias given by the multiplicative factor $(1 - \sum_{i=1}^{N_t} w_{it}^2)$, which can be easily corrected for, since it is available in explicit form. In the end, the bias and variance of the CSV appear to be minima for the EW scheme, which corresponds to taking $w_{it} = 1/N_t$ at each date t . It is easy to see that this bias disappears and the variance tends to 0 for the EW scheme when the number of stocks grows infinitely large, as explained in the following proposition.

Proposition 3: Properties of the Equal-Weighted CSV. The bias and variance of the equal-weighted CSV as an estimator for specific variance disappear in the limit of an increasingly large number of stocks:

$$\begin{aligned} E \left[\text{CSV}_t^{\text{EW}} \right] &\xrightarrow{N_t \rightarrow \infty} \sigma_\varepsilon^2(t). \\ \text{var} \left(\text{CSV}_t^{\text{EW}} \right) &\xrightarrow{N_t \rightarrow \infty} 0. \end{aligned}$$

Proof. See Appendix B for a proof in the slightly more general case when the homogeneous specific variance assumption has been relaxed. \square

The equal-weighted CSV thus appears to be a consistent and asymptotically efficient estimator for idiosyncratic variance. As such, it is the best estimator in the class of CSV estimators defined under any positive weighting scheme, and it dominates in particular the cap-weighted CSV as an estimator for idiosyncratic variance. If we relax the homogeneous residual variance assumption, we still obtain that the CSV appears to be an asymptotically unbiased estimator for the *average* idiosyncratic variance of the stocks in the universe (see Appendix B). Hence, the assumption of homogeneous residual variances comes with no loss of generality.

D. Relaxing the Homogeneity Assumption for Factor Loadings

Relaxing the homogeneous beta assumption involves a bias that remains strictly positive even for an infinite number of stocks and an EW scheme. We characterize this bias in the next proposition in order to gauge its magnitude for given models of returns.

Proposition 4: Bias of CSV as an Estimator for Average Idiosyncratic Variance in the Presence of Heterogeneous Betas. Relaxing the assumptions $\beta_{it} = \beta_t = 1 \forall i, t$ (homogeneous beta assumption) we have, for any strictly positive weighting scheme,

$$(9) \quad E \left[\text{CSV}_t^{(w_i)} \right] = \sum_{i=1}^{N_t} w_{it} \sigma_{\varepsilon_i}^2(t) - \sum_{i=1}^{N_t} w_{it}^2 \sigma_{\varepsilon_i}^2(t) + E \left[F_t^2 \text{CSV}_t^\beta \right],$$

where CSV_t^β denotes the cross-sectional variance of stock betas:

$$\text{CSV}_t^\beta = \sum_{i=1}^{N_t} w_{it} \left(\beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right)^2.$$

Proof. See Appendix C. \square

The first term $\sum_{i=1}^{N_t} w_{it} \sigma_{\varepsilon_i}^2(t)$ in equation (9) represents the average idiosyncratic variance of stocks within the universe under consideration. The second term $-\sum_{i=1}^{N_t} w_{it}^2 \sigma_{\varepsilon_i}^2(t)$ is the negative bias that was also present even in the presence of the homogeneous beta assumption. If we focus on the EW scheme, the sum of these two terms is equal to $1/N_t \sum_{i=1}^{N_t} \sigma_{\varepsilon_i}^2(t) (1 - (1/N_t))$, so that the bias disappears in the limit of an increasingly large number of stocks. The third term $E[F_t^2 \text{CSV}_t^\beta]$ in equation (9) represents, on the other hand, an additional (positive) bias for the CSV as an estimator of average idiosyncratic variance, which is introduced by the cross-sectional dispersion in betas and which does not disappear in the limit of an infinitely large number of stocks.

In Section III.A, we use this explicit expression to directly measure the *beta dispersion bias* using the CAPM and the FF (1993) 3-factor model as benchmark factor models. An extensive analysis of the CSV in the empirical section suggests that the homogeneous beta assumption does not represent a material problem for the CSV as an estimator of idiosyncratic variance as implied by standard asset pricing models (i.e., CAPM and FF).

E. Competing Measures of Idiosyncratic Risk

In this section, we describe measures that have been used in the literature and that will be used for comparison purposes in subsequent sections of the paper. The standard approach consists of considering idiosyncratic variance either relative to the CAPM and/or to the FF (1993) model:

$$(10) \quad r_{it} = b_{0it} + b_{1it} \text{XMKT}_t + b_{2it} \text{SMB}_t + b_{3it} \text{HML}_t + \varepsilon_{it}^{\text{FF}},$$

where r_{it} denotes the excess return at time t of stock i , XMKT is the excess return on the market portfolio, SMB is the size factor, and HML is the value factor. The idiosyncratic variance for asset i is the variance of the residuals of the regression, that is, $\sigma^2(\varepsilon_{it}^{\text{FF}})$. To obtain an estimate for average idiosyncratic variance, Bekaert et al. (2012) use a market capitalization weighting:

$$(11) \quad \text{FF}_t^{\text{CW}} = \sum_{i=1}^{N_t} w_{it} \sigma^2(\varepsilon_{it}^{\text{FF}}).$$

For comparison purposes we also look at the EW average of FF (1993) idiosyncratic variance in what follows.

An alternative approach to average (*mostly*) idiosyncratic risk estimation has been suggested by Goyal and Santa-Clara (GS) (2003), with a measure given by

$$(12) \quad \text{GS}_t^{\text{EW}} = \frac{1}{N_t} \sum_{i=1}^{N_t} \left[\sum_{d=1}^{D_t} r_{id}^2 + 2 \sum_{d=2}^{D_t} r_{id} r_{id-1} \right],$$

where r_{id} is the return on stock i in day d and D_t is the number of trading days in month t .⁹

More recently, Bali, Cakici, and Levy (2008) proposed a model-independent measure of aggregate idiosyncratic risk that does not require estimation of market betas nor correlations and that is based on the concept of gain from portfolio diversification:

$$(13) \quad \sigma_{\varepsilon,t}^2 = \left(\sum_{i=1}^n w_{i,t} \sigma_{i,t} \right)^2 - \text{var}(R_{m,t}),$$

where $(\sum_{i=1}^n w_{i,t} \sigma_{i,t})^2$ is the variance of the nondiversified portfolio and $\text{var}(R_{m,t})$ is the variance of the fully diversified portfolio, that is, the market variance.¹⁰

III. Empirical Implementation

In order to perform an empirical analysis of our measure for idiosyncratic risk, we collect daily U.S. stock returns (common equity shares only) and their

⁹As in GS (2003), when the second term makes the estimate negative, it is ignored. This measure was originally used in French, Schwert, and Stambaugh (1987).

¹⁰Campbell et al. (2001) propose an alternative measure of average idiosyncratic variance. They assume that all betas are equal to one and subtract industry returns in addition to market returns to control for risk. We do not include this measure in our comparison because Bekaert et al. (2012) have shown that it is very closely related to the measure obtained from standard asset pricing models.

market capitalization from the Center for Research in Security Prices (CRSP) database. Our longest sample runs from July 1963 to Dec. 2006. We also extract the FF (1993) factors and the 1-month Treasury bill from the Kenneth French Web site (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html) data library for the same sample period. Each month, we drop stocks with missing returns and with nonpositive market capitalization at the beginning of the month. The number of firms varies between 377 and 7,293 and remains greater than one thousand 75% of the time. The maximum number of stocks is reached during the dot.com bubble. Then, we estimate every month the CW¹¹ and EW average idiosyncratic variances relative to the CAPM and FF, GS, and Bali measures in equations (11), (12), and (13), respectively. We also estimate on a daily basis the equal- and cap-weighted versions of the CSV as in equations (4) and (5). In order to construct the monthly series for our cross-sectional measures, we estimate the average of the daily series at the end of each month. We annualize all figures in order to compare daily and monthly measures. Following Bekaert et al. (2012), we fit a regime-switching model to the monthly and daily series in order to further compare the different measures. Last, we look at the relation between the CSV measures of idiosyncratic variance and selected economic and financial variables.

A. Measuring the CSV Bias

As illustrated in Proposition 4 and discussed in Appendix C, the presence of nonhomogeneous betas introduces a positive bias on the CSV as an estimator for average idiosyncratic variance, which is given by the last term in equation (9). We now measure the impact of this bias with respect to the CAPM as a benchmark model.

First, we compute the bias $E[F_t^2 CSV_t^\beta]$ for every month in the sample using beta estimates for each stock with both the EW and the CW market returns. To gauge its importance, we divide it by the average idiosyncratic variance, also measured with respect to the CAPM.¹²

Table 1 presents a summary of the distribution of the time series of cross-sectional dispersion of betas, its product with the squared return of the market portfolio (hence, the bias itself), and the proportion of this bias with respect to the average idiosyncratic variance at the end of every month. Although the cross-sectional dispersion of betas is sizable, once it is multiplied by the squared return of the market portfolio, the size of the bias remains relatively small. The median of the distribution of $F_t^2 CSV_t^\beta / \sigma_{\varepsilon_t}^2$ is 0.348% for the EW scheme and 0.351% for the CW measure, computed over the whole sample (July 1963 to Dec. 2006). The 97.5 quantiles are 3.24% and 3.47%, respectively.

On the other hand, the formal discussion about the properties of the CSV as a measure of idiosyncratic variance in Section II.C also uncovered the fact that another bias (but negative in sign) coming from the CSV weighting scheme concentration is also introduced.

¹¹We use previous period market capitalization and assume it is constant within the month.

¹²This is measured as in equation (11) with just the market returns with both weighting schemes.

TABLE 1
 Estimates of the Biases Due to the Cross-Sectional Dispersion
 of Betas and Weight Concentration

Table 1 contains a summary of the distribution of the following time series: the cross-sectional dispersion of betas CSV_t^β , estimated with respect to the CAPM at the end of every month using daily returns; the average idiosyncratic variance $\sigma_{\varepsilon_t}^2$ with respect to the CAPM; the product of the average return of the market portfolio squared, F_t^2 , and the beta dispersion, CSV_t^β ; and the proportion of the product $F_t^2 CSV_t^\beta$ to $\sigma_{\varepsilon_t}^2$ and the proportion of $\sum w_{it}^2 \sigma_{\varepsilon_{it}}^2$ to $\sigma_{\varepsilon_t}^2$. Panel A corresponds to the equal-weight scheme (CSV^{EW}) and Panel B to the market-cap weighting (CSV^{CW}). All figures are daily. The period is from July 1963 to Dec. 2006.

	<u>Q2.5</u>	<u>Q25</u>	<u>Q50</u>	<u>Q75</u>	<u>Q97.5</u>
<i>Panel A. Equal-Weighted</i>					
CSV_t^β	0.282	0.970	1.563	3.022	11.437
$\sigma_{\varepsilon_t}^2$ (%)	0.043	0.065	0.103	0.241	0.485
$F_t^2 CSV_t^\beta$ (%)	6.57e-07	6.92e-05	3.84e-04	0.001	0.005
$F_t^2 CSV_t^\beta / \sigma_{\varepsilon_t}^2$ (%)	0.001	0.078	0.348	0.890	3.240
$\sum w_{it}^2 \sigma_{\varepsilon_{it}}^2 / \sigma_{\varepsilon_t}^2$ (%)	0.014	0.020	0.030	0.054	0.154
<i>Panel B. Cap-Weighted</i>					
CSV_t^β	0.075	0.309	0.451	0.704	3.079
$\sigma_{\varepsilon_t}^2$ (%)	0.009	0.020	0.030	0.042	0.153
$F_t^2 CSV_t^\beta$ (%)	1.83e-07	2.30e-05	1.09e-04	2.77e-04	0.001
$F_t^2 CSV_t^\beta / \sigma_{\varepsilon_t}^2$ (%)	4.85e-04	0.080	0.351	0.930	3.472
$\sum w_{it}^2 \sigma_{\varepsilon_{it}}^2 / \sigma_{\varepsilon_t}^2$ (%)	0.173	0.281	0.426	0.637	1.463

Using the explicit expression for this bias provided in Proposition 4, we estimate the proportion of the size of this weights-concentration bias with respect to the average idiosyncratic variances implied by the CAPM.¹³ In the last line of Panels A and B of Table 1 we report quantiles of the distribution of this bias for both weighting schemes. The corresponding medians are 0.030% and 0.426% for the EW and CW schemes, respectively. Since the bias is of opposite sign to the beta cross-sectional dispersion bias, we need to assess the resulting overall bias.

We measure the total bias as the intercept of a regression of the CSV on the average idiosyncratic variance estimated with respect to the CAPM or the FF (1993) 3-factor model:

$$(14) \quad CSV_t^{w_i} = \text{BIAS} + \psi \sigma_{\text{MODEL}}^2(w_i) + \zeta_t,$$

where w_i refers to the weighting scheme (EW or CW) and model stands for either the CAPM or the FF 3-factor model.

Table 2 reports summary statistics for regression (14). The bias of the CSV measured with respect to standard asset pricing models is small in magnitude for both weighting schemes (in the order of 10^{-5}). While it remains statistically significant, we can safely consider that the impact of the bias remains immaterial for any practical purposes. Another interesting finding is the sign of the bias. For the EW quantities, the sign of the bias is positive, while it is negative for the CW quantities. Therefore, the beta bias dominates the weighting bias for EW averages in both models. This is consistent with the prediction made by the theoretical

¹³As noted earlier, it would be straightforward to remove the impact of this bias by dividing the CSV measure by the factor $(1 - \sum_{i=1}^{N_t} w_{it}^2)$, equal to $(1 - 1/N_t)$ in the EW case.

TABLE 2
Total Bias Associated with CSV

	CAPM ^{EW}	FF ^{EW}	CAPM ^{CW}	FF ^{CW}
BIAS	1.29e-05	2.23e-05	-2.09e-05	-3.74e-05
NW <i>t</i> -statistics	(1.986)	(2.382)	(-2.849)	(-4.767)
ψ	0.983	0.988	1.125	1.242
NW <i>t</i> -statistics	(153.819)	(100.162)	(39.259)	(39.226)
\bar{R}^2 (%)	99.866	99.503	98.946	97.117

Table 2 reports the results of the regression $CSV_t^{wt} = \text{BIAS} + \psi \sigma_{\text{MODEL}}^2(w_t) + \zeta_t$, where $\sigma_{\text{MODEL}}^2(w_t)$ represents monthly estimates of the weighted average idiosyncratic variance estimated using the corresponding model (either CAPM or FF). The average measures and the CSV are computed with either the cap-weighted (CW) or the equal-weighted (EW) scheme. The Newey and West (NW) (1987) corrected *t*-statistics are reported in parentheses. The period is from July 1963 to Dec. 2006.

analysis regarding the relative impact of the weighting bias for different weighting schemes. Regarding the model, the bias is larger when the idiosyncratic variance is measured with respect to the FF (1993) model instead of the CAPM for both weighting schemes, as expected, but its magnitude remains negligible.

B. Comparison with Other Measures

In this section we compare the CSV measure to the aforementioned measures of idiosyncratic risk (i.e., FF-based, CAPM-based, GS, and Bali). To obtain the factor-based measures, we need to re-estimate the relevant factor model using a rolling window of 1 month's worth of daily data to allow for time variation in beta estimates (or total-variance variation for the GS measure). In Table 3, we report summary statistics for the monthly time series of annualized idiosyncratic variances based on 516 observations from Jan. 1964 to Dec. 2006.¹⁴

On the monthly series, the annualized means of the equal-weighted CSV, FF-based, and CAPM-based measures are 38.4%, 38.3%, and 38.7%, respectively, while the EW means for the GS and Bali variances are 34.2% and 23.6%, respectively. The standard deviations are 8.5%, 8.6%, and 8.7% for the CSV, FF-based, and CAPM-based measures and 7.0% and 4.7% for the GS and Bali measures, respectively. For the CW version, the CSV, FF, and CAPM idiosyncratic variance measures have annualized means of 8.5%, 7.6%, 8.0%, respectively, and the GS and Bali measures have means of 11.2% and 6.5%, respectively. The standard deviations are also closer for the CSV, FF, CAPM, and Bali measures than for the GS measure. Although GS argue that their measure fundamentally constitutes a measure of idiosyncratic risk, it is strictly speaking an average of total stock variance.

The cross-correlation matrix reported in Table 3 provides further evidence of the CSV closeness to the other model-based measures. Correlations are high between CSV^{EW} and $CAPM^{EW}$ (99.93%), FF^{EW} (99.75%), and $BALI^{EW}$ (97.75%), and between CSV^{CW} and $CAPM^{CW}$ (99.48%), FF^{CW} (98.56%), and $BALI^{CW}$ (99.36%). The high correlations between the CAPM and the FF measures

¹⁴In this section of the paper, we start the sample period in Jan. 1964 to allow for direct comparison with Bekaert et al. (2012). In the predictability section, we instead start the sample in July 1963.

TABLE 3
Comparison of Monthly Measures of Idiosyncratic Variance

Panel A of Table 3 contains the annualized means and standard deviations of the monthly time series for the CSV, the average idiosyncratic variance based on the CAPM and the Fama-French (FF) (1993) models, and the average Goyal and Santa-Clara (GS) (2003) and Bali et al. (2008) idiosyncratic variance measures as in equations (4), (11), (12), and (13) using both EW and CW weighting schemes. Panel B presents the cross-correlation matrix among these variables. The period is from Jan. 1964 to Dec. 2006.

Panel A. Mean and Standard Deviation of Different Measures of Idiosyncratic Variance

	CSV ^{EW}	FF ^{EW}	CAPM ^{EW}	GS ^{EW}	BALI ^{EW}	CSV ^{CW}	FF ^{CW}	CAPM ^{CW}	GS ^{CW}	BALI ^{CW}
Mean	0.384	0.383	0.387	0.342	0.236	0.085	0.076	0.080	0.112	0.065
Std. dev.	0.085	0.086	0.087	0.070	0.047	0.020	0.016	0.018	0.029	0.015

Panel B. Correlation Matrix between Different Measures of Idiosyncratic Variance

Correlation	CSV ^{EW}	FF ^{EW}	CAPM ^{EW}	GS ^{EW}	BALI ^{EW}	CSV ^{CW}	FF ^{CW}	CAPM ^{CW}	GS ^{CW}	BALI ^{CW}
CSV ^{EW}	100.00									
FF ^{EW}	99.75	100.00								
CAPM ^{EW}	99.93	99.88	100.00							
GS ^{EW}	95.44	93.67	94.78	100.00						
BALI ^{EW}	97.75	96.73	97.42	96.00	100.00					
CSV ^{CW}	72.16	68.53	70.62	82.38	81.63	100.00				
FF ^{CW}	74.98	72.16	73.66	82.88	83.76	98.56	100.00			
CAPM ^{CW}	72.83	69.46	71.47	82.36	82.12	99.48	99.18	100.00		
GS ^{CW}	61.25	56.75	59.83	76.60	71.85	92.64	88.17	92.13	100.00	
BALI ^{CW}	69.38	65.81	67.95	80.34	79.42	99.36	98.41	99.44	93.27	100.00

(99.88% and 99.18% for EW and CW, respectively) also indicate that adding factors does not drastically affect the estimation of idiosyncratic variance. Correlations between the GS measures and the other measures are always smaller but remain close to 90% when considering the same weighting scheme. Correlations between measures for different weighting schemes are much lower, irrespective of the estimation method, indicating that the choice over the weighting scheme is fundamentally important for estimating idiosyncratic variance, as stressed in our theoretical analysis in Section II.

Table 4 provides mean and standard-deviation estimates for the daily average idiosyncratic variance measures. The mean of the equal-weighted CSV is 38.4%, practically equal to the mean of EW idiosyncratic variance based on the FF (1993) model.¹⁵ For the CW measures, the CSV has a slightly higher mean than the FF-based model. For the CSV daily series, the standard deviation is higher than for the FF-based measure for both weighting schemes. This is due to the different nature of the two series. The CSV includes only information from the cross section of realized returns, while the FF idiosyncratic variance is a persistent, overlapping, rolling-window estimate.

Panel B of Table 4 presents cross-correlations for the daily series of idiosyncratic variance measures. Although the coefficients are smaller than for the monthly series, the relation remains strong provided the comparison is done for the same weighting scheme: 82.6% and 73.9% for EW and CW measures, respectively. The difference with the monthly series correlations may again be explained by the presence of the smoothed estimation procedure inherent to the FF-based measure.

¹⁵It is estimated based on a rolling window of 1 month.

TABLE 4
Comparison of Daily Measures of Idiosyncratic Variance

Panel A of Table 4 contains the annualized means and standard deviations of the daily time series for the CSV and the average idiosyncratic variance based on the Fama-French (FF) (1993) model as in equations (4) and (11) using both weighting schemes. Panel B presents the cross-correlation matrix among these variables. The period is from Jan. 1964 to Dec. 2006.

Panel A. Mean and Standard Deviation of Different Measures of Idiosyncratic Variance

	CSV ^{EW}	FF ^{EW}	CSV ^{CW}	FF ^{CW}
Mean	0.384	0.383	0.085	0.078
Std. dev.	0.021	0.019	0.005	0.004

Panel B. Correlation Matrix between Different Measures of Idiosyncratic Variance

Correlation	CSV ^{EW}	FF ^{EW}	CSV ^{CW}	FF ^{CW}
CSV ^{EW}	100.00			
FF ^{EW}		100.00		
CSV ^{CW}			100.00	
FF ^{CW}				100.00

C. CSV: Regimes and Relation with Economic and Financial Variables

Bekaert et al. (2012) fit a Markov regime-switching model with a first-order autocorrelation structure for the monthly series of idiosyncratic variance based on the FF (1993) model. We estimate this model with our CSV measure at both the monthly and daily frequencies. In this model, two regimes are indexed by a discrete state variable, s_t , which follows a Markov-chain process with constant transition probabilities. Let the current regime be indexed by i , and let the past regime by j , and let x_t be the original idiosyncratic variance, which follows a first-order autoregressive (AR(1)) model:

$$(15) \quad x_t - \mu_i = \phi(x_{t-1} - \mu_j) + \sigma_i e_t, \quad i, j \in \{1, 2\}.$$

The transition probabilities are denoted by $p = P[s_t = 1 | s_{t-1} = 1]$ and $q = P[s_t = 2 | s_{t-1} = 2]$. The model involves a total of 7 parameters, $\{\mu_1, \mu_2, \sigma_1, \sigma_2, \phi, p, q\}$.

We report the estimation results in Panel A of Table 5 for the equal-weighted CSV, monthly and daily, and for the FF-based measure monthly. We observe that the monthly estimates obtained for the CSV and the FF-based measures are very similar. For both measures, the low-mean, low-variance regime displays a higher probability of remaining in the same state.¹⁶ We estimated the same model for the daily CSV series. The estimated values of the mean parameters μ_1 and μ_2 are quite close to the values obtained with the monthly series. This result suggests that we capture the same process at different frequencies.¹⁷ The resulting filtered probabilities suggest that the CSV measure is countercyclical, the dispersion of returns being high and quite variable when economic growth subsides.

Recently, Bansal and Yaron (2004) have revived consumption-based asset pricing models by showing that two sources of long-run risk (expected consumption

¹⁶Similar results for the cap-weighted CSV are not included for space considerations.

¹⁷This is, of course, not the case for the daily FF-based measure, since it is an overlapping measure of idiosyncratic variance. This results in a high degree of persistence, with an autocorrelation parameter very close to 1, and similar means in both regimes.

TABLE 5

Idiosyncratic Volatility: Regimes and Relations with Macroeconomic and Financial Variables

Panel A Table 5 reports the parameter estimates of the regime-switching model specified in equation (15), first at the monthly frequency for the equal-weighted (EW) CSV and for the average idiosyncratic variance based on the FF (1993) model as in equations (4) and (11), then at the daily frequency for the CSV. Here, μ_i is the average level of the variable on regime i , σ_i is the standard deviation level of the variable on regime i , ϕ is the autocorrelation coefficient, p and q are the probabilities of remaining in regimes 1 and 2, respectively. The estimation period is from Jan. 1964 to Dec. 2006. In Panel B, we run the regression $CSV_t^{EW} = \alpha + \beta_{Evar} Evar_t$, where $Evar_t$ is either Consumption Volatility, Credit Spread, Term Spread, Inflation Volatility, or Risk-Free Rate. The sample period is Jan. 1990 to Dec. 2006.

Panel A. Mean and Standard Deviation of Different Measures of Idiosyncratic Variance

	CSV_m^{EW}	FF_m^{EW}	CSV_d^{EW}
μ_1	0.401	0.363	0.446
μ_2	0.299	0.275	0.304
σ_1	0.067	0.062	0.036
σ_2	0.010	0.009	0.003
ϕ	0.980	0.981	0.965
p	0.839	0.823	0.695
q	0.963	0.951	0.956

Panel B. Correlation Matrix between Different Measures of Idiosyncratic Variance

	Full Sample		Positive Returns		Negative Returns	
	CSV^{EW}	\bar{R}^2 (%)	CSV^{EW+}	\bar{R}^2 (%)	CSV^{EW-}	\bar{R}^2 (%)
Consumption Volatility	18.611	15.659	12.063	2.922	2.930	11.532
t-statistics	(3.242)		(2.403)		(3.505)	
Credit Spread	2.067	2.642	1.647	0.479	0.352	2.239
t-statistics	(1.560)		(1.178)		(1.687)	
Term Spread	-0.157	0.248	-0.279	0.649	-0.073	4.317
t-statistics	(-0.558)		(-1.166)		(-1.390)	
Inflation Volatility	-10.420	13.030	-5.547	1.379	-0.501	0.445
t-statistics	(-3.631)		(-1.901)		(-0.985)	
Risk-Free Rate	5.148	8.687	2.209	0.331	0.508	2.195
t-statistics	(2.457)		(1.235)		(1.269)	

growth and consumption volatility as a measure of economic uncertainty) determine asset returns. Furthermore, Tédongap (2014) provides strong evidence that consumption volatility risk explains a high percentage of the cross-sectional dispersion in average stock returns for the usual set of size and book-to-market portfolios. In tests of the intertemporal CAPM (ICAPM) or the conditional CAPM, the cross section of expected returns is linked to other economic or financial variables such as the term spread, default spread, implied or realized measures of aggregate returns variance, and many others. We relate our CSV measure to some of these variables.

Our first economic variable is consumption-growth volatility as a measure of economic uncertainty. Following Bansal and Yaron (2004) and Tédongap (2014), we filter consumption-growth volatility with a generalized autoregressive conditional heteroskedasticity (GARCH) model. For consumption, we used the Federal Reserve Economic Data's (FRED) (<http://research.stlouisfed.org/fred2/>) personal consumption expenditures of nondurables and services monthly series, divided by the consumer price index and the population values to obtain a per-capita real consumption series. We then compute its growth rate from July 1963 to 2006.¹⁸

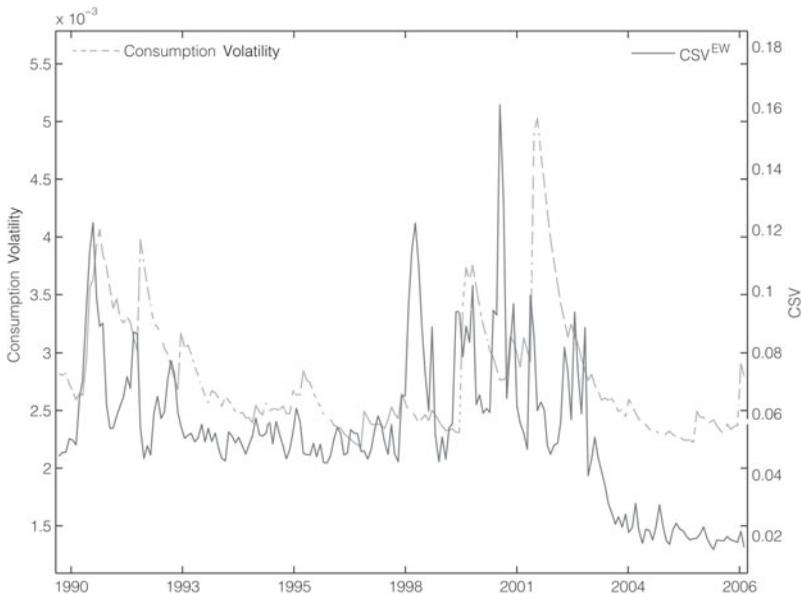
¹⁸The series IDs at the FRED's Web page (<http://research.stlouisfed.org/fred2/>) are PCEND and PCES for "Personal Consumption Expenditures: Nondurable Goods" and "Personal Consumption Expenditures: Services," CPIAUCNS for "Consumer Price Index for All Urban Consumers:

The second economic variable we consider is inflation volatility, which we filter also with a GARCH process.¹⁹ For the financial variables we use Welch and Goyal's (2008) data for corporate bond yields on BAA- and AAA-rated bonds, long-term government bond yield, and the 3-month T-bill rate to estimate the credit spread and term spread (as the difference between the first and the second rate in both cases).²⁰ In Panel B of Table 5 we report the results of univariate regressions of the EW measure of the CSV on the five economic and financial variables during the 1990–2006 period. We also explore some potential asymmetries by computing the CSV^{EW} for the positive and negative returns.

We observe that the R^2 for consumption-growth volatility is close to 16%. In Figure 1 we plot the two series over the same sampling period. While the CSV series is much noisier than consumption-growth volatility, the coincident movements between the two series are quite remarkable. This suggests the presence of a common factor (aggregate economic uncertainty) that affects the idiosyncratic

FIGURE 1
Cross-Sectional Variance and Consumption Volatility

In Figure 1, monthly time series of CSV^{EW} is on the right-hand axis and Consumption Volatility is on the left-hand axis. The sample period is Jan. 1990–Dec. 2006.



All Items,” and POP for “Total Population: All Ages including Armed Forces Overseas.” Bansal and Yaron (2004) used the Bureau of Economic Analysis data available at www.bea.gov/national/consumer_spending.htm on real per-capita annual consumption growth of nondurables and services for the period 1929 to 1998. The series is longer but is available only at annual and quarterly frequencies.

¹⁹For space considerations, we do not report parameter estimates for the two AR(1)-GARCH(1,1) we estimate. They are available from the authors.

²⁰Data available at Amit Goyal's Web page: <http://www.hec.unil.ch/agoyal/>

variance of each security. Aggregating over all securities will make the CSV a function of economic uncertainty.²¹

The next most highly negatively correlated variable is inflation volatility, with an R^2 of 13%. Since 1998, inflation volatility has been on an upward trend, while the CSV of returns has been sharply declining. In the presence of higher inflation uncertainty, investors will allocate more to stocks relative to bonds in their portfolios, generating a general increase in stock returns that reduces their CSV. The T-bill rate is also relatively highly correlated with CSV^{EW} with an R^2 close to 9%. In the type of equilibrium models we have referred to, the risk-free rate, proxied here by the T-bill, will be a function of consumption-growth volatility, hence its positive relation with the CSV.

For the financial variables (credit spread and term spread), the signs obtained are intuitive. Credit risk affects individual firm returns differently and therefore tends to increase CSV, while a pervasive term spread risk will reduce dispersion by being common to many securities due to a move of investors away from bonds into the stock market. However, the relations are not statistically significant, although the relation improves for the CSV computed with negative returns. Therefore, it is in bear periods that these risk variables seem to most affect the dispersion of returns. This asymmetry between positive and negative returns is also present for consumption volatility.

IV. New Evidence on the Predictability of the Market Return

There is an ongoing debate on the predictive power of average idiosyncratic variance for average (or aggregate) stock market returns. GS (2003) find a significantly positive relation between the EW average idiosyncratic stock variance and the CW portfolio returns for the period July 1693–Dec. 1999. They find that their measure of average idiosyncratic variance has a significant relation with next-month return on the CW portfolio. The regression in GS is

$$(16) \quad r_{t+1}^{CW} = \alpha + \beta \nu_t^{EW} + \varepsilon_{t+1},$$

where ν_t^{EW} corresponds to GS_t^{EW} . In a subsequent analysis, Bali et al. (2005) argue that this relation disappeared for the extended sample July 1963–Dec. 2001, and they attribute the relation observed in GS to high-tech-bubble-type stocks (i.e., stocks traded on the NASDAQ) and a liquidity premium. In a similar way, Wei and Zhang (2005) find that the significance of the relation found by GS disappeared for their sample July 1963–Dec. 2002 and argue that the presumably temporary result of GS was driven mainly by the data in the 1990s. Wei and Zhang criticize the fact that GS looked at the relation between an EW average stock variance and the return on a CW average stock return, as opposed to an EW portfolio return. Moreover, both Bali et al. (2005) and Wei and Zhang find no significant relation between the CW measures and the CW portfolio return in all three sample periods (ending in 1999, 2001, and 2002, respectively).

²¹In intertemporal asset pricing models of Bansal and Yaron (2004) and Bollerslev, Tauchen, and Zhou (2009), economic uncertainty is a priced risk factor that affects returns, therefore providing a fundamental rationale for the observed correlation between CSV and consumption-growth volatility.

A. Monthly Evidence

In this section we confirm existing results and extend them in a number of dimensions, including a longer sample period. Panel A in Table 6 presents the predictability regression of the CW return by EW variance measures as in GS (2003) and Bali et al. (2005) for their sample periods, as well as for the extended sample up to Dec. 2006. The regression is as in equation (16), where ν_t^{EW} corresponds to the EW CAPM-based measure and the CSV.²² For comparison purposes we start the sample period in this section in July 1963, as in GS (2003), Bali et al. (2005), and Wei and Zhang (2005).

For the monthly series, we confirm that there is a significant positive relation in the first sample, and also that it weakens for the subsequent extended samples.²³

TABLE 6
Monthly Predictability of Market Returns by Two Measures of Idiosyncratic Variance (CSV and CAPM-based)

	Monthly Series					
	Aug. 1963–Dec. 1999		Aug. 1963–Dec. 2001		Aug. 1963–Dec. 2006	
	CAPM ^{EW}	CSV ^{EW}	CAPM ^{EW}	CSV ^{EW}	CAPM ^{EW}	CSV ^{EW}
<i>Panel A. Forecasting r^{CW}</i>						
Intercept	−0.001	−0.002	0.001	0.001	0.002	0.002
NW <i>t</i> -statistics	(−0.424)	(−0.479)	(0.228)	(0.264)	(0.586)	(0.617)
Coefficient	0.241	0.250	0.123	0.120	0.090	0.087
NW <i>t</i> -statistics	(3.543)	(3.609)	(1.356)	(1.247)	(1.055)	(0.964)
\bar{R}^2 (%)	1.336	1.365	0.275	0.233	0.072	0.044
<i>Panel B. Forecasting r^{EW}</i>						
Intercept	4.51e-04	3.31e-04	−1.67e-04	−1.06e-05	0.001	0.001
NW <i>t</i> -statistics	(0.086)	(0.063)	(−0.033)	(−0.002)	(0.241)	(0.270)
Coefficient	0.247	0.254	0.238	0.235	0.217	0.215
NW <i>t</i> -statistics	(2.175)	(2.189)	(2.395)	(2.329)	(2.331)	(2.271)
\bar{R}^2 (%)	0.774	0.773	0.885	0.824	0.726	0.678
	CAPM ^{CW}	CSV ^{CW}	CAPM ^{CW}	CSV ^{CW}	CAPM ^{CW}	CSV ^{CW}
<i>Panel C. Forecasting r^{CW}</i>						
Intercept	0.001	0.001	0.007	0.007	0.007	0.008
NW <i>t</i> -statistics	(0.150)	(0.334)	(2.226)	(2.459)	(2.551)	(2.823)
Coefficient	0.856	0.688	−0.356	−0.373	−0.404	−0.421
NW <i>t</i> -statistics	(1.192)	(0.995)	(−0.871)	(−1.081)	(−1.080)	(−1.334)
\bar{R}^2 (%)	0.128	0.043	−0.037	0.038	0.035	0.123

²²As explained before, the monthly CSV is the average of its daily values during the month.

²³We found a similar result using the GS measure of equal-weighted average variance. We do not present these regression results for the sake of brevity, given that they produce similar findings, which have also been confirmed in Bali et al. (2005) and Wei and Zhang (2005).

The Newey and West (NW) (1987) autocorrelation-corrected t -statistics for 12 lags of the β coefficient of both CSV and the CAPM-based measures goes from 3.54 for the first sample period down to 0.96 for the largest sample. Consequently, the adjusted R^2 goes from 1.3% down to 0.04%. Therefore, the findings of Bali et al. (2005) and Wei and Zhang (2005) are also present in the extended sample. In Section IV.C we propose a possible explanation for this result.

In Panel B of Table 6 we present the results of the regression between the EW average return and the lagged EW idiosyncratic variance measure, as given by

$$(17) \quad r_{t+1}^{\text{EW}} = \alpha + \beta \nu_t^{\text{EW}} + \varepsilon_{t+1},$$

where ν_t^{EW} is taken as the CAPM-based average idiosyncratic variance or as the CSV measure. In contrast with the former regression, the relation is found to be significantly positive for the three sample periods and for both measures.²⁴

In Panel C of Table 6 we present the results for the three sample periods of the 1-month-ahead predictive regression of the cap-weighted market portfolio using the cap-weighted idiosyncratic variance return as a predictor. In this case, the beta of the idiosyncratic variance is not significant for all three sample periods. This result confirms the findings of Bali et al. (2005) and Wei and Zhang (2005) for the extended sample.

B. New Predictability Evidence at Daily Frequency

Prevailing measures used in the literature require a sample of past data to estimate additional parameters, constraining existing evidence to the monthly estimations. Fu (2009) finds that high idiosyncratic volatilities of individual stocks are contemporaneous with high returns, which tend to reverse in the following month. Huang, Liu, Rhee, and Zhang (2010) find that the negative relation between idiosyncratic variance and expected returns at the stock level uncovered in Ang et al. (2006) and Ang, Hodrick, Xing, and Zhang (2009) becomes positive after controlling for the return reversals. Looking at the predictability relation at a higher frequency than the month may help sort out these effects. Taking advantage of the instantaneous nature of the CSV, we run the same predictability regression (17) on the 1-day-ahead portfolio return using the average idiosyncratic variance.

Panel A of Table 7 shows that at a daily basis, this relation is much stronger, with (NW (1987) corrected) t -statistics of coefficients for the average idiosyncratic variance across the three samples ranging between 4 and 4.7.

In Panel B of Table 7 we report the results for the 1-day-ahead predictive regression on the cap-weighted pairs (CSV and market return). The relation is positive and significant for all three samples, but the t -statistic on the CW idiosyncratic variance coefficient deteriorates as we increase the period, from about 5.91 in the first sample down to 1.97 for the longest sample.

²⁴Wei and Zhang (2005) find a significantly positive relation between the equal-weighted GS measure and the EW market return for the initial sample. They also test the robustness of the relation by using an equal-weighted CSV of monthly returns. They found a significantly positive coefficient for predicting the EW portfolio return mainly for the long samples starting in 1928 but not for the sample going from 1963 to 2002. Note that our cross-sectional measures differ. Ours is an average of the daily CSVs over the month. Theirs is the CSV of the returns computed over the month.

TABLE 7
Daily Predictability of Market Returns by CSV

Panel A of Table 7 presents the results of a 1-day-ahead predictive regression of the equal-weighted daily market excess returns, denoted by r^{EW} , on the daily lagged equal-weighted cross-sectional variance, denoted by CSV^{EW} , estimated as in equation (4) for three sample periods. Panel B presents the results of the predictive regression of the cap-weighted daily excess market returns, denoted by r^{CW} , using the cap-weighted CSV, denoted by CSV^{CW} . The intercept, the regression coefficient corresponding to the CSV, the Newey-West (NW) (1987) corrected t-statistics (shown in parentheses) (30 lags), and the adjusted coefficient of determination, denoted by \bar{R}^2 , are reported. The sample periods are July 1963–Dec. 1999, July 1963–Dec.2001, and July 1963–Dec.2006.

	Daily Series		
	July 1963– Dec. 1999	July 1963– Dec. 2001	July 1963– Dec. 2006
	CSV^{EW}	CSV^{EW}	CSV^{EW}
<i>Panel A. Forecasting r^{EW}</i>			
Intercept	–1.58e-04	–1.40e-04	–1.29e-05
NW t-statistics	(–0.785)	(–0.714)	(–0.071)
Coefficient	0.544	0.483	0.411
NW t-statistics	(4.711)	(4.515)	(4.000)
\bar{R}^2 (%)	0.883	0.788	0.573
	CSV^{CW}	CSV^{CW}	CSV^{CW}
<i>Panel B. Forecasting r^{CW}</i>			
Intercept	–0.001	–1.88e-04	–1.65e-04
NW t-statistics	(–3.521)	(–0.791)	(–0.737)
Coefficient	3.404	1.189	1.151
NW t-statistics	(5.919)	(1.948)	(1.966)
\bar{R}^2 (%)	0.831	0.220	0.186

C. Interpretation of Predictability Results

The relation between the equal-weighted average idiosyncratic variance and the cap-weighted market index observed up to the end of the 1990s can be explained by the heterogeneous and transitory nature of the omitted sources of risk captured by idiosyncratic risk and its relation with the inflated valuation of several NASDAQ companies during that period.²⁵

Some intuition behind the far more robust relation between the EW average idiosyncratic variance and the EW portfolio comes precisely from the logic of standard asset pricing theory. As discussed in Section V, several theories can be invoked to link average idiosyncratic risk to average returns, due to the heterogeneous sources that may compose idiosyncratic risk. According to the CAPM, only systematic risk should explain future returns. However, if during a certain period of time there exist anomalies of any kind (priced omitted risk factors) that, presumably, are not proportionally reflected in the current market capitalization of the companies carrying these factors, then the omitted sources of risk are more likely to explain the returns of an EW portfolio than the returns of a VW portfolio.

²⁵In that sense, we agree with Wei and Zhang (2005) and Bali et al. (2005). The strongest omitted factor in that period (call it the irrational.com factor), partially captured by the EW idiosyncratic variance, started to be increasingly represented in the market-cap index, due to the suddenly higher market capitalization of precisely the group of companies carrying this temporarily strong omitted factor. The subsequent burst of the bubble explains the strong attenuation of the relation between the average idiosyncratic variance and the market-cap portfolio, since most stocks carrying this irrational.com factor lost value and their weight fell in the market-capitalization index.

Along these lines, Pontiff (2006) argues that idiosyncratic risk is the largest holding cost borne by rational arbitrageurs in their pursuit of mispricing opportunities. This theory implies that the current level of idiosyncratic risk should predict returns, since it should measure the amount of current mispricing opportunities present in the market. All things being equal, large-cap stocks are less likely to present mispricing, and hence the predictability implied by this theory would be more likely to be present on the EW index return rather than the CW index return, as we observed in predictive regressions at daily and monthly horizons.²⁶ Moreover, assuming that mispricing opportunities tend to disappear in the long run, it appears more likely to observe this relation between idiosyncratic variance and returns over very short horizons. This is consistent with our results. The sign of the relation is not predicted by Pontiff's theory in general, because it depends on whether the average (EW or CW) portfolio is over- or underpriced (it predicts a positive sign for underpriced stocks and a negative sign for overpriced stocks).

D. Robustness of Predictability

In this section, we test further the predictability documented in the former section. First, we control for some usual macroeconomic and financial predictors. Next, we include other measures of variance in the predictability regression. Finally, we check for the presence of asymmetry in the relation between idiosyncratic variance and future average returns.

1. Inclusion of Macroeconomic and Financial Predictors

Several authors (see, in particular, Bali et al. (2005) and GS (2003)) have suggested to estimate the risk-return trade-off predictability regression with various control variables to account for potential changes in the investment opportunity set.²⁷ We consider four macroeconomic and financial variables: the aggregate dividend yield (DIV), the term spread (TERM), the default or credit spread (DEF), and the relative T-bill rate (RREL). Over the longest sample period we have considered, July 1963 to Dec. 2006, only the variables DIV, TERM, and RREL are available at both daily and monthly frequencies. For this long sample, the default or credit spread is available only at the monthly frequency. At the daily frequency, the latter series starts only in Dec. 1996.

Table 8 reports the daily and monthly regression estimates for predicting the returns of the equal-weighted (r^{EW}) and cap-weighted (r^{CW}) indexes. At the daily frequency, the predictive power of CSV remains unaffected by the presence of the macroeconomic and financial variables. The coefficients are close to the values found in the univariate regression. Of the three added variables, only the relative T-bill rate appears as a strongly significant predictor. This is true for both r^{EW} and r^{CW} .²⁸ At the monthly frequency, the predicting capacity of CSV^{EW} for r^{EW}

²⁶It is well known that large-cap stocks are more liquid than small-cap stocks, which implies a higher number of people trading them and usually a higher number of analysts looking at them. Together with less constraints to short selling, we expect a higher price efficiency for large-cap stocks.

²⁷This is to capture the intertemporal hedging demand components in an ICAPM framework.

²⁸On the reduced sample from Dec. 1996, the default spread does not appear significant. The only significant variable is the dividend yield in the regression forecasting r^{CW} .

TABLE 8
Daily and Monthly Return Predictability by CSV Controlling for Macroeconomic Variables

Panels A and B of Table 8 include the multivariate regression results for daily predictions of the equal-weighted (EW) or cap-weighted (CW) market returns using the equal-weighted CSV (Panel A) or the cap-weighted CSV (Panel B) and a set of macroeconomic variables as predictors. Panels C and D report the corresponding monthly predictions. The set of macroeconomic variables are the relative T-bill rate (RREL), the aggregate dividend yield (DIV), the term spread (TERM), and the default or credit spread (DEF). The Newey and West (NW) (1987) corrected t-statistics are reported in parentheses. The sample period is from July 1963 to Dec. 2006.

	Coeff.	NW t-Stat.	\bar{R}^2 (%)
<i>Panel A. Forecasting r^{EW} (daily)</i>			
CSV ^{EW}	0.380	(3.408)	1.363
RREL	-0.056	(-3.861)	
DIV	0.013	(1.434)	
TERM	0.010	(0.743)	
<i>Panel B. Forecasting r^{CW} (daily)</i>			
CSV ^{CW}	1.337	(2.012)	0.473
RREL	-0.036	(-3.232)	
DIV	0.011	(1.438)	
TERM	0.010	(0.961)	
<i>Panel C. Forecasting r^{EW} (monthly)</i>			
CSV ^{EW}	0.275	(2.262)	5.670
RREL	-1.062	(-3.672)	
DIV	0.647	(1.531)	
TERM	0.078	(0.370)	
DEF	-0.034	(-0.036)	
<i>Panel D. Forecasting r^{CW} (monthly)</i>			
CSV ^{CW}	-0.400	(-1.106)	3.007
RREL	-0.701	(-3.497)	
DIV	0.067	(0.256)	
TERM	0.132	(0.959)	
DEF	0.014	(0.021)	

remains after controlling for the macroeconomic variables, while CSV^{CW} is still a nonsignificant predictor of r^{CW} . These results are consistent with Bali et al. (2005) for a shorter sample ending in Dec. 1999.

2. Inclusion of Return Variance

We now include the variance of the market portfolio in the predictability regression. For the monthly estimations of $\text{var}(r_t^{EW})$, we use the realized sample variance over the month computed from daily returns. For the daily estimations, we fitted an AR(1)-EGARCH(1,1) model on the overall sample.²⁹ In Panels A and B of Table 9, we report regression results at the monthly and daily frequency. In the univariate regression, the variance of the EW portfolio returns does not appear to be significant in explaining the average future returns at the monthly and daily frequencies.

In the bivariate regression, the coefficient of $\text{var}(r_t^{EW})$, ϑ , is negative and nonsignificant at the monthly frequency. At the daily frequency, the coefficient ϑ is still negative and (marginally) significant. For both monthly and daily frequencies, the significance of the CSV coefficient improves slightly after the inclusion of the EW portfolio variance.

²⁹Using the overall sample to estimate the parameters would only give the portfolio variance an advantage to predict future returns. However, we find that even when using such forward-looking estimates for $\text{var}(r_t^{EW})$, the significance of the CSV remains strong.

TABLE 9
Daily and Monthly Return Predictability With and Without Market Volatility

Table 9 presents summary statistics for three predictive regressions. Panels A and B correspond to daily and monthly predictions of the equal-weighted (EW) broad market index using equal-weighted CSV and the realized variance of the EW broad market index, while Panels C and D include the corresponding statistics for the cap-weighted (CW) measures for both returns and CSV. In each panel, the first row corresponds to the regression $r_t^W = \alpha + \beta \text{CSV}_t^W + \epsilon_t$, the second row to $r_t^W = \alpha + \vartheta \text{var}(r_t^W) + \epsilon_t$, and the third row to $r_t^W = \alpha + \beta \text{CSV}_t^W + \vartheta \text{var}(r_t^W) + \epsilon_t$. The t -statistics are reported in parentheses. The sample period is from July 1963 to Dec. 2006.

Panel A. One-Day-Ahead Return Predictions of EW Market Portfolio Using EW CSV

Daily Estimates	Intercept	t -Statistics	CSV ^{EW}	t -Statistics	var(r^{EW})	t -Statistics	\bar{R}^2 (%)
Forecasting r^{EW}	-1.29e-05	(-0.071)	0.411	(4.000)			0.573
Forecasting r^{EW}	0.001	(4.928)			-2.437	(-0.976)	0.034
Forecasting r^{EW}	1.19e-04	(0.629)	0.470	(4.672)	-5.038	(-2.226)	0.737

Panel B. One-Month-Ahead Return Predictions of EW Market Portfolio Using EW CSV

Monthly Estimates	Intercept	t -Statistics	CSV ^{EW}	t -Statistics	var(r^{EW})	t -Statistics	\bar{R}^2 (%)
Forecasting r^{EW}	0.001	(0.270)	0.215	(2.271)			0.678
Forecasting r^{EW}	0.008	(2.673)			0.240	(0.182)	-0.185
Forecasting r^{EW}	0.001	(0.324)	0.226	(2.206)	-0.478	(-0.394)	0.515

Panel C. One-Day-Ahead Return Predictions of CW Market Portfolio Using CW CSV

Daily Estimates	Intercept	t -Statistics	CSV ^{CW}	t -Statistics	var(r^{CW})	t -Statistics	\bar{R}^2 (%)
Forecasting r^{CW}	-1.65e-04	(-0.737)	1.151	(1.966)			0.186
Forecasting r^{CW}	2.10e-05	(0.183)			2.680	(2.003)	0.068
Forecasting r^{CW}	-1.76e-04	(-0.937)	1.084	(1.298)	0.447	(0.198)	0.179

Panel D. One-Month-Ahead Return Predictions of CW Market Portfolio Using CW CSV

Monthly Estimates	Intercept	t -Statistics	CSV ^{CW}	t -Statistics	var(r^{CW})	t -Statistics	\bar{R}^2 (%)
Forecasting r^{CW}	0.008	(2.823)	-0.421	(-1.334)			0.123
Forecasting r^{CW}	0.005	(2.636)			-0.418	(-0.569)	-0.116
Forecasting r^{CW}	0.008	(2.815)	-0.459	(-1.054)	0.130	(0.113)	-0.065

In Panels C and D of Table 9, we report the regression results at monthly and daily frequencies using the cap-weighted index and CSV equivalents. The relation at the daily horizon becomes nonsignificant after the inclusion of the realized variance of the market cap-weighted index. At the monthly horizon, the relation remains nonsignificant.

Other measures of variance have been used in trying to link market returns to a measure of market risk. Option-implied variance (VIX^2) has been used as a forward-looking measure of market variance in addition to realized variance (the sum of squared returns at higher frequency than the targeted frequency for the measure of variance).³⁰ Results with these measures are similar to the ones we just reported.³¹

3. Asymmetry in the Cross-Sectional Distribution of Returns

We now test for an asymmetry in the relation between idiosyncratic variance and future average returns. First, the CSV is split in two and is computed for returns above or below the mean. Second, we compute a robust measure of the cross-sectional skewness.

³⁰For example, for the monthly variance, one will sum the daily squared returns, while for the daily variance, it is customary to use 5- or 1-minute squared returns.

³¹These results are available from the authors.

This asymmetry may be the result of the leverage effect put forward by Black (1976), since we are considering individual firms in the cross section. Consumption volatility risk may also affect differently small and large firms or value and growth firms. Therefore, we test i) whether the predictability power is the same for the CSV of returns to the *left* and *right* of the center of the returns' distribution, ii) whether the relation is driven by one of the sides, and iii) whether the sign of the coefficient is the same on both sides. Therefore, we define CSV_t^+ as the cross-sectional variance of the returns to the right of the cross-sectional distribution (that is, the cross-sectional distribution that includes all stocks such that $r_{it} > r_t^{EW}$) and conversely define CSV_t^- as the cross-sectional variance of the returns to the left of the cross-sectional distribution (i.e., the cross-sectional distribution that includes all stocks such that $r_{it} < r_t^{EW}$). Then we run the following regression:

$$(18) \quad r_{t+1}^{EW} = \alpha + \beta^+ CSV_t^+ + \beta^- CSV_t^- + \epsilon_{t+1}.$$

Table 10 presents the results of regression (18) for daily and monthly estimates. First, we observe that splitting the CSV into right and left sides of the cross-sectional distribution makes the adjusted R^2 of the predictive regression increase from 0.8% to 1.11% in monthly data and from 0.6% to 1.55% in daily data. Second, there is an asymmetric relation between the CSV of the returns to the right and left of the cross-sectional distribution and the expected market return: The coefficient of the CSV_t^+ is positive, while the one of CSV_t^- is negative in both daily and monthly regressions, and, for the daily equal-weighted CSV, the magnitude of the coefficients differs significantly. However, for the monthly regressions, the coefficients are mostly nonsignificant.

The summary statistics of the predictive regression on the cap-weighted index using the equivalent cap-weighted CSV measures, displayed in the right-hand side of Panel A of Table 10, are qualitatively similar to the results on the equal-weighted measures.

These findings suggest that a measure of asymmetry of the cross-sectional distribution would be relevant in the context of exploring the relation between market-expected returns and aggregate idiosyncratic risk. Another key advantage of the cross-sectional nature of the CSV measure is that it can be easily extended to higher-order moments. Therefore, we consider below the skewness of the cross-sectional distribution of returns and assess its predictive power for future returns. To the best of our knowledge, this additional factor, which appears as a natural extension of the CSV for measuring idiosyncratic risk,³² is new in this context.³³ We follow Kim and White (2004) and use a quantile-based estimate (see Bowley (1920)), generalized by Hinkley (1975), as a robust measure of skewness for the cross-sectional distribution of returns:³⁴

³²We show formally in an Appendix available from the authors that there is a link between idiosyncratic skewness and the skewness of the cross-sectional distribution of returns.

³³At the stock level, Kapadia (2009) uses cross-sectional skewness to explain the puzzling finding in Ang et al. (2006) that stocks with high idiosyncratic volatility have low subsequent returns, while Bali, Cakici, and Whitelaw (2011) examine the role of extreme positive returns in the cross-sectional pricing of stocks.

³⁴The usual nonrobust skewness measure of the cross section of returns is highly noisy compared to the proposed robust measure, especially at the daily frequency.

TABLE 10
 Predictability Regression on Market Returns with Right and Left CSV Measures, and Skewness

In Panel A of Table 10, we report the results of daily and monthly predictive regressions of the excess equal-weighted and cap-weighted portfolio returns, denoted by r^{EW} and r^{CW} , respectively. We use as predictors the corresponding daily or monthly measures of the returns cross-sectional variance to the right and the left of the cross-sectional distribution mean, denoted as CSV^+ and CSV^- , respectively. In Panel B, we report results for similar regressions with the returns cross-sectional skewness as the predictive variable, at the daily and monthly frequencies. We include the intercept, the regression coefficients, and the Newey-West (NW) (1987) corrected t -statistics (shown in parentheses), as well as the adjusted coefficient of determination, denoted by \bar{R}^2 . The sample period is from July 1963 to Dec. 2006.

Panel A. Predictability with Right and Left CSV Measures (EW and CW)

	Forecasting r^{EW}		Forecasting r^{CW}	
	Daily ^{EW}	Monthly ^{EW}	Daily ^{CW}	Monthly ^{CW}
Intercept	0.001	0.003	-1.12e-04	0.008
NW t -statistics	(3.944)	(0.727)	(-0.588)	(2.753)
CSV^+	0.488	0.375	4.942	0.071
NW t -statistics	(3.360)	(2.400)	(2.546)	(0.054)
CSV^-	-1.200	-0.486	-2.842	-1.302
NW t -statistics	(-3.551)	(-1.129)	(-2.736)	(-0.879)
\bar{R}^2 (%)	1.552	1.114	0.785	-0.069

Panel B. Predictability with Cross-Sectional Skewness

	Daily r^{EW}			Monthly r^{EW}		
	Coefficients	t -Statistics	\bar{R}^2 (%)	Coefficients	t -Statistics	\bar{R}^2 (%)
Intercept	0.001	(5.200)	5.345	0.008	(3.403)	3.626
Skewness	0.004	(19.989)		0.004	(4.915)	
	Daily r^{CW}			Monthly r^{CW}		
	Coefficients	t -Statistics	\bar{R}^2 (%)	Coefficients	t -Statistics	\bar{R}^2 (%)
Intercept	2.18e-004	(2.576)	0.952	0.006	(2.840)	0.047
Skewness	0.002	(11.601)		0.001	(1.180)	

$$(19) \quad RCS = \frac{F^{-1}(1 - \alpha_1) + F^{-1}(\alpha_1) - 2Q_2}{F^{-1}(1 - \alpha_1) + F^{-1}(\alpha_1)}$$

for any α_1 between 0 and 0.5 and $Q_2 = F^{-1}(0.5)$. The Bowley coefficient of skewness is a special case of Hinkley’s coefficient when $\alpha_1 = 0.25$ and satisfies Groeneveld and Meeden’s (1984) properties for reasonable skewness coefficients. It has upper and lower bounds $\{-1, 1\}$.

In Panels C and D of Table 10 we report the results of a univariate regression of daily and monthly returns on the equal-weighted and cap-weighted market indexes on the robust measure of cross-sectional skewness. The coefficients of skewness in the daily regressions are highly significant. The R^2 for the equal-weighted market returns is above 5% at the daily level. At the monthly frequency, skewness is a significant predictor only for the equal-weighted index. The sign of the coefficient of skewness is positive, which means that a shock resulting in an increase of the positive skewness of the cross-sectional distribution of returns will translate into higher expected returns next period.³⁵ We can invoke the

³⁵Several theoretical papers predict lower expected returns for stocks with idiosyncratic skewness; see, among recent contributions, Mitton and Vorkink (2007), Barberis and Huang (2008), and Brunnermeier, Gollier, and Parker (2007). This literature refers to the pricing of individual securities that exhibit skewness. Our result concerns the relation between aggregate idiosyncratic skewness and the expected aggregate return.

same mispricing arguments as in Section IV.C. Our findings are also consistent with the differentiated impact of investor sentiment on the prices of individual securities. Baker and Wurgler (2006) run predictability regressions for equal-weighted and cap-weighted market returns using proxies for investor sentiment at the monthly frequency. They find that when sentiment is low small stocks earn particularly high subsequent returns. When sentiment is high there is no size effect. This would explain both the positive sign for skewness and the fact that capitalization weighting obscures the pattern.

In Table 11, we report the results of predictive regressions at the daily and monthly frequencies where we add the robust measure of the cross-sectional skewness to the CSV measures and macroeconomic variables included in Table 8. The first observation is that the magnitude and the significance of the CSV measures are very close to the values reported in Table 8. However, the striking fact is that skewness remains the major contributor to the predictability of returns since the R^2 increases significantly compared to the regressions including CSV and macroeconomic variables. At the daily frequency, the adjusted R^2 increases to close to 6% (from 1.4%). At the monthly frequency, it goes up to almost 8% (from 5.7%). This robustness check shows that the skewness effect goes beyond good and bad times defined by macroeconomic conditions.

TABLE 11
Daily and Monthly Predictability of Market Returns with
Skewness and Macroeconomic Variables

Panels A and B of Table 11 include the multivariate regression results for daily predictions of the equal-weighted (EW) broad market index using the equal-weighted CSV (Panel A) or the cap-weighted (CW) CSV (Panel B), cross-sectional skewness and a set of macroeconomic variables as regressors. Panels C and D report the corresponding monthly predictions. The set of macroeconomic variables are the relative T-bill rate (RREL), the aggregate dividend yield (DIV), the term spread (TERM), and the default or credit spread (DEF). The Newey and West (NW) (1987) corrected t -statistics are reported in parentheses. The sample period is from July 1963 to Dec. 2006.

	Coefficient	NW t -Statistics	\bar{R}^2 (%)
<i>Panel A. Forecasting r^{EW} (daily)</i>			
CSV ^{EW}	0.331	(3.205)	6.081
Skewness	0.004	(19.450)	
RREL	-0.045	(-3.475)	
DIV	0.007	(0.901)	
TERM	0.014	(1.142)	
<i>Panel B. Forecasting r^{CW} (daily)</i>			
CSV ^{CW}	1.278	(1.912)	1.241
Skewness	0.002	(10.340)	
RREL	-0.030	(-2.916)	
DIV	0.009	(1.181)	
TERM	0.011	(1.175)	
<i>Panel C. Forecasting r^{EW} (monthly)</i>			
CSV ^{EW}	0.277	(2.403)	7.800
Skewness	0.003	(3.819)	
RREL	-0.869	(-3.108)	
DIV	0.546	(1.445)	
TERM	0.114	(0.586)	
DEF	0.043	(0.053)	
<i>Panel D. Forecasting r^{CW} (monthly)</i>			
CSV ^{CW}	-0.393	(-1.092)	2.829
Skewness	1.47e-004	(0.271)	
RREL	-0.691	(-3.426)	
DIV	0.064	(0.245)	
TERM	0.134	(0.966)	
DEF	0.015	(0.022)	

V. Is Average Idiosyncratic Risk Priced?

Several theoretical papers predict a positive relation between idiosyncratic risk and expected returns. Levy (1978), Merton (1987), and Malkiel and Xu (2002) pricing models relate stock returns to their beta with the market and their beta to market-wide measures of idiosyncratic risk. In these models, an important portion of investors' portfolios may differ from the market. Their holdings may be affected by corporate compensation policies, borrowing constraints, and heterogeneous beliefs and include nontraded assets that add background risk to their traded portfolio decisions (e.g., human capital and private businesses). These theoretical predictions are also in line with Campbell et al.'s (2001) argument that investors holding a limited number of stocks hoping to approximate a well-diversified portfolio would end up being affected by changes in idiosyncratic volatility just as much as by changes in market volatility. More recently, Guo and Savickas (2008) argue that changes in average idiosyncratic volatility provide a proxy for changes in the investment opportunity set and that this proxy is closely related to the book-to-market factor.³⁶

However, there is an ongoing empirical debate on the cross-sectional relation between idiosyncratic volatility and future stock returns. Ang et al. (2006), (2009) find results that are opposite to these theories since stocks with high idiosyncratic volatility earn low average returns, but they cannot fully rationalize this result. Recently, Huang et al. (2010) have found that the negative sign in the relation between idiosyncratic variance and expected returns at the stock level becomes positive after controlling for return reversals. Similarly, Fu (2009) documented that high idiosyncratic volatilities of individual stocks are contemporaneous with high returns, which tend to reverse in the following month. Bali et al. (2011) find that the negative effect of idiosyncratic volatility is driven by its close association with the maximum daily return in a month, proxying for demand for lottery-like stocks. Finally, Bali and Cakici (2008) show that there is no robust, significant relation between idiosyncratic volatility and the cross section of expected returns. The relation is sensitive to portfolio construction methodology and data frequency.

In order to use the standard set of assets in the asset pricing literature, we extract daily returns data from the Kenneth French data library on their 100 (10×10) and 25 (5×5) size/book-to-market portfolios for the period July 1963 to Dec. 2006. Then we run every calendar month the following regression for each portfolio:³⁷

$$(20) \quad r_{it} = \alpha + \beta_{i, \text{XMKT}} \text{XMKT}_t + \beta_{i, \text{SMB}} \text{SMB}_t + \beta_{i, \text{HML}} \text{HML}_t + \beta_{i, \text{CSV}} \text{CSV}_t^{\text{EW}}.$$

To estimate the factor loadings (the β s), we use the past month or the past 3 months of daily data (to reduce measurement error) and run the following

³⁶Alternative explanations of the relation between idiosyncratic risk and return are the firm's assets' call-option interpretation by Merton (1974), where equity is a function of total volatility, as well as the Barberis, Huang, and Santos (2001) prospect theory asset pricing model with loss aversion over (owned) individual stock's variance.

³⁷As before, XMKT stands for excess market return, SMB and HML are the size and book-to-market FF (1993) factors, also directly extracted from the Kenneth French data library (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

cross-sectional regression every month on the next month's excess returns and record the γ coefficients:

$$(21) \quad r_{it+1}^m = \gamma_0 + \gamma_{\text{XMKT}}\beta_{i,\text{XMKT}}(t) + \gamma_{\text{SMB}}\beta_{i,\text{SMB}}(t) + \gamma_{\text{HML}}\beta_{i,\text{HML}}(t) + \gamma_{\text{CSV}}\beta_{i,\text{CSV}}(t).$$

We finally test whether the average γ coefficients are statistically different from 0. In order to take into account possible serial correlation in the coefficients, we compute the t -statistics using NW (1987) standard errors with 4 lags (same number of lags as in Ang et al. (2009)). We use four sets of assets: 100 (10×10) size/book-to-market EW and CW portfolios, and 25 (5×5) size/book-to-market EW and CW portfolios. For each of them, we use the CSV^{EW} as the fourth risk factor. In Table 12 we report the corresponding Fama-MacBeth (1973) regression results. The table displays the annualized coefficients and standard errors (multiplied by 12 from the original monthly values), as well as their corresponding autocorrelation-corrected t -statistics and the average R^2 . In Panel A, for the monthly beta estimates, the γ coefficient for CSV^{EW} is positive and significant when we use the 100 and 25 size/book-to-market FF (1993) equal-weighted portfolios. However, it is positive and not significant when we use the

TABLE 12
Pricing of Cross-Sectional Variances

Table 12 displays the average values and Newey-West (NW) (1987) corrected t -statistics (shown in parentheses) for the coefficients in Fama-MacBeth (1973) regressions, run every month in the sample, using the betas on the 3 Fama-French (FF) (1993) factors and on CSV^{EW} for the 100 and 25 size and book-to-market FF equal-weighted (EW) (first two rows of each panel) and cap-weighted (CW) (last two rows of each panel) portfolios. Panel A uses 1 month of daily data (21 days) to estimate the betas of the factor model regressions every month, while Panel B uses 63 days of daily returns (3 months) to run the factor model regressions every 21 days. The table also displays the average R^2 across subsamples of the Fama-MacBeth regressions. γ is as defined in equation (21). The sample period is July 1963 to Dec. 2006.

	Intercept	XMKT	SMB	HML	CSV ^{EW}	\bar{R}^2 (%)
<i>Panel A. Fama-MacBeth Monthly Regressions</i>						
<i>100-EW Portfolios</i>						
γ	0.223	-0.067	0.029	0.048	0.005	24.657
t -statistics	(9.222)	(-4.173)	(2.320)	(3.916)	(2.847)	
<i>25-EW Portfolios</i>						
γ	0.278	-0.133	0.042	0.064	0.009	51.962
t -statistics	(9.969)	(-6.447)	(2.480)	(3.927)	(2.703)	
<i>100-CW Portfolios</i>						
γ	0.155	-0.007	0.004	0.037	0.003	24.262
t -statistics	(6.896)	(-0.421)	(0.323)	(2.762)	(1.909)	
<i>25-CW Portfolios</i>						
γ	0.175	-0.034	0.010	0.048	0.004	50.815
t -statistics	(7.417)	(-1.624)	(0.625)	(2.910)	(1.395)	
<i>Panel B. Fama-MacBeth Daily Regressions</i>						
<i>100-EW Portfolios</i>						
γ	0.691	-0.297	0.152	0.174	0.033	31.163
t -statistics	(7.574)	(-4.480)	(2.872)	(3.340)	(2.468)	
<i>25-EW Portfolios</i>						
γ	0.891	-0.512	0.175	0.205	0.065	59.323
t -statistics	(7.615)	(-5.915)	(2.839)	(3.105)	(2.164)	
<i>100-CW Portfolios</i>						
γ	0.424	-0.042	0.047	0.149	0.017	30.022
t -statistics	(5.653)	(-0.709)	(0.940)	(2.723)	(1.436)	
<i>25-CW Portfolios</i>						
γ	0.518	-0.147	0.063	0.148	0.031	56.889
t -statistics	(5.914)	(-2.002)	(1.103)	(2.450)	(1.198)	

25 market cap-weighted portfolios and marginally significant for the 100 market cap-weighted portfolios. The latter result is not entirely surprising, considering that the CSV in returns is reduced through the market-capitalization adjustment. In Panel B, we report results for the 3-month beta estimates to account for some potential measurement errors. The coefficient for the CSV loading is now significant for the 100 cap-weighted portfolios but not quite for the 25.

VI. Conclusion

We formally introduced a cross-sectional dispersion measure as a proxy for aggregate idiosyncratic risk that has the distinct advantage of being readily computable at any frequency, with no need to estimate other parameters. It is therefore a model-free measure of idiosyncratic risk. We extensively show how this measure is related to previous proxies of idiosyncratic variance, such as the measures proposed by Goyal and Santa-Clara (2003) and Bali et al. (2008), and the ones based on the Fama and French (1993) and CAPM models, which have been previously shown to be very close to the Campbell et al. (2001) proxy as well. We confirm previous findings of Goyal and Santa-Clara (2003), Bali et al. (2005), and Wei and Zhang (2005) on the monthly predictability regressions for an extended sample period using our cross-sectional measure and more standard measures of idiosyncratic variance. We find that the results are robust across these measures. Thanks to the instantaneous nature of our measure, we are able to extend to daily data the evidence on the predictability power of idiosyncratic variance on the future market portfolio return. We provide a statistical argument to support the choice of an equal-weighted measure of average idiosyncratic variance as opposed to a market-cap weighted measure and explain why both empirically and theoretically such a measure should better forecast the equal-weighted market return. We also showed that this cross-sectional measure displays a sizable correlation with economic uncertainty, as measured by consumption-growth volatility, and with several economic and financial variables. One additional advantage of the cross-sectional nature of our measure is that it generalizes to higher moments. We showed that the asymmetry of the cross-sectional distribution is a very good predictor for future returns. We leave for further research an exhaustive analysis of the properties of the skewness of cross-sectional return distribution as a measure of average idiosyncratic skewness as well as an empirical analysis of the CSV measure using international data.

Appendix A. Proof of Proposition 1

Consider the factor model decomposition

$$r_t^{(w_r)} = \sum_{i=1}^{N_t} w_{it} \beta_{it} F_t + \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}$$

and

$$r_{it} - r_t^{(w_r)} = \left(\beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right) F_t + \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt}.$$

Under the homogeneous betas assumption, we have

$$(A-1) \quad r_{it} - r_t^{(w_t)} = \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt}$$

and therefore

$$\left[r_{it} - r_t^{(w_t)} \right]^2 = \varepsilon_{it}^2 + \left(\sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} \right)^2 - 2\varepsilon_{it} \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt}$$

so that

$$\begin{aligned} \text{CSV}_t^{(w_t)} &= \sum_{i=1}^{N_t} w_{it} \left(r_{it} - r_t^{(w_t)} \right)^2 \\ &= \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2 + \left(\sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} \right)^2 - 2 \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} w_{jt} w_{it} \varepsilon_{it} \varepsilon_{jt} \\ &= \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2 - \left(\sum_{i=1}^{N_t} w_{jt} \varepsilon_{jt} \right)^2. \end{aligned}$$

We now argue that the term $\sum_{i=1}^{N_t} w_{jt} \varepsilon_{jt}$ converges to 0 for increasingly large numbers of stocks. To show this, we need to use recent results regarding the Marcinkiewicz–Zygmund strong law of large numbers for weighted sums of independent and identically distributed (i.i.d.) variables (see Sung (2002)):

$$(A-2) \quad \frac{1}{N} \sum_{i=1}^N a_{Ni} X_i \xrightarrow{N \rightarrow \infty} 0 \quad \text{almost surely (a.s.)},$$

when $\{X, X_N, N \geq 1\}$ is a sequence of i.i.d. random variables with $E(X) = 0, E|X|^2 < \infty$ and $\{a_{Ni}, 1 \leq i \leq N, N \geq 1\}$ is an array of constants satisfying³⁸

$$(A-3) \quad \sum_{i=1}^N a_{Ni} = o\left(\frac{1}{N}\right).$$

Here, we take $a_{Ni} \equiv w_{it}$ and $X_i \equiv \varepsilon_{it}$ at all dates t . Thus, for a positive weighting scheme satisfying condition (A-3), it follows from the strong law of large numbers for weighted sums of i.i.d. variables that

$$\sum_{i=1}^{N_t} w_{it} \varepsilon_{it} \xrightarrow{N_t \rightarrow \infty} 0 \quad \text{a.s.}$$

Using similar arguments and the homogeneous idiosyncratic second moment assumption, $E[\varepsilon_{it}^2] \equiv \sigma_\varepsilon^2(t)$, we obtain that for a strictly positive weighting scheme, w_t , and i.i.d. ε_i ,

$$\sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2 \xrightarrow{N_t \rightarrow \infty} \sigma_\varepsilon^2(t) \quad \text{a.s.}$$

³⁸See corollary 2 in Sung (2002) (with $p = q = 2$). Note that the convergence result holds under slightly weaker assumptions (see Cuzick (1995) or corollary 2 from Sung (2002)).

Using these results, we finally have that

$$\text{CSV}_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} \left(r_{it} - r_t^{(w_t)} \right)^2 \xrightarrow{N_t \rightarrow \infty} \sigma_\varepsilon^2(t) \quad \text{a.s.}$$

Appendix B. Properties of the CSV Estimator

1. Bias of the CSV Estimator

Under the factor model decomposition (1) and equation (2) and using the homogeneous beta assumption, we have

$$(B-1) \quad r_{it} - r_t^{(w_t)} = \left(\beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right) F_t + \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} = \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt}.$$

Replacing result (B-1) in equation (3), we have as before

$$(B-2) \quad \text{CSV}_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2 - \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} w_{jt} w_{it} \varepsilon_{it} \varepsilon_{jt}.$$

By definition of a strict factor model, $E[\varepsilon_{it} \varepsilon_{jt}] = 0$ for $i \neq j$, and $E(\varepsilon_{it}^2) = \sigma_{\varepsilon_i}^2$. Applying the expectation operator in equation (B-2), we get

$$(B-3) \quad E \left[\text{CSV}_t^{(w_t)} \right] = \sum_{i=1}^{N_t} w_{it} \sigma_{\varepsilon_i}^2(t) - \sum_{i=1}^{N_t} w_{it}^2 \sigma_{\varepsilon_i}^2(t).$$

The second term in equation (B-3) implies that the CSV would tend to underestimate the average idiosyncratic variance. Considering the EW scheme where $w_{it} = 1/N_t \forall i$, equation (B-3) simplifies into

$$E \left[\text{CSV}_t^{\text{EW}} \right] = \left(1 - \frac{1}{N_t} \right) \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_{\varepsilon_i}^2(t),$$

and we obtain, as N_t goes to infinity,

$$E \left[\text{CSV}_t^{\text{EW}} \right] - \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_{\varepsilon_i}^2 \rightarrow 0.$$

2. Variance of the CSV Estimator

Let w_t and ε_t be column vectors of the weighting scheme and residuals, respectively, and $\Omega_t = w_t w_t'$, $\Lambda_t = \text{diag}(w_t)$, $N_t \times N_t$ matrices, and denote Σ^ε the variance covariance matrix of the residuals, which is diagonal for a strict factor model.

For a finite number of stocks in the case where $F_t \neq r^{(w_t)}$, we have from equation (B-2)

$$\text{CSV}_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2 - \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} w_{jt} w_{it} \varepsilon_{it} \varepsilon_{jt}.$$

Letting $Q_t = \Lambda_t - \Omega_t$, CSV_t can be written in matrix form, as

$$(B-4) \quad \text{CSV}_t^{(w_t)} = \varepsilon_t' Q_t \varepsilon_t.$$

Using the quadratic structure of the CSV and assuming normal residuals, we have (see, e.g., Kachman (1999)):³⁹

$$(B-5) \quad \text{var} \left(\varepsilon_t' Q_t \varepsilon_t \right) = 2\text{tr} \left(Q_t \Sigma_t^\varepsilon Q_t \Sigma_t^\varepsilon \right).$$

Under the assumption of a strict factor model (i.e., $\rho_{ijt}^\varepsilon = 0 \forall i \neq j$), equation (B-5) simplifies to

$$(B-6) \quad \text{var} \left(\text{CSV}_t^{(w_t)} \right) = 2 \sum_{i=1}^{N_t} \sigma_{\varepsilon_{it}}^4 w_{it}^2 (1 - w_{it})^2 + 2 \sum_{i=1}^{N_t} \sum_{j \neq i}^{N_t} w_{it}^2 w_{jt}^2 \sigma_{\varepsilon_{it}}^2 \sigma_{\varepsilon_{jt}}^2.$$

Assuming an upper bound for the individual idiosyncratic variances, denoted as $\bar{\sigma}_{\varepsilon_t}^2$, equation (B-6) yields to the following inequality (replacing each variance for its upper bound):

$$(B-7) \quad \text{var} \left(\text{CSV}_t^{(w_t)} \right) < 2 \bar{\sigma}_{\varepsilon_t}^4 \left(\left(\sum_{i=1}^{N_t} w_{it}^2 \right)^2 + \sum_{i=1}^{N_t} w_{it}^2 - 2 \sum_{i=1}^{N_t} w_{it}^3 \right).$$

When $w_t = 1/N_t$, equation (B-7) simplifies to

$$(B-8) \quad \text{var} \left(\text{CSV}_t^{(w_t)} \right) < 2 \bar{\sigma}_{\varepsilon_t}^4 \left(\frac{N_t - 1}{N_t^2} \right) < 2 \bar{\sigma}_{\varepsilon_t}^4 \left(\frac{1}{N_t} \right).$$

For a large number of stocks,

$$(B-9) \quad \text{var} \left(\text{CSV}_t^{(w_t)} \right) \rightarrow 0.$$

Appendix C. Relaxing the Assumption of Homogeneous Betas

The assumption that $\beta_{it} = \beta_t$ for all i is obviously a simplistic one and is done only for exposure purposes. Starting with the single-factor decomposition on the definition of the CSV, we have

$$\begin{aligned} \text{CSV}_t^{(w_t)} &= \sum_{i=1}^{N_t} w_{it} \left(r_{it} - r_t^{(w_t)} \right)^2 \\ &= \sum_{i=1}^{N_t} w_{it} \left[\left(\beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right) F_t + \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} \right]^2. \end{aligned}$$

After simple rearrangement of terms, we get

$$\begin{aligned} \text{CSV}_t^{(w_t)} &= F_t^2 \sum_{i=1}^{N_t} w_{it} \left(\beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right)^2 + \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2 - \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} w_{jt} w_{it} \varepsilon_{it} \varepsilon_{jt} \\ &\quad + 2 F_t \sum_{i=1}^{N_t} w_{it} \varepsilon_{it} \left(\beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right). \end{aligned}$$

³⁹The operator tr stands for the trace of a matrix, which is the sum of the diagonal terms.

Applying the expectation operator and assuming a strict factor model, the last expression simplifies to yield

$$E \left[\text{CSV}_t^{(w_i)} \right] = \sum_{i=1}^{N_t} w_{it} \sigma_{\varepsilon_{it}}^2 - \sum_{i=1}^{N_t} w_{it}^2 \sigma_{\varepsilon_{it}}^2 + E \left[F_t^2 \text{CSV}_t^\beta \right].$$

References

- Allen, L., and T. Bali. "Cyclicality in Catastrophic and Operational Risk Measurements." *Journal of Banking and Finance*, 31 (2007), 1191–1235.
- Ang, A.; R. Hodrick; Y. Xing; and X. Zhang. "The Cross-Section of Volatility and Expected Returns." *Journal of Finance*, 61 (2006), 259–299.
- Ang, A.; R. Hodrick; Y. Xing; and X. Zhang. "High Idiosyncratic Volatility and Low Returns: International and Further U.S. Evidence." *Journal of Financial Economics*, 91 (2009), 1–23.
- Baker, M., and J. Wurgler. "Investor Sentiment and the Cross-Section of Stock Returns." *Journal of Finance*, 61 (2006), 1645–1680.
- Bali, T., and N. Cakici. "Idiosyncratic Volatility and the Cross Section of Expected Returns." *Journal of Financial and Quantitative Analysis*, 43 (2008), 29–58.
- Bali, T.; N. Cakici; and H. Levy. "A Model-Independent Measure of Aggregate Idiosyncratic Risk." *Journal of Empirical Finance*, 15 (2008), 878–896.
- Bali, T.; N. Cakici; and R. Whitelaw. "Maxing Out: Stocks as Lotteries and the Cross-Section of Expected Returns." *Journal of Financial Economics*, 99 (2011), 427–446.
- Bali, T.; N. Cakici; X. Yan; and Z. Zhang. "Does Idiosyncratic Risk Really Matter?" *Journal of Finance*, 60 (2005), 905–929.
- Bansal, R., and A. Yaron. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance*, 59 (2004), 1481–1509.
- Barberis, N., and M. Huang. "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices." *American Economic Review*, 98 (2008), 2066–2100.
- Barberis, N.; M. Huang; and T. Santos. "Prospect Theory and Asset Prices." *Quarterly Journal of Economics*, 116 (2001), 1–53.
- Bekaert, G.; R. J. Hodrick; and X. Zhang. "Aggregate Idiosyncratic Volatility." *Journal of Financial and Quantitative Analysis*, 47 (2012), 1155–1185.
- Black, F. "The Pricing of Commodity Contracts." *Journal of Financial Economics*, 3 (1976), 167–179.
- Bollerslev, T.; G. Tauchen; and H. Zhou. "Expected Stock Returns and Variance Risk Premia." *Review of Financial Studies*, 22 (2009), 4463–4492.
- Bowley, S. *Elements of Statistics*, Vol. 2, London, UK: P. S. King and Son (1920).
- Brandt, M.; A. Brav; J. Graham; and A. Kumar. "The Idiosyncratic Volatility Puzzle: Time Trend or Speculative Episodes?" *Review of Financial Studies*, 23 (2010), 863–899.
- Brunnermeier, M. K.; C. Gollier; and J. A. Parker. "Optimal Beliefs, Asset Prices, and the Preference for Skewed Returns." *American Economic Review*, 97 (2007), 159–165.
- Campbell, J.; M. Lettau; B. Malkiel; and Y. Xu. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk." *Journal of Finance*, 56 (2001), 1–43.
- Cuzick, J. "A Strong Law for Weighted Sums of I.I.D. Random Variables." *Journal of Theoretical Probability*, 8 (1995), 625–641.
- Fama, E., and K. French. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, 33 (1993), 3–56.
- Fama, E. F., and J. D. MacBeth. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy*, 81 (1973), 607–636.
- French, K. R.; G. W. Schwert; and R. F. Stambaugh. "Expected Stock Returns and Volatility." *Journal of Financial Economics*, 19 (1987), 3–29.
- Fu, F. "Idiosyncratic Risk and the Cross-Section of Expected Stock Returns." *Journal of Financial Economics*, 91 (2009), 24–37.
- Goyal, A., and P. Santa-Clara. "Idiosyncratic Risk Matters!" *Journal of Finance*, 58 (2003), 975–1008.
- Groeneveld, R., and G. Meeden. "Measuring Skewness and Kurtosis." *Statistician*, 33 (1984), 391–399.
- Guo, H., and R. Savickas. "Average Idiosyncratic Volatility in G7 Countries." *Review of Financial Studies*, 21 (2008), 1259–1296.
- Hinkley, D. "On Power Transformations to Symmetry." *Biometrika*, 62 (1975), 101–111.
- Huang, W.; Q. Liu; S. Rhee; and L. Zhang. "Return Reversals, Idiosyncratic Risk, and Expected Returns." *Review of Financial Studies*, 23 (2010), 147–168.

- Kachman, S. D. "Distribution of Linear Transformations and Quadratic Forms." Working Paper, University of Nebraska, Lincoln (1999).
- Kapadia, N. "The Next Microsoft? Skewness, Idiosyncratic Volatility, and Expected Returns." Working Paper, University of North Carolina at Chapel Hill (2009).
- Kim, T., and H. White. "On More Robust Estimation of Skewness and Kurtosis." *Finance Research Letters*, 1 (2004), 56–73.
- Levy, H. "Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio." *American Economic Review*, 68 (1978), 643–658.
- Malkiel, B., and Y. Xu. "Idiosyncratic Risk and Security Returns." Working Paper, University of Texas at Dallas (2002).
- Merton, R. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance*, 29 (1974), 449–470.
- Merton, R. "A Simple Model of Capital Market Equilibrium with Incomplete Information." *Journal of Finance*, 42 (1987), 483–510.
- Mitton, T., and K. Vorkink. "Equilibrium Underdiversification and the Preference for Skewness." *Review of Financial Studies*, 20 (2007), 1255–1288.
- Newey, W., and K. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica: Journal of the Econometric Society*, 55 (1987), 703–708.
- Pontiff, J. "Costly Arbitrage and the Myth of Idiosyncratic Risk." *Journal of Accounting and Economics*, 42 (2006), 35–52.
- Ross, S. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory*, 13 (1976), 341–360.
- Sharpe, W. "A Simplified Model for Portfolio Analysis." *Management Science*, 9 (1963), 277–293.
- Sung, S. H. "Strong Laws for Weighted Sums of I.I.D. Random Variables." *Bulletin of the Korean Mathematical Society*, 39 (2002), 607–615.
- Tédongap, R. "Consumption Volatility and the Cross-Section of Stock Returns." *Review of Finance*, forthcoming (2014).
- Wei, S., and C. Zhang. "Idiosyncratic Risk Does Not Matter: A Re-Examination of the Relationship between Average Returns and Average Volatilities." *Journal of Banking and Finance*, 29 (2005), 603–621.
- Welch, I., and A. Goyal. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies*, 21 (2008), 1455–1508.