# Idiosyncratic volatility: the myth-busting effect of measurement error and market structure*

Markus Leippold† Michal Svatoň‡

August 17, 2018

**Abstract**

We reexamine the evolution of idiosyncratic volatility (IV) of US firms between 1962 and 2016, and its relation to expected returns. The unusually high variance of new firmly listed stocks from 1962 to 1997 materializes in a trend in the IV and, contrary to the conclusions of related studies, in systemic risk. A decreasing market concentration attenuates the trend in the latter before 1997. The subsequent reversal of IV stems from the bias in correlations before 1997 due to measurement errors among existing illiquid stocks. Young and distressed stocks are crucial for obtaining the negative IV-expected return relation.

**Keywords:** Idiosyncratic volatility, measurement error, asynchronicity, IV-puzzle.

**JEL classification:** G11, G12, G14.

---

*We thank Per Ostberg, Alexandre Ziegler, Catherine Bruneau, and the seminar participants at the University of Zurich and the AFFI Conference 2018 (Paris).

†University of Zurich, Department of Banking and Finance, Plattenstrasse 14, 8032 Zurich, Switzerland; e-mail: markus.leippold@bf.uzh.ch.

‡Corresponding author. University of Zurich, Department of Banking and Finance, Plattenstrasse 14, 8032 Zurich, Switzerland; e-mail: michal.svaton@bf.uzh.ch.

# 1 Introduction

The empirical finance literature provides abundant evidence on the failure of seemingly reasonable economic theories. These puzzles give rise to suspicions that something must be wrong with these established theories. However, what if the data contains some hidden features that, if neglected, could be misleading the analysis? What if quantities related to market structure, such as market concentration or changes in the cross-sectional sample size, and related to return and volatility measurements, such as the share of non-traded stocks or bid-ask spreads, are time-varying and impact statistical inferences?

In this paper, we argue that many results from the empirical finance literature, which rely on daily returns of US small stocks from the CRSP database before 1997, should be carefully revisited or, at least, taken with a grain of salt. In particular, we challenge two empirical findings on idiosyncratic volatility, which have emerged as pervasive stylized facts in empirical asset pricing. The first one is concerned with the trends in aggregate and idiosyncratic (or firm-level) volatility, as discussed in Campbell et al. (2001) (henceforth CLMX).[1] The second one is related to the puzzling negative relationship between stock returns and idiosyncratic volatility, as advocated by Ang et al. (2006).[2]

There seems to be a consensus that the US equity market in the 1980s and 1990s was characterized by a strengthening role of idiosyncratic volatility relative to the systemic component of market risk. At the same time, more recent studies indicate that the aggregate level of idiosyncratic risk reverted around 1997 and systemic risk regained importance.[3] We challenge the assertion of such a reversal and show that it stems from a systematic bias in correlation estimates before 1997 due to measurement errors in the CRSP database, rather than from fundamental factors. Furthermore, we challenge the view that the increase in variance was restricted to the idiosyncratic component, as indicated by CMLX, and argue that the systemic risk had a similar trend, partially attenuated by decreasing market concentration.

[1] See, also Xu and Malkiel (2003) and Comin and Philippon (2005), among many others.

[2] While the literature on the trend of firm-level volatility has been less controversial, the literature on the idiosyncratic volatility puzzle is more so, as we discuss below.

[3] See, e.g., Brown and Kapadia (2007), Bali et al. (2008), Brandt et al. (2010), and Bekaert et al. (2012). These studies provide various explanations of the change in idiosyncratic volatility, such as changes in institutional ownership and growth options, among others.

The structure of the US equity market has undergone many significant changes over the last decades. At least two of them might have a significant impact on IV measurement. First, we have witnessed a dramatic change in the market size over the sample period, which complicates the interpretation of the time series of volatility components obtained from the CLMX decomposition. In particular, it is a priori not clear whether the CLMX idiosyncratic volatility is driven by changes in the covariance structure of already listed firms, by the potentially different characteristics of newly listed firms, or by the dilution of market weights. Our second concern is that the definition of idiosyncratic volatility by CLMX highly depends on the applied industry classification.[4] To address these issues, we therefore provide a generalization of the variance decomposition used in CLMX. Our method helps us to explicitly take into account the potential effect of new listings and delistings, and industry reclassification. Moreover, we can theoretically identify the three main drivers that influence the IV estimate: average variances, the market concentration, and the correlation structure.

From our empirical analysis, we learn that a significant portion of the documented changes in the idiosyncratic volatility (IV) is an artifact of using daily CRSP price data, which are inappropriate for return correlation measurement before 1997.[5] We show that the high-IV period coincides with a high frequency of zero returns in the database. These zero returns are concentrated among illiquid stocks with high bid-ask spreads. They often occur, because a significant portion of price entries (and hence returns) in the CRSP database, and all NASDAQ SmallCap stocks before 1997 in particular, are computed as an average of the bid and ask prices. As the percentage of zero daily returns peaks around 50% shortly after inclusion of NASDAQ stocks, it is apparent that the effect on the estimated correlations is quantitatively significant. Our observation is also in line with the findings of Bali et al. (2008) who argue that the increase in the idiosyncratic volatility concentrates among NASDAQ stocks with a low price, for which the measurement issue is the most severe.

Switching to a lower frequency, which is more robust to measurement error, we make three observations. First, the trend in value-weighted IV is weaker and significant only on a 90% level of confidence. Second, the reversal after 1997 essentially disappears. Third, the idiosyncratic volatility

[4]A motivation to design a methodology that takes into account a potentially dynamic industry classification is given by the work of Hoberg and Phillips (2010) and Hoberg and Phillips (2017), who apply text-based industry classification schemes.

[5]While previous studies acknowledged the differences of IV estimates based on returns of different frequency, the historical increase and the subsequent reversal remains a common conclusion.

becomes relatively more important after 1997, contrary to what previous literature has found in the daily data. The inconsistency between the IV series based on different frequencies poses a challenge for standard fundamental explanations of the IV evolution.

Our results not only have far reaching implications for analyzing the trend in idiosyncratic volatility, but also for many other empirical applications and research questions that build on historical daily stock returns from the CRSP database before 1997. Hence, we additionally put a focus on the ongoing and prominent debate over the role of the idiosyncratic risk on asset prices. Ang et al. (2006, 2009) (henceforth AHXZ) document a systematic underpricing of stocks with low idiosyncratic volatility and strong overpricing of firms with high ones. The effect is found to be the strongest in the 1980s and 1990s, in line with the documented surge in the idiosyncratic volatility and rapidly changing market composition.

There are various attempts to reconcile the puzzling findings of AHXZ with theory. The first group of studies tries to explain the negative relation by an omitted pricing factor, where the volatility proxies for the missing loading on a systemic factor like the VIX in Ang et al. (2006), the average volatilities as in Chen and Petkova (2012), or the interest rate exposure as in Beilo et al. (2017). A second group explains the mispricing as a consequence of market frictions.[6] The negative relation between IV and expected returns is challenged by Fu (2009) and Branger et al. (2017), who find a positive one.[7] We contribute to this debate by showing that the negative relation is mostly driven by young stocks and pre-delisting effects, and disappears with their exclusion. Alphas of the IV-portfolios remain decreasing in the IV, but are insignificant and captured by the exposure to the betting against beta (BAB) factor of Frazzini and Pedersen (2014).

We organize the paper as follows. In Section 2, we review the CLMX framework and idiosyncratic volatility measurement, which we extend and generalize in Section 3. This generalization allows us

[6]For instance, Stambaugh et al. (2015) suggest that high volatility translates to a higher risk of exploiting pricing anomalies, thus reducing the willingness of arbitrageurs to trade in these stocks. Han and Lesmond (2011) highlight the role of market microstructure on the IV measurement, but Chen et al. (2012) conclude that the stocks most prone to these effects, in fact, make the IV-puzzle weaker. Hou and Loh (2016) summarize a range of other explanations and show that they can only partially explain the IV effect.

[7]Fu (2009) used a GARCH model to substantiate his claim. To come up with reasonable estimates for the GARCH model, he required a minimum of 30 monthly returns. Thus, he (unintentionally) eliminates young stocks. Hence, in light of our results, Fu's conclusion contradicting that of AHXZ is not surprising. A similar argument applies to Branger et al. (2017), as they use only stocks on which there are liquid options traded.

to identify a range of long-term drivers of idiosyncratic volatility, which we analyze empirically in Section 4. We then turn to the relation between IV and expected return. Section 5, provides empirical evidence on the role of sampling, post-listing, and pre-delisting effects. In Section 6 we discuss the main results and their implications. Section 7 concludes.

## 2 Revisiting CMLX

As a starting point, we first review the methodology of CLMX, which starts from a decomposition of stock returns into three components, a market return, an industry-specific return, and a firm-specific return. Following their notation, we denote the excess market return at time $t$ by $R_{mt}$, the excess return of industry $i$ by $R_{it}$, and the excess return of firm $j$ belonging to industry $i$ by $R_{jit}$. For the market return, CMLX assume that the capital asset pricing model (CAPM) holds. Hence, we can write industry and firm returns as:

$$R_{it} = \beta_{im} R_{mt} + \tilde{\epsilon}_{it}, \tag{1}$$

$$R_{jit} = \beta_{ji} R_{it} + \tilde{\eta}_{jit} = \beta_{ji}\beta_{im} R_{mt} + \beta_{ji}\tilde{\epsilon}_{it} + \tilde{\eta}_{jit}, \tag{2}$$

where $\tilde{\epsilon}_{it}$ and $\tilde{\eta}_{jit}$ denote the industry and firm innovations, and $\beta_{im}$ and $\beta_{ji}$ are the corresponding beta coefficients. To avoid estimation of market betas, CLMX follow Campbell et al. (1997, Chapter 4) and use the "market-adjusted-return model" with the following modified residuals:

$$\epsilon_{it} = R_{it} - R_{mt}, \tag{3}$$

$$\eta_{jit} = R_{jit} - R_{it}. \tag{4}$$

By $w_{it}$ we denote the weight of industry $i$ in the total market and by $w_{jit}$ the weight of firm $j$ in industry $i$. Furthermore, we define

$$\mathrm{Var}(\tilde{\epsilon}_t) = \sigma^2_{\tilde{\epsilon}t} = \sum_i w_{it}\sigma^2_{\tilde{\epsilon}it}, \tag{5}$$

$$\mathrm{Var}(\tilde{\eta}_t) = \sigma^2_{\tilde{\eta}t} = \sum_i w_{it} \sum_{j\in i} w_{jit}\sigma^2_{\tilde{\eta}jit}, \tag{6}$$

as the variance of the industry and firm-level residuals, respectively. We refer to the variance of the firm level residuals interchangeably as idiosyncratic variance. We follow CLMX and estimate the individual variances $\sigma^2_{\eta jit}$ ($\sigma^2_{\epsilon it}$) by realized variance measure, summing squared daily returns of firm (industry) within a given month. For the market-adjusted model, the terms in equations (5) and (6) are defined accordingly to obtain $\sigma^2_{\epsilon t}$ and $\sigma^2_{\eta t}$. The estimates of these variances are related by

$$\sigma^2_{\epsilon t} = \sigma^2_{\tilde{\epsilon} t} + \mathrm{CSV}_t(\beta_{im})\sigma^2_{mt}, \tag{7}$$

$$\sigma^2_{\eta t} = \sigma^2_{\tilde{\eta} t} + \mathrm{CSV}_t(\beta_{im})\sigma^2_{mt} + \mathrm{CSV}_t(\beta_{ji})\sigma^2_{\tilde{\epsilon} t}, \tag{8}$$

where $\sigma_{mt}$ is the market volatility, and

$$\mathrm{CSV}_t(\beta_{im}) := \sum_i w_{it}(\beta_{im} - 1)^2, \quad \mathrm{CSV}_t(\beta_{ji}) := \sum_i w_{it} \sum_{j \in i} w_{jit}(\beta_{ji} - 1)^2$$

denote the cross-sectional variations of betas, which drive the deviations of the actual average variances from the model and the ones from the "market-adjusted" model. Clearly, the advantage of using the market-adjusted model is that it avoids the explicit estimation of market betas.[8]

To be confident that we build our analysis on the same subset of data like CLMX, we revisit their main empirical result. We use stock price data from CRSP spanning the period from January 1962 to December 2016. Hence, compared to the study of CLMX, which spans the period from July 1962 to December 1997, our sample period starts six months earlier and ends nine years later. As a consequence, our dataset includes a jump in the number of firms in July 1962 when AMEX stocks were added to the database. A second spike occurs in December 1972 with the inclusion of NASDAQ stocks. For our analysis, we include all common stocks listed on NYSE, AMEX, and NASDAQ with share codes 11 and 12.[9] For the risk-free rate, we use the annualized one-month Treasury rate provided by CRSP and convert it to a daily rate by assuming 260 business days per year.[10]

[Figure 1 about here.]

[8] Its use is justified by the small magnitude of the cross-sectional variations of betas. In particular, we find that the correlation between the increments of the exact and the "market-adjusted" computation is 0.89 and 0.97 for industry and firm variances, respectively. Hence, we find the quality of the approximation to be adequate, especially for idiosyncratic variance, our main object of interest.

[9] This restriction is standard and its impact on the final result is negligible. Moreover, the notion of idiosyncratic volatility for funds (e.g., index ETFs) has a very different interpretation to that for common stocks.

[10] It might be more appropriate to consider the actual number of business days in each month and adjust the risk-free rate accordingly. However, we find that in the context of equity returns any difference from this adjustment for the daily risk-free rate is quantitatively negligible.

In Figure 1, we plot the resulting time series of constructed measures of industry and idiosyncratic volatility, $\sigma_{\epsilon t}$ and $\sigma_{\eta t}$, as well as the volatility of market returns within each month. We can confirm the result for the CLMX sample and observe a clear upward trend in the idiosyncratic volatility.[11] However, consistent with Bekaert et al. (2012), we find the opposite in the post-CLMX sample. The IV drops even below levels observed for most of the CLMX sample period. This trend is robust to alternative IV measures.[12] Therefore, we use the CLMX measure throughout the study.

# 3 Generalization of the CLMX Framework

Having replicated the findings of the previous literature on the IV trend, we first provide a generalization of the CLMX methodology for identifying IV. As argued in the introduction, our goal is to find a decomposition that is robust to changes in market size and industry classifications. To do so, we first need to introduce some definitions.

## 3.1 Classification matrices

As a first definition, we present the concept of a classification matrix, which introduces considerable flexibility when we map firms to industries.

**Definition 1.** *Let $F$, $S \in \mathbb{N}$, $F \geq S$. For each $s \in \{1, \dots, S\}$, let $\mathcal{S}_s$ denote a nonempty set of elements $f \in \{1, \dots, F\}$. An $S \times F$ classification matrix $C = [c_{sf}]$ has the following properties:*

*a) The elements $c_{sf}$ are given by $c_{sf} = \mathbb{1}_{\{f \in \mathcal{S}_s\}}$, i.e., they are either zero or one.*

*b) Each $f$ belongs to exactly one $\mathcal{S}_s$, i.e., $e_S^\top C = e_F$, where by $e_N$ we denote the one-vector of size $N$.*

*c) Classes are nonempty, i.e., $C_{s\cdot} e_F > 0$, where $C_{s\cdot}$ is the $s$-th row of $C$.*

[11] Throughout this study we use "trend" in reference to a deterministic trend. CLMX rejected the hypothesis of unit root for all three volatility components, but also rejected the absence of a deterministic trend.

[12] In particular, similar patterns emerge when we measure IV relative to the three-factor model of Fama and French (1993), as a residual from a principal component approximation as in Herskovic et al. (2016), or using the model-free approach of Bali et al. (2008) and Bekaert et al. (2012).

In the above definition, we can think of $F$ as firms and $S$ as sectors or industries. Two extreme cases are possible. In the first case, we call $C_F = I_F$ the firm classification, where by $I$ we denote the identity matrix. Each industry will only have one firm. In the second case, there is only one industry, i.e., $C_M = e_F^\top$ which we call the market classifier.

**Definition 2.** *Let $C_C \in \mathbb{R}^{S_C \times F}$ and $C_F \in \mathbb{R}^{S_F \times F}$ be two classification matrices. We call $C_F$ the refinement of $C_C$, if there exists a classification matrix $C_{C,F} \in \mathbb{R}^{S_C \times S_F}$ such that $C_C = C_{C,F} C_F$.*

The refinement means that whenever we know the firm category on a finer level, we can infer the category on a coarser level, but not vice versa. For example, the firm classifier refines the industry classification, which further refines the market classifier. This definition is useful for extending market-industry-firm stratification to one with multiple layers, and for describing the relation between IV relative to industry and IV measured directly relative to the market.

**Definition 3.** *We call a sequence of classification matrices $C_t = \{C_{l,t}\}_{l=0}^{L+1}$ the local classification of degree $L$ at time $t$, if $C_{0,t} = C_{M_t}$, $C_{L+1,t} = C_{F_t}$ and $C_{l,t}$ is a refinement of $C_{l+1,t}$ for $l = 0, \ldots, L$. A sequence of local classifications $C = \{C_t\}_{t \in \mathbb{T}}$ with an index set $\mathbb{T}$ is called full classification (over $\mathbb{T}$).*

By allowing the size of classification matrices to vary with time, we acknowledge the fact that the classification may change due to new listings and delistings or as a consequence of reclassification of some firms.

### 3.2 Layer variances and weighting functions

Given a classification matrix and a stratification into, e.g., a market, industry, a firm layer, we can now define for each of these layers a variance measure. To this end, we consider an $\mathbb{R}^F$-valued random vector $r$ of simple returns and $R = \mathrm{diag}(r)$. Furthermore, by $v \in \mathbb{R}^F_{++}$ we denote the vector of lagged market capitalizations and by $w \in (0,1]^F$ the corresponding relative market weights.[13]

[13]We consider the market capitalization $v$ to be nonzero. This assumption is not very restrictive as exchanges have minimum price requirements.

**Definition 4.** *For a full classification $C$ with degree $L$, period $T = (T_1, T_2]$, and return and variance weighting functions $\psi : T \to T$ and $\nu : T \to T$, a layer $l \in \{1, \ldots, L+1\}$ variance is defined as*

$$\hat{\sigma}^2_{l-1:l,T}(\psi, \nu) = \sum_{t \in T} \epsilon_{l,t}^\top (C_{l,\nu(t)} V_{\nu(t)} C_{l,\nu(t)}^\top) \epsilon_{l,t} (e_{F_t}^\top V_{\nu(t)} e_{F_t})^{-1}, \tag{9}$$

*where $V = \mathrm{diag}(v)$ and*

$$\begin{aligned} \epsilon_{l,t} &= R_{l,t} e_{S_l} - C_{l-1,l,\psi(t)}^\top R_{l-1,t} e_{S_{l-1}} && (10) \\ &= \left( C_{l,\psi(t)} V_{\psi(t)} R_t C_{l,\psi(t)}^\top (C_{l,\psi(t)} V_{\psi(t)} C_{l,\psi(t)}^\top)^{-1} e_{S_l} \right) \\ &\quad - C_{l-1,l,\psi(t)}^\top \left( C_{l-1,\psi(t)} V_{\psi(t)} R_t C_{l-1,\psi(t)}^\top (C_{l-1,\psi(t)} V_{\psi(t)} C_{l-1,\psi(t)}^\top)^{-1} e_{S_{l-1}} \right). && (11) \end{aligned}$$

For our empirical analysis, we focus on idiosyncratic variance and industry-layer variances. These variances are obtained by setting $L = 1$ in equation (9). Hence, we have $\hat{\sigma}^2_{1:2,T}(\psi, \nu)$ for idiosyncratic variance and $\hat{\sigma}^2_{0:1,T}(\psi, \nu)$ industry-layer variance.

For the weighting functions $\psi$ and $\nu$, we can make different choices. For the CLMX specification, we have

$$\psi^{CLMX}(t) = t, \quad \nu^{CLMX}(t) = T_1 + 1. \tag{12}$$

Hence, CMLX compute industry (market) returns based on the previous day's weight but use the weights at the beginning of the month to weight individual firms' (industries') variances. We consider different specifications, which we refer to as time consistent and which enable us to derive additional properties of idiosyncratic variance. We can define a dynamic weight specification by setting

$$\psi^{DW}(t) = t, \quad \nu^{DW}(t) = t, \tag{13}$$

i.e., we vary the weights with time, but we always use the same weights for the return computation and variance weighting. Similarly, we can define a constant weight specification as

$$\psi^{CW}(t) = T^*, \quad \nu^{CW}(t) = T^*, \tag{14}$$

for some $T^* \in \mathbb{N}$, i.e., we keep the weights and classifications fixed at the values prevailing at time $T^*$. For our empirical application, we focus on the case $T^* = T_1 + 1$, using the lagged market capitalization associated with the first observation in period $T$.

In practice, the choice of the weighting scheme has a negligible impact on the results of the empirical analysis, especially when IV is computed over one month or shorter periods. The correlation of innovations (differences) of the monthly IV series with different combinations of weighting used for the construction of industry returns and to weight individual stock variance is above 0.98. Intuitively, a typical change in market weights over the course of one month is relatively small to drive the results.

**Proposition 1.** *Let $\mathbb{T}$ be an index set, $C = (C_t)_{t\in\mathbb{T}}$ be a full classification of degree $L$. With consistent weighting $\psi$, the layer $l$ variance for period $T = (T_1, T_2]$ satisfies*

$$\begin{aligned}\hat{\sigma}^2_{l-1:l,T} =& \operatorname{tr}\Bigg[\sum_{t\in T} W_{\psi(t)}(r_t r_t^\top)W_{\psi(t)}\left[C^\top_{l,\psi(t)}(C_{l,\psi(t)}W_{\psi(t)}C^\top_{l,\psi(t)})^{-1}C_{l,\psi(t)}\right.\\ &\left.- C^\top_{l-1,\psi(t)}(C_{l-1,\psi(t)}W_{\psi(t)}C^\top_{l-1,\psi(t)})^{-1}C_{l-1,\psi(t)}\right]\Bigg],\end{aligned} \tag{15}$$

*where $W = \operatorname{diag}(w)$ is the diagonal matrix of lagged market weights. If further $\psi = \psi^{CW}$, then*

$$\begin{aligned}\hat{\sigma}^2_{l-1:l,T} =& \operatorname{tr}\Bigg[W_{T^*}\hat{\Sigma}_T W_{T^*}\left[C^\top_{l,T^*}(C_{l,T^*}W_{T^*}C^\top_{l,T^*})^{-1}C_{l,T^*}\right.\\ &\left.- C^\top_{l-1,T^*}(C_{l-1,T^*}W_{T^*}C^\top_{l-1,T^*})^{-1}C_{l-1,T^*}\right]\Bigg],\end{aligned} \tag{16}$$

$$=\bar{\hat{\sigma}}^2_{l,T} - \bar{\hat{\sigma}}^2_{l-1,T}, \tag{17}$$

*where we use $\hat{\Sigma}_T = \sum_{t\in T} r_t r_t^\top$ as an estimator for $\Sigma_T$, the (conditional) covariance matrix of period $T$ excess returns.*[14] *Furthermore, $\bar{\hat{\sigma}}^2_{l,T} = \sum_{i=1}^{S_l} w_{T^*,i,l}\hat{\sigma}^2_{T,i,l}$ denotes the value-weighted average variance of layer $l$, and $w_{T^*,i,l}$, $\hat{\sigma}^2_{T,i,l}$ are the weight and the variance estimate of the $i$-th member of layer $l$ in period $T$.*

Proposition 1 obviates that even though the betas are not explicitly estimated, the layer variance measures still depend on the (estimated) covariance matrix. Consequently, the gain from the CLMX methodology is merely computational and negligible, given today's computing power. Nevertheless, for better comparability to the CLMX results, we use the market-adjusted model, which can also be motivated by the small magnitude of the resulting error. Furthermore, for $T^*$ we use the start of the interval $T_1$ to ensure that, in a given month, the market weights are not affected by the returns and $\hat{\sigma}^2_{l-1:l,T}$ depends on the returns only via $\hat{\Sigma}_T$.

[14] We implicitly assume that the mean is relatively small on the considered interval. Neglecting the mean has an advantage that it avoids its estimation which is notoriously difficult, especially over short samples.

In addition, Proposition 1 has the following consequences. Let $\hat{\sigma}^2_{l-k:l,T}$ denote the variance at layer $l$ computed relative to layer $l-k$. Then,

$$\hat{\sigma}^2_{l-2:l-1,T} + \hat{\sigma}^2_{l-1:l,T} = \hat{\sigma}^2_{l-2:l,T}, \tag{18}$$

i.e., the layer variances add up. This property implies that idiosyncratic volatility measured with a finer industry classification is smaller than or equal to the one based on a coarser classification (e.g., SIC four-digit codes versus two-digit codes). Hence, if we measure the IV directly relative to the market and drop the industry layer, the resulting variance will be the sum of firm-layer and industry-layer variance in our setting.

The layer variance measure in equation (16) equals the trace of the product of a weighted covariance matrix, $W\hat{\Sigma}W$, and a block-diagonal matrix of the form

$$[C_{l,t}^\top(C_{l,t}W_tC_{l,t}^\top)^{-1}C_{l,t} - C_{l-1,t}^\top(C_{l-1,t}W_tC_{l-1,t}^\top)^{-1}C_{l-1,t}], \tag{19}$$

with block sizes determined by the coarser classification $C_{l-1,t}$. By the properties of the trace operator, it follows that the layer variance depends only on the covariances within that block. The firm-level volatility measured relative to the industry is therefore independent of covariance structure between individual industries. Alternatively, the independence of the IV of inter-industry correlations becomes evident from equation (17), which expresses the IV as a difference between average stock and industry variances, which are independent of those correlations.

The inclusion of an additional firm, while keeping everything else fixed, has an ambiguous effect on the idiosyncratic volatility estimate. We can clarify this ambiguity in a market with a single industry. If the market had initially consisted of only one stock (meaning that the idiosyncratic volatility is zero), expanding the market by an imperfectly correlated stock will increase idiosyncratic volatility. In contrast, by including a stock and arbitrarily increasing its size, the idiosyncratic volatility will shrink towards zero as the stock will become systemic. The latter case is improbable under realistic circumstances, as will be seen in Section 4.2. The opposite holds for firm exclusions.

Next, we define the following measures of variance and correlation, which we call the constant equivalent variance and correlation, respectively. These measures reflect more accurately the impact of both the variance and correlation channels on IV. For the definition, we abandon matrix notation, and we suppress time-dependence for notational convenience. Then, the idiosyncratic variance given in equation (9) with $L = 1$ reads as follows:

$$\hat{\sigma}^2_{L:L+1} = \sum_{i=1}^{F} w_i \hat{\sigma}_i^2 - \sum_{s=1}^{S} \frac{\sum_{i,j \in \mathcal{S}_s} w_i w_j \hat{\sigma}_i \hat{\sigma}_j \hat{\rho}_{ij}}{\sum_{i \in \mathcal{S}_s} w_i}, \tag{20}$$

with $\mathcal{S}_s$ denoting the set of firms in industry $s$.

**Definition 5.** *For period $T$ and given sets of firms in industries $(\mathcal{S}_s)_{s=1}^S$, let $\hat{\Sigma} = (\hat{\sigma}_{ij}) \in \mathbb{R}^{F \times F}$ denote the covariance matrix estimate, $\hat{\sigma}^2_{L:L+1}$ the firm-level variance with consistent weighting and weights $w \in (0,1]^F$. The period $T$ constant equivalent (CE) variance and correlation are defined as scalars $\hat{\sigma}^2_{CE}$ and $\hat{\rho}_{CE}$ that satisfy*

$$\hat{\sigma}^2_{L:L+1} = \hat{\sigma}^2_{CE} \sum_{i=1}^{F} w_i - \hat{\sigma}^2_{CE} \sum_{s=1}^{S} \frac{\sum_{i,j \in \mathcal{S}_s} w_i w_j \hat{\rho}_{ij}}{\sum_{i \in \mathcal{S}_s} w_i}, \tag{21}$$

$$\hat{\sigma}^2_{L:L+1} = \sum_{i=1}^{F} w_i \hat{\sigma}_i^2 - \sum_{s=1}^{S} \frac{\sum_{i \in \mathcal{S}_s} w_i^2 \hat{\sigma}_i^2 + \sum_{i,j \in \mathcal{S}_s, i \neq j} w_i w_j \hat{\sigma}_i \hat{\sigma}_j \hat{\rho}_{CE}}{\sum_{i \in \mathcal{S}_s} w_i}. \tag{22}$$

These measures provide a useful scalar summary of the covariance structure. The CE correlation represents a correlation value for which the IV with equicorrelated returns would match the observed one. Similarly, the CE variance is a scalar variance which matches the firm-level volatility, in case the variances of all stocks are equal. The CE correlation depends on variance estimates, and the CE variance depends on the correlations within the industry.

### 3.3 A simplified example

To gain some additional insights on the drivers of the IV estimate, we consider the following simplified example of a one-period market. We assume that the market consists of $F$ firms distributed among $S$ equally sized industries with $F/S$ (assumed integer) firms in each, i.e., $C = I_S \otimes e_{F/S}^\top$ in terms of the classification matrix $C$.[15] Next, we assume that the weight distribution, characterized by a diagonal weighting matrix $W$, satisfies $W = W_S \otimes W_{F/S}$ for some diagonal matrices $W_S$ and $W_{F/S}$. $W_S$ governs the capital distribution between industries and $W_{F/S}$ describes the firms' weights within each industry, i.e., we assume that each industry has an identical capital distribution. By

[15]This formulation assumes that, without loss of generality, the firms are sorted according to their industry.

$\Sigma = \mathbb{E}[\hat{\Sigma}]$ we denote the true covariance matrix and assume it takes the form

$$\Sigma = \Sigma_S \otimes ((1-\rho)I_{F/S} + \rho e_{F/S} e_{F/S}^\top), \tag{23}$$

which implies that each industry has an identical correlation matrix, corresponding to the equicorrelated returns within the industry. The specification of $\Sigma_S$ in equation (23) allows us to introduce inter-industry correlations and different variances for different industries. These assumptions imply

$$\mathbb{E}\left[\hat{\sigma}^2_{L:L+1}\right] = \operatorname{tr}\left(\Sigma_S W_S\right)\left(1 - \operatorname{tr}[W^2_{F/S}]\right)(1-\rho)\,. \tag{24}$$

Hence, we can identify three main drivers that influence the IV estimate, average variances $\operatorname{tr}\left(\Sigma_S W_S\right)$, market concentration $\operatorname{tr}[W^2_{F/S}]$, and correlations $\rho$.

Under the assumptions of Example 3.3, the expected value of the estimator $\mathbb{E}[\hat{\sigma}^2_{L:L+1}]$ depends multiplicatively on the average variance $\operatorname{tr}\left(\Sigma_S W_S\right)$, the capital concentration within industries measured by weighted industry Herfindahl-Hirschman index $\bar{\mathcal{H}} = \operatorname{tr}[W^2_{F/S}]$, and the stock correlation $\rho$ in each industry. This observation hints already at a potential source of the observed trend in idiosyncratic volatility: the market concentration decreases over the CLMX sample, but not afterward. The number of listed firms follows a similar pattern. With equal weights within equally sized industries, we have $\bar{\mathcal{H}} = \frac{S}{F}$. Hence, the idiosyncratic variance explicitly depends on the number of firms per industry. In a more realistic setting, the IV will depend on a nontrivial interaction of these effects.

### 3.4 Decomposing the volatility series

The above example shows that IV measurement critically depends on three distinct channels, namely market concentration, correlations, and average variances. Changes in these channels may either originate from existing firms or from changes in market composition due to listing of new firms and delisting of existing firms. Therefore, to shed further light on the impact of a changing market composition on IV, we decompose changes in IV to changes in market size (listing and delisting), reclassification of stocks (to a different industry), and a part reflecting changes in the market weights or the covariance structure among existing stocks, which from now on we refer to as 'fundamentals.'[16]

Let $F_T = \{V_T, \hat{\Sigma}_T\}$ denote the fundamentals in period $T$, i.e., a pair of the firms' (initial) capitalizations $V_T$ and the estimated covariance matrix $\hat{\Sigma}_T$, and let $C_T$ denote the local classification in period $T$. Slightly abusing notation, we write

$$\gamma(F_T, C_T) = \hat{\sigma}_{1:2,T}(\psi, \nu) \tag{25}$$

as the idiosyncratic volatility given in Definition 4 for $L = 1$. Further, we denote those variables restricted to companies available in both $T-1$ and $T$ using the superscript '$\star$'. Then, we can decompose the change in IV, defined as

$$\epsilon_T = \gamma(F_T, C_T) - \gamma(F_{T-1}, C_{T-1}), \tag{26}$$

in the following way:

$$\epsilon_T = \epsilon_T^{NEW} + \epsilon_T^{OLD} + \epsilon_T^{RC}(w^{RC}) + \epsilon_T^{F}(w^{F}), \tag{27}$$

where $\epsilon_T^{NEW}$ is the component attributable to newly listed companies given as

$$\epsilon_T^{NEW} = \gamma(F_T, C_T) - \gamma(F_T^\star, C_T^\star). \tag{28}$$

Similarly, $\epsilon_T^{OLD}$ is the delisting component given by

$$\epsilon_T^{OLD} = \gamma(F_{T-1}^\star, C_{T-1}^\star) - \gamma(F_{T-1}, C_{T-1}). \tag{29}$$

The term $\epsilon_T^{RC}$ captures the effect of reclassifying of a company between industries and the term $\epsilon_T^{F}$ is the fundamental component, capturing the effect of changes in the covariance structure and the

[16]The advantage of our approach, compared to the regression analysis of Brown and Kapadia (2007), is that our method is also directly applicable to the industry-layer and the market volatility. Furthermore, the regression method of Brown and Kapadia (2007) will likely overstate the listing effect around events of high (idiosyncratic) volatility, especially if the lifespan of the firms listed in that period is short. Furthermore, because our method compares the IV of newly listed firms to that of existing firms in the same or the previous month, it is less prone to the heteroskedasticity effects.

market weights among existing firms. These two terms are given by

$$\begin{aligned}
\epsilon_T^F(w^F) &= w^F(\gamma(F_T^\star, C_T^\star) - \gamma(F_{T-1}^\star, C_T^\star)) \\
&\quad +(1-w^F)(\gamma(F_T^\star, C_{T-1}^\star) - \gamma(F_{T-1}^\star, C_{T-1}^\star)), \qquad (30)\\
\epsilon_T^{RC}(w^{RC}) &= w^{RC}(\gamma(F_T^\star, C_T^\star) - \gamma(F_T^\star, C_{T-1}^\star)) \\
&\quad +(1-w^{RC})(\gamma(F_{T-1}^\star, C_T^\star) - \gamma(F_{T-1}^\star, C_{T-1}^\star)), \qquad (31)
\end{aligned}$$

and they depend on the weights $w^{RC} \in [0,1]$ and $w^F \in [0,1]$, which determine whether the effect is computed at the initial or terminal fundamentals and classification, respectively. In our application, we consider only the balanced case with $w^{RC} = w^F = 0.5$. Finally, we define

$$\epsilon_T^{SIZE} := \epsilon_T^{NEW} + \epsilon_T^{OLD} \qquad (32)$$

to capture the joint effect of changing market size and composition.[17] Analogously, we decompose market and industry-layer volatility.

# 4 Empirical Analysis of idiosyncratic volatility

Having laid out the theoretical framework in the previous section, we can now empirically analyze in detail the three channels influencing IV, i.e., market concentration, correlations, and average variances. Since these channels are influenced by changes in the market composition, we first explore the role of existing firms and the impact of listing and delisting of firms using the volatility decomposition from Section 3.4.

## 4.1 The role of existing, newly listed, and delisted firms

In equation (27), we decomposed changes in IV to changes in market size (listing and delisting), reclassification of stocks (to a different industry), and a part related to the underlying fundamentals.

[17]For the combined effect an alternative definition

$$\tilde{\epsilon}_T^{SIZE} := \epsilon_T^{NEW} + \epsilon_{T+1}^{OLD} = \gamma(F_T^\star, C_T^\star) - \gamma(F_T^{\star\star}, C_T^{\star\star}),$$

can be considered, with the superscript '$\star\star$' indicating a restriction to firms available in periods $T$ and $T+1$. The advantage of this measure is that both the impact of newly listed and delisted companies is measured with the fundamentals belonging to the same period. However, for our application this is not critical as we focus on longer-term trends.

To visualize the contributions of the individual components, we construct the recovered “level series” from the component series. This construction has to be understood as a mere graphical device; the level series is not a valid standard deviation, because each IV increment (and its components) is conditional on the previous value of measured idiosyncratic volatility, not the level series of a given component. As a consequence, the recovered level series may become negative even for a variance measure with nonnegative support.[18]

In Figure 2, we plot the volatility decomposition for different volatility layers. We can draw three conclusions. First, the decomposition of the firm-layer volatility in Panel A suggests that we can attribute the IV increase in the CLMX period to the size component and not to the fundamental component, which exhibits a stable mean in that period. Hence, the positive trend originates from newly listed and delisted firms, not from the existing firms in the market, confirming the conclusions of Brown and Kapadia (2007). We note that the size component also captures well the events with size impacts, like the inclusion of NASDAQ stocks into the sample, the dot-com bubble, and the financial crisis. The effect of reclassification is negligible - the cumulative series of the reclassification component is mostly flat.

[Figure 2 about here.]

Second, in Panel A of Figure 2, there is a significant decrease in the fundamental component at the beginning of the post-CLMX period. At the same time, we observe in Figure 1 a decrease in IV during this period, which was also confirmed by Bekaert et al. (2012), among others. As the size component continuous to grow, we conclude that the arguments of Brown and Kapadia (2007) must be held irrelevant for the post-CLMX period. Instead, the decrease of IV must be related to a change in the covariance structure of already listed companies. We will get to the bottom of the cause of this change in Section 4.3.

Third, Panels B and C of Figure 2 indicate that the size component in industry-layer and market volatility is weaker than in the IV case, and is driven by the delisting of firms rather than the

[18] As a robustness check, we applied the decomposition also on other variances measures $\gamma$, like variance levels and their logarithms. The resulting components are qualitatively similar to our baseline measure, the volatility, and are available upon request.

arrival of new firms. We explain this difference by a diversification effect. Both industry and market variances are variances of a portfolio and, hence, decrease with an addition of a stock, unless the stock has abnormally large weight or variance. Because the OLD component is a difference between the volatility of a portfolio with fewer and one with more constituents, we can expect it to be positive on average. The converse implication holds for the NEW component. Therefore, what surprises us is the relatively flat level series of the NEW component, not the increase in the OLD component. We interpret this finding as evidence that the diversification in the NEW component is offset by either higher variances or correlations of newly listed companies.

As our analysis of Figure 2 suggests, listed and delisted stocks play an important role for changes in IV, particularly if their variances and correlations differ substantially from those of other stocks in the market. These differences may occur due to a fundamentally different nature of stocks listed or delisted in that period, or because stocks after listing or delisting are affected by transitory effects related to those events. While Brown and Kapadia (2007) have argued that the increase in IV during the CLMX period is solely due to newly listed firms with higher volatility and is unrelated to the firms' age, we revisit both the transitory age-related post-listing and pre-delisting effects below.[19]

There are many reasons a firm will be delisted from the exchange. Inspecting the historical counts and shares of delistings, as specified by the different CRSP delisting categories, we find that delistings are dominated by two categories, namely mergers and bankruptcies.[20] For a significant part of our sample, mergers constitute more than 70% of delistings, with a notable exception of the first half of the 1990s and during the financial crisis. Value-weighted metrics are almost entirely driven by mergers, which on average represent 85% of the delisted market capitalization.

Mergers and bankruptcies will have different impacts on the stock's volatility behavior before it gets delisted. We expect a firm before bankruptcy to be more volatile due to increased leverage. However, the bankruptcy category is, value-wise, negligible as it is composed of distressed stocks

[19] Pástor and Veronesi (2003) show that younger firms have, on average, higher volatility than their competitors. This pattern is consistent with their model, which predicts such behavior due to higher uncertainty about their profitability. The substantial variability of IPOs also supports the increased uncertainty about pricing of young firms' stock returns in the cross-section (see, Lowry et al. (2010)).

[20] The categories are specified by their first digit in the CRSP delisting codes: Mergers (code $2**$), Exchanges for different assets (code $3**$), Liquidations (code $4**$), and "Bankruptcy" (code $5**$), which covers events such as bankruptcies, insufficiency of equity or assets, drop in price to unacceptable level, among others. For full description see `http://www.crsp.com/products/documentation/delisting-codes`.

with low value. Therefore, its impact on IV is negligible as well. Before a merger, the volatility of the stock depends on how the merger is settled and when the merger deal is announced. After the announcement of a cash-settled merger, the stock of the acquired company essentially becomes a bond, thereby decreasing the stock's volatility, which in most cases leads to a decrease of IV.[21] In case the merger is settled in stocks of the acquiring company, the target's stock functionally turns into a forward contract on the acquirer's stock, leading to an increased correlation between both stocks, particularly towards the deal's closure. As indicated by equation (24), an increase in correlation will lead to a decrease in IV. However, as with cash-settled mergers, the impact on IV is in general ambiguous, but under realistic conditions, we expect the IV to decrease with the announcement.

To investigate whether the data reflects our reasoning above, we compare the variance of stocks in different age baskets. We use the sample variance using one month of daily returns to estimate the stocks' variances in given month. To capture the post-listing and pre-delisting effects, we express the firm's age either as the number of months after listing or before delisting.[22] A direct comparison of variances is prone to endogeneity issue, since both listing and delisting might be more likely in periods of high (or low) volatility. Therefore, we mitigate the effect of time-varying volatility by using volatility ranks in a given month. Still, it is possible that a priori more (or less) volatile firms are more likely to go bankrupt or become a merger target. Hence, to control for firm-specific effects as well, we estimate a fixed-effect model with dummy variables for each firm and month and with stock variances (logs and levels) as the dependent variable.

[Figure 3 about here.]

In Figure 3, we plot the residuals against the firms' age, measured both from the listing and backward from the delisting.[23] Clearly, and in line with Brown and Kapadia (2007), significant post-listing effects are absent (Panels A and C). However, Panels B and D indicate some significant

[21] The IV will increase only if $\sigma_i < \sum_{j \neq i, j \in \mathcal{S}_s} \tilde{w}_j \sigma_j \rho_{ij}$ with $\tilde{w}_j$ being the normalized weights of other firms in the relevant industry. This criterion requires that correlations before announcement were high and the volatility of the stock low relative to (weighted) variances of other stock within the industry.

[22] We compute the firm's age on a calendar basis. Therefore, the first month of the company is often incomplete. Similarly, we calculate the number of months before delisting for each month-company observation.

[23] One limitation of our analysis is that for companies with a short lifetime it is challenging to disentangle firm-specific, post-listing, and delisting effects. An alternative is to directly include the dummy variables for different age baskets in the regression. The results are similar and available upon request.

patterns before the delisting event. For variance levels, a positive spike appears in the last few months before delisting. A fixed-effect regression on log-variances shows the opposite pattern with very low volatility before delisting. This finding indicates the presence of both extremely high variances, driving the levels residuals up, and extremely low variances (i.e., strongly negative log-variances), reverting the spike for log-residuals.

Splitting the residuals by the delisting code sheds further light on the mixed results about the variance behavior before delisting. Panels E and F in Figure 3 show that mergers (bankruptcies) have abnormally low (high) volatilities. Due to strong positive skewness of (realized) variance levels, the effect of bankruptcies is quantitatively large and dominates the merger effect in computation of the overall pre-delisting effect on levels. Taking logarithm of variances reduces the skewness of their distribution, dampens the positive effect of bankruptcies, and amplifies the negative effect of mergers on variance. The negative effect of mergers on log-variances, in conjunction with their higher count, outweighs the positive effect of bankruptcies in average pre-delisting effect computation. The hump in pre-delisting effect stems from the fact, that the effect of mergers materializes only during the last few months before delisting.[24]

Summarizing the above discussion, we find no evidence in favor of post-listing effects on variance. Firms after the announcement of a merger have a low variance if the merger is cash-settled and a high correlation with the acquirer if it is stock-settled. Firms before bankruptcy have an abnormally high variance. Because IV is positively related to variances and negatively to correlations, the presence of pre-merger firms has an overall negative effect on IV. In contrast, firms before bankruptcy tend to drive IV up. Because the merger effect is weak on levels, bankruptcy category will dominate when firms are weighted equally. This effect is most pronounced in 1980s and early 1990s when bankruptcies were more frequent, but the impact of bankruptcies on value-weighted IV is negligible due to their low capitalization. Thus, the only pre-delisting effect that could fuel the increase in value-weighted IV would be a significant drop in pre-merging firms. However, during the CMLX period, the percentage of mergers was actually increasing from less than 0.1% to more than 0.8% of

[24]We also analyze the distribution of log-residuals after listing and before delisting and manifest that the means are not driven by few outliers, with the proportion of the negative residuals decreasing towards delisting for bankruptcies, and increasing for mergers. Furthermore, the merger impact on variances is particularly substantial for cash-settled mergers. Stock-settled mergers essentially have nil effect on variances, but as expected, the correlation with the acquirer's stock increases towards delisting. These results can be obtained from the authors.

total firms.

### 4.2 Market concentration

We next analyze the role of market concentration, one of the channels, which according to equation (24) have a direct impact of IV measurement. Theoretically, equation (24) connotes a negative impact of market concentration on measured idiosyncratic volatilities. Intuitively, this impact stems from a weaker diversification effect in industry variances, driving the IV down. We consider two measures of market concentration. The first measure is a market-wide Herfindahl-Hirschman index $\mathcal{H}$,

$$\mathcal{H} = \sum_i w_i^2, \tag{33}$$

with $w_i$ denoting capitalization-based weights of firms. The second measure is a weighted industry concentration index $\bar{\mathcal{H}}$,

$$\bar{\mathcal{H}} = \sum_s \sum_{j \in \mathcal{S}_s} w_j \sum_{i \in \mathcal{S}_s} \left( \frac{w_i}{\sum_{k \in \mathcal{S}_s} w_k} \right)^2 = \sum_s w_s \sum_{i \in \mathcal{S}_s} \left( \frac{w_i}{w_s} \right)^2, \tag{34}$$

with $w_s$ denoting the weight of industry $s$. The former is more appropriate if we want to measure the diversification effect in the market portfolio, which affects market variance and, by equation (17), also industry-layer variance. The latter is more suitable for studying an average diversification effect in industry portfolios, whose variance drives the industry-layer variance and the IV.

[Figure 4 about here.]

In Figure 4, we plot the different concentration indices for value-weighted and equal-weighted volatilities. In Panel A, the period of the increase in value-weighted IV as perceived by CLMX coincides with a period of decreasing market and industry concentration. The concentration stabilizes in the post-CLMX period when the IV in Figure 1 tends to decline again. A similar pattern emerges when we treat the firms in the sample equally, assigning them equal weights. In Panel B, the decrease in the concentration based on equal weights is negatively related to the number of listed firms in the market and industries. Indeed, in the CLMX period, we witness a substantial increase in the number of listed firms and a decrease afterward.[25]

In the simplified example of Section 3.3, the expected squared IV in equation (24) depends multiplicatively on the industry concentration via $1 - \bar{\mathcal{H}}$. While over the CLMX period the IV roughly doubled, the values for $1-\mathcal{H}$ and $1-\bar{\mathcal{H}}$ increased by less than 1.3% and 17.8% peak-to-trough, respectively. Therefore, we conclude that the contribution of a change in market concentration is quantitatively small and insufficient to explain most of the observed trend in the CLMX sample. Similarly, the relatively stable market concentration after 1997 cannot capture the drop in IV during that period. Other forces must be at work.

### 4.3 Correlation

The next candidate for driving the evolution of IV is the correlation channel, with correlations negatively related to the IV as suggested by equation (24). We start our analysis by computing an equal-weighted average of all pairwise correlations in each month. We estimate the correlations on a rolling window of one year by their sample counterpart, using daily and monthly returns. Figure 5 displays the resulting correlation series. The average correlation with daily data increases significantly after CLMX sample period. We presume that the recent drop in the IV stems from the observed increase in correlations. Such a conjecture resonates nicely with our observation in Figure 2 that the decrease in idiosyncratic volatility in the post-CLMX period is attributable to fundamentals, i.e., changes in the covariance matrix structure and market weights of existing firms.

[Figure 5 about here.]

What strikes us in Figure 5 is the wedge between average correlations based on different frequency in the CLMX period, which disappears afterwards. Because we use identical set of firms for both series, the wedge cannot be explained by exclusion of a subset of firms. Therefore, the difference between both series must stem from negative (cross-)autocorrelations of daily returns in the CLMX sample, which shrunk in the post-CLMX period.

[25] Doidge et al. (2017) provide an excellent discussion of the causes for the changing trend in the number of listed firms, highlighting a decreased propensity to listing, partially explained by regulatory changes and developments in financial markets.

Both the magnitude and the disappearance of the wedge after 1997 seem difficult to justify on grounds of changing fundamental factors. Instead, we argue that the wedge arises due to microstructure effects, specifically due to asynchronicity bias in sample correlations based on daily data. The bias arises due to the fact that computation of sample covariance requires pairs of returns covering identical period, while in practice trades of different assets occur at different times. Often, prices are "synchronized" to a fixed grid by using the last known price value. This procedure biases sample correlation towards zero, and more so with a fine grid which generates flat segments in measured prices. In high-frequency literature, the phenomenon of decreasing correlations between stocks at higher frequencies is referred to as Epps effect.[26]

We label zero-return observations as "non-trades", despite the fact that in most cases the trade volume in the stock on a given day is nonzero. Often, non-trades occur due to the way prices are recorded in the CRSP database: If the closing price is not available, the price is computed as a midpoint of bid and ask quotes. In such case, the price will remain flat unless the inside quote is revised or filled, so zero returns are also consistent with low volumes.

One of factors contributing to low volume are high transaction costs, e.g. wide bid-ask spreads (BAS, see Bollen et al. (2004)). Price discreteness (minimum tick requirements) further widens BAS,[27] and in lower volumes as a consequence, especially for low-price stocks. For NASDAQ stocks, Christie and Schultz (1994) show that in 1991 (and likely before), when the tick size was 1/8 USD, dealers avoided odd eights. Therefore, the role of price discreteness is likely amplified in NASDAQ sample.

First, we discuss the plausibility of the illiquidity explanation of the frequency of non-trades. In Panels A and B of Figure 6 we plot equal-weighted share of zero-return observations and average percentage BAS[28], respectively. Both series are strikingly similar, shooting up with the inclusion of NASDAQ, and decreasing around major tick-size reductions: AMEX in September 1992, NYSE, NASDAQ and AMEX in June 1997, and the "decimalization" of quotes in April 2001 ordered by Security and Exchange Commission. The relation between percentage BAS and non-trades holds

[26] See, Epps (1979).
[27] See, Hasbrouck (1999), Zhang et al. (2008), Bessembinder (2000).
[28] See Appendix B for details of BAS computation.

also in cross-section, with correlation of 0.6 in the pooled sample of firm-month pairs. Both BAS and non-trade are low in post-CLMX sample, compared to the CLMX period after inclusion of NASDAQ.

Second, we argue that asynchronicity bias due to non-trades has a non-trivial effect on measured correlations and IV. The proportion of prices computed as a quote midpoint amounts to 20% of all day-firm observations in the pooled sample, and the percentage of affected prices varies strongly over time. In particular, all prices in the NASDAQ SmallCap Market before June 15, 1992 are computed this way. These two facts are reflected in Panel A of 6 – percentage of non-trades fluctuates around 40% in the CLMX sample once NASDAQ is included, and starts decreasing after 1992. Furthermore, the fact that small NASDAQ stocks are the ones most prone to asynchronicity aligns with evidence of Bali et al. (2008), who show that stocks with those characteristics drive the trend in the IV in the CLMX sample.

One possible concern is that the equal-weighted share of non-trades overstates the impact of zero returns on value-weighted measure of IV, as the affected stocks are small in size. The VW series in Panel A, which measures the capitalization share of non-trades, confirms that non-trades contribute to smaller share of capitalization, relative to their count. However, capitalization share understates the impact of asynchronicity bias on the IV, because each non-trade generates bias in all correlations with affected asset.

To gain better estimate of the asynchronicity bias, we consider the following simplified setup. We assume that non-trades for asset $i$ occur with probability $\lambda_i$, and independently of other assets' non-trades and returns. Then, the estimated correlation $\hat{\rho}_{ij}$ is approximately[29] related to the true correlation of monthly returns via

$$\hat{\rho}_{ij} \approx \rho_{ij}(1-\lambda_i)(1-\lambda_j) := \rho_{ij}B_{ij}. \tag{35}$$

We do not apply Equation (35) to adjust individual correlations, since the adjustment may violate positive semi-definiteness of the resulting covariance matrix. Next, we note that the variance of the portfolio with weights $w$ equals $\sum_{i,j}^{N} w_i w_j \hat{\sigma}_i \hat{\sigma}_j \hat{\rho}_{ij}$. Applying approximation (35), the covariance

[29]The deviation stems from the return dependence on past non-trades, when the return contains information content of multiple periods. However, simulation exercise (available upon request) shows that the approximation does not exceed 5% in long sample.

contribution to portfolio variance may be expressed as

$$\sum_{i \neq j}^{N} w_i w_j \hat{\sigma}_i \hat{\sigma}_j \hat{\rho}_{ij} \approx \sum_{i \neq j}^{N} w_i w_j \hat{\sigma}_i \hat{\sigma}_j \rho_{ij} B_{ij}, \tag{36}$$

where $w$ and $\hat{\sigma}$ denote the weights and volatility estimates of individual assets, $\rho$ stands for pairwise correlation, and $N$ is the number of stocks in the portfolio. The contribution of covariance terms can be further expressed as

$$\frac{\sum_{i \neq j}^{N} w_i w_j \hat{\sigma}_i \hat{\sigma}_j \hat{\rho}_{ij}}{\sum_{i \neq j}^{N} w_i w_j \hat{\sigma}_i \hat{\sigma}_j} \approx \frac{\sum_{i \neq j}^{N} w_i w_j \hat{\sigma}_i \hat{\sigma}_j B_{ij}}{\sum_{i \neq j}^{N} w_i w_j \hat{\sigma}_i \hat{\sigma}_j} \frac{\sum_{i \neq j}^{N} w_i w_j \hat{\sigma}_i \hat{\sigma}_j B_{ij} \rho_{ij}}{\sum_{i \neq j}^{N} w_i w_j \hat{\sigma}_i \hat{\sigma}_j B_{ij}} \tag{37}$$

$$\sum_{i \neq j}^{N} \tilde{w}_{ij} \hat{\rho}_{ij} \approx \bar{B} \sum_{i \neq j}^{N} \tilde{\tilde{w}}_{ij} \rho_{ij} \tag{38}$$

$$\bar{B} := \frac{\sum_{i \neq j}^{N} w_i w_j \hat{\sigma}_i \hat{\sigma}_j B_{ij}}{\sum_{i \neq j}^{N} w_i w_j \hat{\sigma}_i \hat{\sigma}_j}. \tag{39}$$

Finally, by neglecting the cross-sectional covariation between correlations and the difference in two sets of weights, we may approximate the weighted correlation to

$$\sum_{i \neq j}^{N} \tilde{w}_{ij} \hat{\rho}_{ij} = \bar{B} \left( \sum_{i \neq j}^{N} \tilde{w}_{ij} \rho_{ij} + \sum_{i \neq j}^{N} \left( \tilde{\tilde{w}}_{ij} - \tilde{w}_{ij} \right) \rho_{ij} \right) \approx \bar{B} \sum_{i \neq j}^{N} \tilde{w}_{ij} \rho_{ij}. \tag{40}$$

Equal-weighted correlation, presented in Figure 5, corresponds to the case when weights are constant across assets and returns are standardized ($\hat{\sigma} = 1$). In that case the bias-correction term $\bar{B}$ (denoted $\bar{B}_{EW}$) boils down to the average of $B_{ij}$.

Analogously, we define $\bar{B}_{VW}$ as a $w_i w_j$-weighted average of individual $B_{ij}$s. To measure the effect of the bias on industry portfolios, hence on the value-weighted IV, it would be preferable to weight also by stocks' volatilities, and to consider intra-industry correlations only. Empirically, the bias-correction factors are almost identical to $\bar{B}_{VW}$. Figure 6 shows that the percentage bias in average correlations, approximated by $(1 - \bar{B}_{EW})$ goes up to 75%. In Figure 5 we include a series of average correlations of daily returns, scaled by the correction factor ($\bar{B}_{EW}$). Reassuringly, the corrected series is very close to values obtained with monthly data, for which asynchronicity bias is likely limited. The series $(1 - \bar{B}_{VW})$ indicates that (VW) industry covariance is biased by approximately

30%. Because the variance of a diversified portfolio is dominated by average covariance, similar bias can be expected for portfolio variances, which enter in computation of IV, industry-layer and market variance.

[Figure 6 about here.]

In what follows we study the asynchronicity effects on CE correlations (see, Def. 5), which reflects more accurately the impact of correlations on the value-weighted IV, by taking into account size of the firms, as well as independence of inter-industry correlations. First possible solution to asynchronicity is to exclude most affected stocks from analysis. In the CLMX period, the CE correlations increase monotonically with the number of nonzero return required for inclusion. However, requirement of having no more than 25% daily returns are non-trades in given month, result in loss of 36% month-firm observations in pooled sample, which is further heavily concentrated in the CLMX period.

To avoid data loss stemming from exclusion of firms with frequent non-trades, a preferable solution is to use an estimator robust to asynchronicity (or correct for the bias), such as those proposed by Hayashi and Yoshida (2005), Zhang (2011), Aït-Sahalia et al. (2010). A challenge is that in our setting (up to 7,460 firms) these methods are computationally burdensome, and typically do not ensure positive semi-definiteness of the resulting covariance matrix. Moreover, the percentage of non-trade days after the inclusion of NASDAQ is close to 50%, which makes many of the methods inapplicable. Figure 5 suggests that switching to a lower frequency, which we explore next, is a simple solution that eliminates the bias, and retains positive semi-definiteness.

We estimate the CE correlations on a rolling of one year, using daily, weekly, and monthly returns. By using identical window for all frequencies, we ensure that the difference in measured correlations is not driven by differences in sample of used firms. Figure 7 shows the results. The correlations in the post-CLMX sample, when the non-trade observation share is low, depend only weakly on the applied sampling. Conversely, in the CLMX sample, the differences are substantial. The series based on monthly returns seems to have a roughly constant long-term mean over the sample except for the period of low correlation around dot-com bubble.

[Figure 7 about here.]

Next, we compute the idiosyncratic volatility based on 24 monthly returns, so that the observation count roughly matches the number of observations when one month of daily returns. Panel A of Figure 8 shows the final output.[30] The change of sampling frequency has a sizable impact on measured IV. One obvious concern is that the requirement of complete two-year history for the IV based on monthly returns might eliminate nontrivial portion of stocks and thus drive the difference. To mitigate this concern, we also compute additional series of IV based on two years of daily data. Over majority of the sample, two IV measures based on daily data are similar. On the contrary, both series based on two years of data differ by larger magnitude, which supports the assertion that asynchronicity is the driving force behind the difference of the original series and the one based on two years of monthly returns.

Next, we ask how the choice of sampling frequency affects the relative importance of individual volatility components. Panels B in Figure 8 shows the percentage of total variance (i.e., sum of firm-layer, industry-layer, and market variance, which equals average firm variance) each component of the original series represents. Panel C presents the results of the decomposition based on two years of monthly data. The daily-data decomposition shows a strong structural break after the CLMX sample due to the correlation bias. Somewhat surprisingly, the monthly-data decomposition not only eliminates the IV drop but even turns it into a mild increase. A partial explanation of recent increase observed with monthly returns is that the industries became less concentrated after 1980, making the industry and market portfolios less volatile.

[Figure 8 about here.]

A possible concern with the stabilization of the value-weighted idiosyncratic volatility, when measured on a more reliable monthly frequency, is that the result might be driven by few large companies, while the frequency change has little impact on the small ones. To address this concern, we construct the CLMX IV measure with equal weights in place of the weights based on market capitalization. We apply the equal weighting both for construction of industry (market) returns, and

[30] For comparability, the IV estimates corresponding to one-month and two-year period have to be converted to equal footing. We translate the two-year IV ($\hat{\sigma}^2_m$) to monthly basis measure $\hat{\sigma}^2_{ma}$ via $\hat{\sigma}^2_{ma} = (1 + \hat{\sigma}^2_m)^{1/24} - 1$, which is based on variance calculation with different frequencies under the assumption of iid returns with zero mean within given period.

for weighting of the IVs of individual firms (industries).[31] The resulting equal-weighted measures, as well as the contribution of its components to total variance, are displayed in Figure 9.

[Figure 9 about here.]

A first notable feature of the equal-weighted IV is that the difference between the daily and the monthly sampling frequency is more pronounced, but a significant portion of the change stems from the exclusion of stocks without a complete two-year history. This difference is partially driven by distressed stocks with high variance (see, section 4.1), part of which has a lifetime shorter than two years and thus does not appear in the sample. For stocks with such a short lifetime, the measurement of exposure to systemic factors as well as disentangling firm-specific and age-specific effects is difficult. Among the stocks with a complete history, the impact of sampling on IV is still significant. A second feature that stands out is that the EW IV based on monthly data still has clearly visible upward "trend" in the CLMX sample. Third, as in the value-weighted case, the IV represents a more significant portion of the total variance in the early part of the sample, not in the latter.

To elaborate on the potential impact of industry concentration on the share of IV on total variance, we first note that by Equation (17) the IV share satisfies

$$\mathrm{IVS}_T := \frac{\hat{\sigma}^2_{1:2,T}}{\hat{\sigma}^2_{0,T} + \hat{\sigma}^2_{0:1,T} + \hat{\sigma}^2_{1:2,T}} = 1 - \frac{\bar{\hat{\sigma}}^2_{1,T}}{\bar{\hat{\sigma}}^2_{2,T}}, \tag{41}$$

with $\hat{\sigma}^2_{0,T}$, $\bar{\hat{\sigma}}^2_{1,T}$, and $\bar{\hat{\sigma}}^2_{2,T}$ denoting the market, average industry, and average firm variance. In case the returns were uncorrelated, we could write

$$1 - \mathrm{IVS}_T = \frac{\sum_{s=1}^{S} w_s \sum_{i \in \mathcal{S}_s} \left( \frac{w_i}{\sum_{i \in \mathcal{S}_s} w_i} \right)^2 \hat{\sigma}_i^2}{\sum_{i=1}^{F} w_i \hat{\sigma}_i^2} = \bar{\mathcal{H}} \frac{\sum_{i=1}^{F} \tilde{w}_i \hat{\sigma}_i^2}{\sum_{i=1}^{F} w_i \hat{\sigma}_i^2}, \tag{42}$$

where $\bar{\mathcal{H}}$ denotes the average industry Herfindahl-Hirschman index (see, Equation (34)), and $\tilde{w}$ are the normalized weights on the individual firms' variances in the average industry variance computation.

Motivated by Equation (42), we regress the logarithm of $1 - \mathrm{IVS}_T$ on the logarithm of $\bar{\mathcal{H}}_T$. The regressions yield an R-squared of 0.13 and 0.39 for value-weighted and equal-weighted IV share,

[31] This is equivalent to replacing $W$ in the definition of the layer variance (Equation (16)) by $I_F \frac{1}{F}$, with $F$ denoting the number of firms.

respectively. While for VW-IV the explanatory power of industry concentration is limited, for the EW IV the fit is significantly better. First possible explanation of the better fit is that the correlations in the EW case are closer to zero, so the deviation from Equation 42 will be smaller. A second possible explanation is that there is an increase in variances of small stocks, which have a higher weight ($w$) in the average variance than in the value-weighted portfolio ($\tilde{w}$). This discrepancy essentially disappears with equal weights.

Summarizing the above discussion, we demonstrated that the decrease in IV and its relative importance in total variance after CLMX sample is mostly an artifact of the bias in correlations stemming from asynchronicity among illiquid stocks. A trivial solution we propose is to switch to a lower frequency. With monthly data, the drop in the IV mostly disappears. The share of IV to total variance even increases in the most recent period, contrary to what we observe in IV based on daily data. Changes in the industry concentration can to some extent explain the increase in IV share on total variance.

### 4.4 Average variances

Section 4.3 highlights several essential features of the IV. First, the drop in the IV, as well as its relative importance, observed after 1997 is to a nontrivial extent driven by bias in correlations, stemming from illiquidity of a subset of stocks. For the value-weighted IV, a change in sampling frequency to monthly eliminates the structural break a weakens the measured "trend" in the CLMX period. For equal-weighted measure, the impact is similar but overshadowed by the effect of discarding firms with an incomplete two-year history. Still, a part of the trend in the IV in the CLMX period remains, which leads us to investigate the evolution of average variance, the last channel driving IV identified in Example 3.3.

Figure 10 displays the evolution of average variance, both equal-weighted (analyzed by Xu and Malkiel (2003)) and value-weighted. Both series evolve similarly, with the EW variance dominating the VW measure due to higher variances of small stocks. The figure indicates a presence of a trend in the CLMX period. To corroborate this finding, we follow CLMX and test for the presence of a deterministic trend in the CLMX sample using the $\mathrm{PS}^1 - t$ test of Vogelsang (1998), which is robust both to stationary and unit root error term. Table 1 presents the results for logarithm of average variances, whose increments are closer in distribution to Gaussian than those of raw series, thus the critical values (derived from Gaussian assumption) are more accurate.

[Figure 10 about here.]

[Table 1 about here.]

On daily data, the hypothesis of no trend is rejected, confirming to the conclusion of previous studies. Surprisingly, we cannot reject the hypothesis of no trend on a 95% level of confidence in the case of VW average variance based on the monthly data. However, the inability to reject the hypothesis stems from a low power of the test due to limited sample size (445 observations) and strong persistence in the error term, amplified by the overlapping window used in the IV construction, which inflates the variance of the estimator. Moreover, it seems unlikely that there was a trend in the EW average, but not in the VW measure.

If the trend in the IV in the CLMX period stems from an increase in the average variance, as argued above, is it possible that the industry and the market components have no trend? CLMX conclude that these components "have a small positive but insignificant trend coefficient whereas the trend in FIRM is much larger." After the CLMX publication numerous studies[32] emerged trying to explain the trend in the IV prior to 1997. We argue that industry and market components trended in that period as well, and the inability to reject the absence of trend stems from a low power of the test.

First, if the variance of all stocks increased in the same proportion, then, keeping other things fixed, the industry and market component would increase proportionally as well. Of course, we can expect the industry and market component are growing less than one-to-one in response to an increase in the average variance, as they put relatively lower weight on small stocks, the primary driver of the trend (see, Bali et al. (2008)). However, the concentration of the trend among small stocks cannot eliminate the trend unless it is offset by a decrease in variances of large stocks, for

[32]See, e.g., Xu and Malkiel (2003), Comin and Philippon (2005), Bartram et al. (2018)

which we find no support in the data. Indeed, the variances of large stocks have a stable mean in the long run (not presented). Second, if there was a trend in the IV, but not in the industry and market components, then the share of the IV on total volatility would increase over time as well. For value-weighted IV (Figure 8) no such trend in the CLMX period shows up, regardless of the applied sampling frequency. Similar conclusions hold in a formal estimation: Vogelsang's test cannot reject the hypothesis of no trend in the IV share, though admittedly, the underlying normality assumption of the test is violated even more than in the case of variance series. When weighted equally, the IV share trends upwards. The EW measure is less affected by a disproportionate change in the variance of small and large firms, so the trend more likely originates from a change in correlation or market concentration. We have already shown the relevance of the latter channel in Section 4.3.

In this section, we have shown that average variances, one of the drivers of the IV, trended over the CLMX sample. The only series for which we cannot reject the hypothesis of no trend is the value-weighted series based on monthly data, likely due to a lack of power of the test with sample size at hand. Next, we argue that if the IV increased due to an increase in variances, industry and market components should increase as well, contrary to the conclusion of CLMX. The analysis of the share of IV in total variance supports this line of argument. We show that the share was, in value-weighted case, remarkably stable, up to the structural breaks stemming from the correlation measurement bias (see, Section 4.3) or the industry concentration effects, depending on the applied sampling frequency. In the equal-weighted case, the IV share is trending, but the trend is well captured by changing industry concentration. We conclude that period before 1997 is not characterized by a trend in the idiosyncratic risk alone, but by an increase in the overall risk. As a consequence, the focus of related studies on idiosyncratic component seems unnecessary, and also requires controlling for changing market concentration and proper measurement of correlations.

## 5 Idiosyncratic volatility and expected returns

In this section, we investigate the effect of sampling frequency and delisting effects on the relation between IV and expected returns. The first indication of the viability of these channels is the fact that the underperformance of high-IV portfolio is most severe in the 1980s and 1990s, where these effects are most pronounced. Han and Lesmond (2011) show that the pricing ability of IV mostly disappears once we control for microstructure effects. On the contrary, Chen et al. (2012) conclude that the IV anomaly is robust to the exclusion of the stocks prone to the microstructure effects, or even becomes stronger, which leads us to further investigation. As a baseline, we consider the IV measure of AHXZ, computed as the standard deviation of residuals from Fama-French 3-factor model[33], estimated on a month of daily data. As alternatives, we compute the IV from the last one and two years of monthly data, respectively. For better comparability to their results, we restrict our analysis to the AHXZ sample from July 1963 to December 2000.

First, we ask to what extent does the choice of the IV measure affect to which IV-quintile portfolio the stock belongs. Table 2 shows the migration matrices of stocks between quintile portfolios based on different measures of IV. Despite the significant effect of sampling frequency on correlation estimates demonstrated in Section 4.3, the impact on ranking, thus on portfolio formation, is modest. Still, the effect on expected returns can be material, as the return differential between the adjacent quintiles is significant, especially in the high-IV region. The secondary effect of switching to a different lookback period is the exclusion of a substantial portion of observations, more so for the high-IV quintile.

[Table 2 about here.]

The second aspect is the extent to which the post-listing and pre-delisting effects drive the portfolio performance. We start by extending the delisting analysis of Section 4.1 for expected returns. Figure 11 shows the results. As in the case of variances, there is no support for post-listing effects, and evidence in favor of pre-delisting effects, which are again specific to delisting category. A low variance of the merger category is accompanied by abnormally high returns. The converse holds for the “Bankruptcy” category. However, the effect on means is delayed (hence shorter) compared to the variance effects.

[Figure 11 about here.]

Table 3 shows how many firm-month observations, which are less than one year before delisting in each category, are contained in each IV portfolio, and their average returns in the pooled sample. The

[33] The FF factors are obtained from WRDS.

results are in line with the intuition that more distressed stocks are allocated in the high-IV portfolio. Presence of mergers among portfolios is either decreasing with the IV or U-shaped, depending on the applied measure and the required number of observations. The returns on the distressed stocks are abysmally low, regardless of the IV portfolio. The opposite holds for the merger category.

The results provide a rough confirmation of the negative IV-expected return relation in the baseline case. Once we move towards the IV measures based on monthly data and longer lookback periods (hence exclude new stocks), the negative relation disappears or even reverts. Exclusion of pre-delisting period has a similar effect: the negative relationship disappears and is positive when both exclusions are applied simultaneously. The breakdown suggests that the stocks (one year) after listing and before delisting are crucial for the negative relation. The lack of negative effect of post-listing on the mean (see, Figure 11) likely stems from the difficulty to disentangle the age-related impact from firm-specific and time-specific in some cases.

[Table 3 about here.]

One caveat of Table 3 is that the pooled sample overweights the periods with a high number of shares. We repeat the analysis, computing the monthly average returns in each category and portfolio, and then averaging over time. This exercise yields similar results (not presented), so the overweighting of periods with many firms does not drive the results. The second difference compared to the portfolio analysis is that the stocks are equal-weighted. To investigate the direct impact of the exclusions, we construct the IV portfolios on samples where we exclude the last years of “bankruptcy” stocks and merger and “bankruptcy” category at the same time, for all IV-measures under consideration.

Table 4 summarizes the results. The difference between the IV measure based on one month of data and the remaining two measures, based on one and two years of data, respectively, confirms the central role of young firms. The impact of young firms is in line with a low expected return and high volatility found by Pástor and Veronesi (2003). Further exclusion of the pre-delisting stocks eliminates the residual negative IV-relation. The positive relation found in the equal-weighted analysis (Table 3) indicates that high-IV portfolio contains many micro-caps which have high returns on average but have limited impact on the value-weighted measure. Bali and Cakici (2008) also stress the importance of the choice of the weighting scheme, and conclude that replacing value-weighting with equal-weighting, or by weights proportional to the inverse IV has a significant effect on the resulting IV-expected return relation.

[Table 4 about here.]

In Table 5 we show monthly-return alphas of the IV portfolios for different IV measures and exclusions of the pre-delisting period. Similarly to expected returns, alphas are smaller once we exclude young stocks (by using a more extended lookback period), or distressed stocks before delisting. Contrary to the expected returns, alphas retain the ordering and their economic significance, though they are no longer statistically significant. The results are robust to the inclusion of Fama-French factors (Panel B) and momentum factor of Jegadeesh and Titman (1993) (Panel C). The systemic difference between the patterns of expected returns and CAPM alphas stems from the fact that betas are monotonically increasing along the IV dimension. This observation is consistent with betting against beta (BAB) factor of Frazzini and Pedersen (2014):[34] Adding BAB to the factor model specification results in economically small alphas (Panel D).

[Table 5 about here.]

An obvious concern is that excluding stocks based on their future delisting events suffers from lookahead and survivorship biases. First, this objection does not affect the post-listing period, which we have shown to be more critical. Second, the purpose of our analysis is not to formulate a valid investment strategy, but rather to identify the drivers of underperformance of the high-IV portfolio. Our analysis shows that the primary source of return differentials in the IV-based portfolios are newly listed firms and, to a lesser extent, firms affected by pre-delisting effects. A stronger effect of newly listed firms is surprising in light of findings of Avramov et al. (2013), who show that returns of the IV strategy derive from firms with deteriorating credit conditions, especially around credit rating downgrades of low-rated firms. Even though these firms are most likely to delist, our results suggest

[34] We obtain the BAB factor from AQR website.

that delisting is insufficient to drive the IV puzzle. One possible reason for the different conclusion stems from a different sample composition due to the requirement of availability of credit rating. To what extent this hypothesis holds remains an open research question.

## 6 Discussion

In Sections 4.3 and 4.4 we show that before 1997 there was a trend in the idiosyncratic risk and, in contrast to CLMX, also in the systemic component. As a consequence, the relative importance of value-weighted idiosyncratic risk remained almost unchanged. As argued by CLMX, for an investor, who wanted to reduce idiosyncratic risk to a fixed level, the trend means that more stocks were needed over time. On the contrary, an investor, whose goal was to reduce the idiosyncratic risk to a given percentage of the market risk, could have achieved the goal with an almost constant number of shares in the portfolio. Section 4.3 further shows that in the period before 1997, the idiosyncratic volatility, as well as its contribution to total variance, are overestimated when using daily data. The overestimation stems from the illiquidity-driven bias in correlations. Daily data also imply that the systemic risk increased its relevance after 1997. The increase in relevance is in stark contrast to what is observed with more reliable monthly returns: The idiosyncratic risk is more important after 1997, partially due to the observed change in market concentration.

In light of our sampling analysis, we recommend avoiding using daily returns before 1997, whenever estimates of the covariance structure are needed, and when small stocks, especially those listed on NASDAQ, are considered. One possibility is to use the sample after 1997 only, thus avoiding the illiquidity issue, or to use a lower sampling frequency. This comes at the cost of reduced sample size. Unfortunately, most of the asset pricing applications revolve around estimation of means (alphas), which require long samples to be done accurately. An alternative is to reformulate the problem to a mixed frequency, if applicable. A sub-sample analysis (sample before vs. after 1997), and testing different sampling frequencies are useful points of departure for assessing the impact of asynchronicity in empirical studies.

The sampling effect is relevant for studies of the relation between idiosyncratic volatility and expected returns. Underestimation of correlations results in false attribution of a part of the systemic risk to the idiosyncratic component. Herskovic et al. (2016) use the average idiosyncratic volatility as a pricing factor. When they compute it from daily returns, the series of averages of total and idiosyncratic volatility essentially coincide which is no longer valid for monthly returns. Underestimation of correlations also explains the robustness of Ang et al. (2006) results to changing their IV measure with total volatility, documented by Blitz and Van Vliet (2007). We have shown that despite the sizable impact of sampling frequency on the measured magnitude of IV, the effect on the IV quintile to which the stock belongs is modest. In turn, the correlation bias is unable to eliminate a significant portion of mispricing of the IV portfolios - an opposite conclusion to that of Han and Lesmond (2011). The estimation of the IV based on a more extended lookback period has a significant secondary effect of dropping young firms with an insufficient history. The result is similar as in Bali and Cakici (2008), and Fu (2009), who find a flat or even positive IV-expected return relation in specific configurations, often involving truncation of young stocks.

Second aspect we highlight is the role of corporate events such as mergers on the return distribution. As documented in Section 4.1, cash-settled mergers essentially convert the stock to a bond, thus reduce volatility after the merger announcement. Stock-settled mergers shift the covariance properties to those of the stock of the acquiring company. Overall, these effects as well as high sensitivity of these stocks to bond returns found by Beilo et al. (2017), are consistent with the low-IV portfolio. On the contrary, distressed stock prior to delisting share characteristics with the high-IV portfolio. We show that the high-IV portfolio indeed loads heavily on distressed stocks, which have abysmally low returns. The presence of mergers in the low-IV portfolio is sensitive to the choice of IV measure. Both of these events have a significant effect on expected returns which raises a question to what extent are well-known pricing anomalies driven by these events, and how these stocks are priced.

# 7 Conclusion

In this study we revisit the question of the existence of a transitory increase in idiosyncratic volatility (IV), perceived as a trend in earlier studies (see, e.g., Campbell et al. (2001), Xu and Malkiel (2003) and Comin and Philippon (2005)), and as a transitory increase more recently (see, Bekaert et al. (2012)), and discuss its primary drivers. The importance of these questions stems from the role of idiosyncratic volatility in investment decisions of agents who are unable to diversify fully. In light of our findings on the drivers of the historical IV evolution, we also review the link between the IV and expected returns, studied both theoretically (see, Merton (1987)) and empirically (see, e.g., Ang et al. (2006, 2009), Fu (2009), Stambaugh et al. (2015)).

Computing the value-weighted IV along the lines of Campbell et al. (2001), we first show that the trend breaks down after 1997, and the IV reverts to a lower level, confirming the results of Bekaert et al. (2012). Second, by isolating the effect of newly listed firms, we show that the shift in the IV originates from newly listed firms, confirming the conclusion of Brown and Kapadia (2007). Because the period of high IV corresponds to a period of strong growth in the number of listed firms, we ask whether the impact is related to the specifics of the covariance structure around the events of listing and delisting, and conclude that this is not the case. Third, we show that it is inappropriate to use daily data in the context of IV estimation before 1997. The reason is that illiquidity influences such correlation estimates and biased toward zero. Switching to monthly frequency results in higher correlations, and a lower IV before 1997. Another consequence of the switch is that the IV became relatively more important in recent decades - an opposite conclusion to what we find using daily data. Further, we show, that the pre-1997 period was characterized by an increase in the average variance, which has materialized in a trend not only in the IV but also in the industry and market components of variance. A decreasing market concentration partially attenuated the trend in the latter two due to diversification effect. This motivates to study the fundamental drivers of two distinct phenomena, evolution of average variance, and market concentration, rather than the evolution of idiosyncratic volatility itself, for which a tremendous effort was made (Xu and Malkiel (2003), Comin and Philippon (2005), Brown and Kapadia (2007), Bali et al. (2008), Bekaert et al. (2012)).

Finally, we analyze the role of sampling, post-listing, and pre-delisting effects on the performance of IV portfolios. Despite the significant role of the sampling frequency in correlation measurement, the change of frequency has a modest impact on portfolio classification. We show that the abysmally low returns of high-IV returns are strongly driven by newly listed and distressed stocks. Without these stocks, the IV-expected return relation is mostly flat or even upward sloping, depending on the applied weighting. Pricing of distressed stocks remains an open area for future research, as their characteristics are difficult to reconcile with standard asset pricing models.

## References


Aït-Sahalia, Y., Fan, J. and Xiu, D.: 2010, High-frequency covariance estimates with noisy and asynchronous financial data, *Journal of the American Statistical Association* **105**(492), 1504–1517.

Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X.: 2006, The cross-section of volatility and expected returns, *The Journal of Finance* **61**(1), 259–299.

Ang, A., Hodrick, R. J., Xing, Y. and Zhang, X.: 2009, High idiosyncratic volatility and low returns: International and further u.s. evidence, *Journal of Financial Economics* **91**(1), 1–23.

Avramov, D., Chordia, T., Jostova, G. and Philipov, A.: 2013, Anomalies and financial distress, *Journal of Financial Economics* **108**(1), 139–159.

Bali, T. G. and Cakici, N.: 2008, Idiosyncratic volatility and the cross section of expected returns, *Journal of Financial and Quantitative Analysis* **43**(1), 29–58.

Bali, T. G., Cakici, N. and Levy, H.: 2008, A model-independent measure of aggregate idiosyncratic risk, *Journal of Empirical Finance* **15**(5), 878–896.

Bartram, S. M., Brown, G. W. and Stulz, R. M.: 2018, Why has idiosyncratic risk been historically low in recent years?, *Working Paper 24270*, National Bureau of Economic Research.

Beilo, R., Driessen, J. and Kuiper, I.: 2017, Does interest rate exposure explain the low-volatility anomaly?

Bekaert, G., Hodrick, R. J. and Zhang, X.: 2012, Aggregate idiosyncratic volatility, *Journal of Financial and Quantitative Analysis* **47**(6), 1155–1185.

Bessembinder, H.: 2000, Tick size, spreads, and liquidity: An analysis of nasdaq securities trading near ten dollars, *Journal of Financial Intermediation* **9**(3), 213–239.

Blitz, D. C. and Van Vliet, P.: 2007, The volatility effect: Lower risk without lower return, *Journal of Portfolio Management* pp. 102–113.

Bollen, N. P., Smith, T. and Whaley, R. E.: 2004, Modeling the bid/ask spread: measuring the inventory-holding premium, *Journal of Financial Economics* **72**(1), 97–141.

Brandt, M. W., Brav, A., Graham, J. R. and Kumar, A.: 2010, The idiosyncratic volatility puzzle: Time trend or speculative episodes?, *Review of Financial Studies* **23**(2), 863–899.

Branger, N., Hülsbusch, H. and Middelhoff, T. F.: 2017, Idiosyncratic volatility, its expected variation, and the cross-section of stock returns.

Brown, G. and Kapadia, N.: 2007, Firm-specific risk and equity market development, *Journal of Financial Economics* **84**(2), 358–388.

Campbell, J. Y., Lettau, M., Malkiel, B. G. and Xu, Y.: 2001, Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk, *The Journal of Finance* **56**(1), 1–43.

Campbell, J. Y., Lo, A. W.-C. and MacKinlay, A. C.: 1997, *The econometrics of financial markets*, Princeton University Press.

Chen, L. H., Jiang, G. J., Xu, D. and Yao, T.: 2012, Dissecting the idiosyncratic volatility anomaly.

Chen, Z. and Petkova, R.: 2012, Does idiosyncratic volatility proxy for risk exposure?, *The Review of Financial Studies* **25**(9), 2745–2787.

Christie, W. G. and Schultz, P. H.: 1994, Why do nasdaq market makers avoid odd-eighth quotes?, *The Journal of Finance* **49**(5), 1813–1840.

Comin, D. and Philippon, T.: 2005, The rise in firm-level volatility: Causes and consequences, *NBER Macroeconomics Annual* **20**, 167–201.

Doidge, C., Karolyi, G. A. and Stulz, R. M.: 2017, The u.s. listing gap, *Journal of Financial Economics* **123**(3), 464–487.

Epps, T. W.: 1979, Comovements in stock prices in the very short run, *Journal of the American Statistical Association* **74**(366), 291–298.

Fama, E. F. and French, K. R.: 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* **33**(1), 3–56.

Frazzini, A. and Pedersen, L. H.: 2014, Betting against beta, *Journal of Financial Economics* **111**(1), 1–25.

Fu, F.: 2009, Idiosyncratic risk and the cross-section of expected stock returns, *Journal of Financial Economics* **91**(1), 24–37.

Han, Y. and Lesmond, D.: 2011, Liquidity biases and the pricing of cross-sectional idiosyncratic volatility, *The Review of Financial Studies* **24**(5), 1590–1629.

Hasbrouck, J.: 1999, The dynamics of discrete bid and ask quotes, *The Journal of Finance* **54**(6), 2109–2142.

Hayashi, T. and Yoshida, N.: 2005, On covariance estimation of non-synchronously observed diffusion processes, *Bernoulli* **11**(2), 359–379.

Herskovic, B., Kelly, B., Lustig, H. and Van Nieuwerburgh, S.: 2016, The common factor in idiosyncratic volatility: Quantitative asset pricing implications, *Journal of Financial Economics* **119**(2), 249–283.

Hoberg, G. and Phillips, G.: 2010, Product market synergies and competition in mergers and acquisitions: A text-based analysis, *The Review of Financial Studies* **23**(10), 3773–3811.

Hoberg, G. and Phillips, G.: 2017, Text-based industry momentum, *Journal of Financial and Quantitative Analysis* **forthcoming**.

Hou, K. and Loh, R. K.: 2016, Have we solved the idiosyncratic volatility puzzle?, *Journal of Financial Economics* **121**(1), 167–194.

Jegadeesh, N. and Titman, S.: 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of Finance* **48**(1), 65–91.

Lowry, M., Officer, M. S. and Schwert, G. W.: 2010, The variability of IPO initial returns, *The Journal of Finance* **65**(2), 425–465.

Merton, R. C.: 1987, A simple model of capital market equilibrium with incomplete information, *The Journal of Finance* **42**(3), 483–510.

Pástor, Ĺ. and Veronesi, P.: 2003, Stock valuation and learning about profitability, *The Journal of Finance* **58**(5), 1749–1789.

Stambaugh, R. F., Yu, J. and Yuan, Y.: 2015, Arbitrage asymmetry and the idiosyncratic volatility puzzle, *The Journal of Finance* **70**(5), 1903–1948.

Vogelsang, T.: 1998, Trend function hypothesis testing in the presence of serial correlation, *Econometrica* **66**(1), 123–148.

Xu, Y. and Malkiel, B. G.: 2003, Investigating the behavior of idiosyncratic volatility, *The Journal of Business* **76**(4), 613–645.

Zhang, L.: 2011, Estimating covariation: Epps effect, microstructure noise, *Journal of Econometrics* **160**(1), 33–47.

Zhang, M. Y., Russell, J. R. and Tsay, R. S.: 2008, Determinants of bid and ask quotes and implications for the cost of trading, *Journal of Empirical Finance* **15**(4), 656–678.

# A Proof of Proposition 1

First, we consider a single $t \in T$ and suppress the time subscripts for brevity. Observe that for any diagonal matrix $D \in \mathbb{R}^{F \times F}$ and classification matrix $C \in \mathbb{R}^{S \times F}$ matrix $B = CDC^\top$ has a typical element $b_{uv} = \sum_{x=1}^{F} C_{ux} D_{xx} C_{vx}$. As classification matrices have exactly one nonzero element (equal to one) in each column, $B$ is also a diagonal with elements equal to sums of $D$ entries in each class. Using the commutativity of a product of diagonal matrices and the invariance of the matrix trace under a cyclic permutation, a direct calculation gives

$$\begin{aligned}
\epsilon_l^\top (C_l V C_l^\top) \epsilon_l = &\left( C_l V R C_l^\top (C_l V C_l^\top)^{-1} e_{S_l} \right)^\top (C_l V C_l^\top) \left( C_l V R C_l^\top (C_l V C_l^\top)^{-1} e_{S_l} \right) \\
&+ \left( C_{l-1,l}^\top \left( C_{l-1} V R C_{l-1}^\top (C_{l-1} V C_{l-1}^\top)^{-1} e_{S_{l-1}} \right) \right)^\top (C_l V C_l^\top) \\
&\times C_{l-1,l}^\top \left( C_{l-1} V R C_{l-1}^\top (C_{l-1} V C_{l-1}^\top)^{-1} e_{S_{l-1}} \right) \\
&- 2 \left( C_l V R C_l^\top (C_l V C_l^\top)^{-1} e_{S_l} \right)^\top (C_l V C_l^\top) C_{l-1,l}^\top \left( C_{l-1} V R C_{l-1}^\top (C_{l-1} V C_{l-1}^\top)^{-1} e_{S_{l-1}} \right) \\
=: &S_1 + S_2 - 2S_3.
\end{aligned} \tag{A.1}$$

Computing the first term:

$$\begin{aligned}
S_1 &= e_{S_l}^\top (C_l V C_l^\top)^{-1} \left( C_l V R C_l^\top \right) \left( C_l V C_l^\top \right) \left( C_l V R C_l^\top \right) (C_l V C_l^\top)^{-1} e_{S_l} \\
&= e_{S_l}^\top \left( C_l V R C_l^\top \right) (C_l V C_l^\top)^{-1} \left( C_l V R C_l^\top \right) e_{S_l} \\
&= e_F^\top V R C_l^\top (C_l V C_l^\top)^{-1} C_l V R e_F \\
&= \operatorname{tr} \left[ V (r r^\top) V C_l^\top (C_l V C_l^\top)^{-1} C_l \right].
\end{aligned} \tag{A.2}$$

Similarly for the second term.

$$\begin{aligned}
S_2 &= \left( C_{l-1,l}^\top \left( C_{l-1} V R C_{l-1}^\top (C_{l-1} V C_{l-1}^\top)^{-1} e_{S_{l-1}} \right) \right)^\top (C_l V C_l^\top) C_{l-1,l}^\top \left( C_{l-1} V R C_{l-1}^\top (C_{l-1} V C_{l-1}^\top)^{-1} e_{S_{l-1}} \right) \\
&= e_{S_{l-1}}^\top C_{l-1} V R C_{l-1}^\top (C_{l-1} V C_{l-1}^\top)^{-1} C_{l-1,l} (C_l V C_l^\top) C_{l-1,l}^\top C_{l-1} V R C_{l-1}^\top (C_{l-1} V C_{l-1}^\top)^{-1} e_{S_{l-1}} \\
&= e_F^\top R V C_{l-1}^\top (C_{l-1} V C_{l-1}^\top)^{-1} (C_{l-1} V C_{l-1}^\top) (C_{l-1} V C_{l-1}^\top)^{-1} C_{l-1} V R e_F \\
&= \operatorname{tr} \left[ V (r r^\top) V C_{l-1}^\top (C_{l-1} V C_{l-1}^\top)^{-1} C_{l-1} \right].
\end{aligned} \tag{A.3}$$

And the cross-term

$$
\begin{aligned}
S_3 &= e_{S_l}^\top (C_l V C_l^\top)^{-1} \left(C_l V R C_l^\top\right) (C_l V C_l^\top) C_{l-1,l}^\top \left(C_{l-1} V R C_{l-1}^\top (C_{l-1} V C_{l-1}^\top)^{-1} e_{S_{l-1}}\right) \\
&= e_{S_l}^\top \left(C_l V R C_l^\top\right) C_{l-1,l}^\top \left(C_{l-1} V R C_{l-1}^\top (C_{l-1} V C_{l-1}^\top)^{-1} e_{S_{l-1}}\right) \\
&= e_F^\top R V C_l^\top C_{l-1,l}^\top (C_{l-1} V C_{l-1}^\top)^{-1} C_{l-1} V R e_F \\
&= \operatorname{tr}\left[V(rr^\top) V C_{l-1}^\top (C_{l-1} V C_{l-1}^\top)^{-1} C_{l-1}\right] = S_2.
\end{aligned}
\tag{A.4}
$$

From the above it follows that

$$
\begin{aligned}
\hat{\sigma}_{l,T}^2 &= \sum_{t \in T} \operatorname{tr}\left[V_t (r_t r_t^\top) V_t [C_l^\top (C_l V_t C_l^\top)^{-1} C_l - C_{l-1}^\top (C_{l-1} V_t C_{l-1}^\top)^{-1} C_{l-1}]\right] (e^\top V_t e)^{-1} \\
&= \sum_{t \in T} \operatorname{tr}\left[W_t (r_t r_t^\top) W_t [C_{l,t}^\top (C_{l,t} W_t C_{l,t}^\top)^{-1} C_{l,t} - C_{l-1,t}^\top (C_{l-1} W_t C_{l-1,t}^\top)^{-1} C_{l-1,t}]\right].
\end{aligned}
\tag{A.5}
$$

With constant weights this simplifies to

$$
\hat{\sigma}_{l,T}^2 = \operatorname{tr}\left[W_{T^*} \left(\sum_{t \in T} r_t r_t^\top\right) W_{T^*} [C_{l,T^*}^\top (C_{l,T^*} W_{T^*} C_{l,T^*}^\top)^{-1} C_{l,T^*} - C_{l-1,T^*}^\top (C_{l-1,T^*} W_{T^*} C_{l-1,T^*}^\top)^{-1} C_{l-1,T^*}]\right].
\tag{A.6}
$$

# B Bid-ask spread computation

For the computation of the bid-ask spread we use the daily series of CRSP data: BID - last available closing bid, ASK - last available closing ask, BIDLO - lowest trading price during the day or closing bid (if trading price unknown), ASKHI - highest trading price during the day or closing ask (if trading price unknown), and PRC - closing price or bid-ask average (if trading price not available). While the former series, BID and ASK, are clearly preferable, the series have a limited availability so we utilize on the information in the latter whenever BID/ASK are unavailable. We compute the daily bid-ask spread (BAS) in the following steps:

1. We set $BAS \coloneqq ASK - BID$.

2. If $BAS$ from 1. is negative, we set $BAS \coloneqq ASKHI - \max(BID, BIDLO)$.

3. If $BAS$ from 1. and 2. is not available (missing $BID/ASK$), we set it equal to $BAS \coloneqq ASKHI - BIDLO$.

4. If $ASK/ASKHI > 5$, we set $BAS \coloneqq ASKHI - BID$. This is done to clean the effect of extremely high quoted asks from market makers when he is not interested in trading.

5. If $BAS$ from 4. is negative, we set $BAS \coloneqq ASKHI - BIDLO$.

6. For any observation with percentage $BAS$ ($BAS$ scaled by $PRC$) greater than 5, we set $BAS = 2\min\{PRC - BIDLO, ASKHI - PRC\}$. This is done to mitigate the impact of extreme measurements and possible measurement errors from previous step.

Finally, we compute the daily percentage $BAS$ by scaling the $BAS$ by the price series $PRC$.

CRSP database offers own series of bid-ask spread on a monthly basis. Without any data cleaning the resulting percentage spreads range between -0.95 and 31.4. The major difference between our construction outlined above and the CRSP spread series is that the former provides a value of bid-ask spread for all firm-month observation, but the CRSP monthly values cover only 21% of total observations. A final difference in the two constructions stems from the fact that we consider monthly averages of the daily percentage BAS, while the CRSP computation uses monthly closing values.

The resulting time series of cross-sectional averages are qualitatively similar, with the CRSP-based series being above our computation in the majority of the sample. Both series are also strongly correlated: 0.86 on levels and 0.47 on differences. Graphical comparison of both series and an analysis of cross-sectional medians are available upon request.

## Tables

| Weighting | Frequency | 0.95 Lower CI | 0.95 Upper CI |
|---|---|---:|---:|
| VW | daily | 0.52 | 3.19 |
| EW | daily | 1.19 | 5.88 |
| VW | monthly | -2.24 | 5.51 |
| EW | monthly | 1.38 | 3.86 |

**Table 1.** Vogelsang $\mathrm{PS}^1 - t$ test of logarithm of average variance series. Weighting column describes the weighting scheme with EW and VW for equal an value weighting. “Daily” and “Monthly” values in Frequency field indicate whether one month of daily data, or two years of monthly data were used in IV estimation. The series based on one month are converted to two-year period. Values of the confidence intervals are multiplied by $1,000$.

| **Panel A: 1 month daily vs 2 years monthly** | | | | | | |
|---|---|---|---|---|---|---|
| | | | **1 year, monthly** | | | |
| **1 month, daily** | **NA** | **IV1** | **IV2** | **IV3** | **IV4** | **IV5** |
| **IV1** | 0.10 | 0.46 | 0.24 | 0.12 | 0.06 | 0.03 |
| **IV2** | 0.10 | 0.24 | 0.28 | 0.21 | 0.12 | 0.05 |
| **IV3** | 0.12 | 0.11 | 0.20 | 0.24 | 0.21 | 0.12 |
| **IV4** | 0.15 | 0.05 | 0.11 | 0.19 | 0.26 | 0.24 |
| **IV5** | 0.17 | 0.01 | 0.05 | 0.11 | 0.22 | 0.44 |

| **Panel B: 1 year monthly vs 2 years monthly** | | | | | | |
|---|---|---|---|---|---|---|
| | | | **2 years, monthly** | | | |
| **1 year, monthly** | **NA** | **IV1** | **IV2** | **IV3** | **IV4** | **IV5** |
| **NA** | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **IV1** | 0.08 | 0.66 | 0.19 | 0.05 | 0.01 | 0.00 |
| **IV2** | 0.09 | 0.21 | 0.44 | 0.19 | 0.06 | 0.02 |
| **IV3** | 0.11 | 0.01 | 0.24 | 0.40 | 0.19 | 0.05 |
| **IV4** | 0.15 | 0.00 | 0.01 | 0.23 | 0.43 | 0.18 |
| **IV5** | 0.19 | 0.00 | 0.00 | 0.00 | 0.18 | 0.63 |

| **Panel C: 1 month daily vs 2 years monthly** | | | | | | |
|---|---|---|---|---|---|---|
| | | | **2 years, monthly** | | | |
| **1 month, daily** | **NA** | **IV1** | **IV2** | **IV3** | **IV4** | **IV5** |
| **IV1** | 0.18 | 0.43 | 0.22 | 0.10 | 0.05 | 0.03 |
| **IV2** | 0.18 | 0.21 | 0.26 | 0.19 | 0.11 | 0.04 |
| **IV3** | 0.23 | 0.08 | 0.17 | 0.22 | 0.19 | 0.11 |
| **IV4** | 0.27 | 0.03 | 0.09 | 0.17 | 0.23 | 0.21 |
| **IV5** | 0.31 | 0.01 | 0.03 | 0.08 | 0.19 | 0.38 |

**Table 2.** Migration matrices: The numbers presented are percentages from row category which belong to each column basket. NA are missing values which occur because some stocks have an incomplete history, so that the corresponding measure cannot be computed.

**Panel A: One month, daily data**

| | All | | W/o new | | Merger | | Bankruptcy | | W/o delisting | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Count | Mean | Count | Mean | Count | Mean | Count | Mean | Count | Mean |
| **IV1** | 596,343 | 0.0116 | 558,663 | 0.0122 | 31,714 | 0.0264 | 9,494 | -0.0089 | 555,135 | 0.0111 |
| **IV2** | 596,075 | 0.0132 | 555,520 | 0.0137 | 24,584 | 0.0408 | 4,572 | -0.0423 | 566,919 | 0.0124 |
| **IV3** | 596,068 | 0.0131 | 544,300 | 0.0139 | 25,000 | 0.0452 | 7,686 | -0.0561 | 563,382 | 0.0126 |
| **IV4** | 596,075 | 0.0114 | 533,439 | 0.0127 | 24,602 | 0.0462 | 15,980 | -0.0664 | 555,493 | 0.0121 |
| **IV5** | 596,456 | 0.0099 | 532,633 | 0.0121 | 23,191 | 0.0442 | 57,725 | -0.0491 | 515,540 | 0.0149 |

**Panel B: One year, monthly data**

| | All | | W/o new | | Merger | | Bankruptcy | | W/o delisting | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Count | Mean | Count | Mean | Count | Mean | Count | Mean | Count | Mean |
| **IV1** | 519,992 | 0.0128 | 519,992 | 0.0128 | 16,769 | 0.0394 | 3,342 | -0.0076 | 499,881 | 0.0120 |
| **IV2** | 519,747 | 0.0129 | 519,747 | 0.0129 | 22,899 | 0.0404 | 4,052 | -0.0361 | 492,796 | 0.0120 |
| **IV3** | 519,764 | 0.0128 | 519,764 | 0.0128 | 26,012 | 0.0408 | 7,923 | -0.0505 | 485,829 | 0.0123 |
| **IV4** | 519,747 | 0.0129 | 519,747 | 0.0129 | 26,783 | 0.0421 | 17,042 | -0.0538 | 475,922 | 0.0137 |
| **IV5** | 520,126 | 0.0123 | 520,126 | 0.0123 | 23,197 | 0.0418 | 45,758 | -0.0491 | 451,171 | 0.0171 |

**Panel C: Two years, monthly data**

| | All | | W/o new | | Merger | | Bankruptcy | | W/o delisting | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Count | Mean | Count | Mean | Count | Mean | Count | Mean | Count | Mean |
| **IV1** | 455,693 | 0.0125 | 455,693 | 0.0125 | 13,859 | 0.0392 | 2,186 | -0.0008 | 439,648 | 0.0117 |
| **IV2** | 455,425 | 0.0130 | 455,425 | 0.0130 | 21,316 | 0.0391 | 2,335 | -0.0270 | 431,774 | 0.0120 |
| **IV3** | 455,444 | 0.0131 | 455,444 | 0.0131 | 24,554 | 0.0399 | 5,910 | -0.0515 | 424,980 | 0.0125 |
| **IV4** | 455,425 | 0.0139 | 455,425 | 0.0139 | 24,301 | 0.0432 | 13,533 | -0.0574 | 417,591 | 0.0145 |
| **IV5** | 455,814 | 0.0146 | 455,814 | 0.0146 | 18,252 | 0.0471 | 39,530 | -0.0444 | 398,032 | 0.0189 |

**Table 3.** Statistics of pre-delisting stocks, IV-portfolios, and IV-portfolios with exclusion of new stocks, or pre-delisting period (one year after listing or before delisting). “All” denotes the full portfolio of stocks, “W/o new” excludes the stocks in their first year on exchange (irrelevant for measures based on a longer history). “Merger” and “Bankruptcy” contain only stocks in the last year before delisting due to the corresponding cause, “W/o delisting” excludes the last year of pre-merger and pre-bankruptcy stocks. Panel A: IV portfolios based on one month of daily returns. Panel B: IV portfolios based on one year of monthly returns. Panel C: IV portfolios based on two years of monthly returns.

| Drop last | HP | 1 month, daily IV1 | IV2 | IV3 | IV4 | IV5 | 1 year, monthly IV1 | IV2 | IV3 | IV4 | IV5 | 2 years, monthly IV1 | IV2 | IV3 | IV4 | IV5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0Y** | **1M** | 0.92 | 1.01 | 1.04 | 0.74 | 0.23 | 0.91 | 0.96 | 0.97 | 1.00 | 0.78 | 0.89 | 1.02 | 1.05 | 1.00 | 0.87 |
| | **1Q** | 0.91 | 1.00 | 1.00 | 0.84 | 0.36 | 0.90 | 0.97 | 0.96 | 0.99 | 0.81 | 0.89 | 1.00 | 1.01 | 1.03 | 0.83 |
| | **2Q** | 0.92 | 0.97 | 0.99 | 0.88 | 0.45 | 0.90 | 0.96 | 0.96 | 0.99 | 0.83 | 0.89 | 0.98 | 1.01 | 1.02 | 0.80 |
| | **1Y** | 0.90 | 0.94 | 0.96 | 0.91 | 0.60 | 0.89 | 0.95 | 0.94 | 0.96 | 0.80 | 0.89 | 0.96 | 0.98 | 1.01 | 0.75 |
| | **2Y** | 0.89 | 0.95 | 0.97 | 0.94 | 0.75 | 0.88 | 0.96 | 0.96 | 0.98 | 0.79 | 0.89 | 0.97 | 0.98 | 1.05 | 0.74 |
| **1YB** | **1M** | 0.92 | 1.02 | 1.05 | 0.83 | 0.43 | 0.91 | 0.96 | 1.00 | 1.02 | 0.95 | 0.89 | 1.03 | 1.07 | 1.03 | 0.99 |
| | **1Q** | 0.91 | 1.00 | 1.02 | 0.92 | 0.54 | 0.90 | 0.98 | 0.98 | 1.01 | 0.94 | 0.89 | 1.01 | 1.03 | 1.05 | 0.96 |
| | **2Q** | 0.92 | 0.98 | 1.01 | 0.94 | 0.60 | 0.91 | 0.96 | 0.97 | 1.02 | 0.92 | 0.89 | 0.99 | 1.03 | 1.04 | 0.93 |
| | **1Y** | 0.90 | 0.95 | 0.97 | 0.94 | 0.71 | 0.89 | 0.94 | 0.95 | 0.99 | 0.86 | 0.88 | 0.97 | 0.98 | 1.03 | 0.85 |
| | **2Y** | 0.89 | 0.95 | 0.97 | 0.96 | 0.81 | 0.88 | 0.96 | 0.96 | 0.99 | 0.85 | 0.89 | 0.97 | 0.98 | 1.06 | 0.81 |
| **3YB** | **1M** | 0.93 | 1.02 | 1.07 | 0.94 | 0.59 | 0.92 | 0.96 | 1.02 | 1.07 | 1.08 | 0.89 | 1.04 | 1.10 | 1.06 | 1.14 |
| | **1Q** | 0.92 | 1.00 | 1.06 | 0.99 | 0.71 | 0.91 | 0.98 | 1.00 | 1.05 | 1.07 | 0.89 | 1.02 | 1.06 | 1.09 | 1.11 |
| | **2Q** | 0.92 | 0.99 | 1.03 | 1.01 | 0.78 | 0.91 | 0.97 | 0.99 | 1.06 | 1.06 | 0.89 | 1.00 | 1.05 | 1.08 | 1.09 |
| | **1Y** | 0.90 | 0.96 | 1.00 | 0.99 | 0.88 | 0.89 | 0.95 | 0.97 | 1.04 | 1.01 | 0.89 | 0.98 | 1.00 | 1.07 | 1.00 |
| | **2Y** | 0.89 | 0.96 | 1.00 | 1.01 | 0.96 | 0.88 | 0.96 | 0.98 | 1.03 | 0.98 | 0.89 | 0.98 | 0.99 | 1.10 | 0.95 |
| **1YBM** | **1M** | 0.90 | 0.98 | 0.99 | 0.73 | 0.24 | 0.89 | 0.91 | 0.94 | 0.93 | 0.81 | 0.88 | 0.97 | 1.02 | 0.90 | 0.89 |
| | **1Q** | 0.89 | 0.96 | 0.96 | 0.80 | 0.39 | 0.88 | 0.93 | 0.91 | 0.93 | 0.82 | 0.87 | 0.96 | 0.98 | 0.93 | 0.87 |
| | **2Q** | 0.90 | 0.94 | 0.94 | 0.84 | 0.47 | 0.89 | 0.92 | 0.91 | 0.95 | 0.83 | 0.87 | 0.95 | 0.97 | 0.94 | 0.84 |
| | **1Y** | 0.89 | 0.92 | 0.92 | 0.87 | 0.62 | 0.88 | 0.91 | 0.91 | 0.94 | 0.80 | 0.87 | 0.94 | 0.95 | 0.98 | 0.78 |
| | **2Y** | 0.89 | 0.94 | 0.96 | 0.94 | 0.79 | 0.88 | 0.95 | 0.95 | 0.98 | 0.82 | 0.89 | 0.97 | 0.97 | 1.05 | 0.79 |
| **3YBM** | **1M** | 0.91 | 0.99 | 1.01 | 0.85 | 0.45 | 0.90 | 0.92 | 0.96 | 1.02 | 0.93 | 0.88 | 0.99 | 1.05 | 1.01 | 1.03 |
| | **1Q** | 0.90 | 0.97 | 1.00 | 0.91 | 0.60 | 0.89 | 0.93 | 0.95 | 1.00 | 0.98 | 0.87 | 0.98 | 1.01 | 1.05 | 1.01 |
| | **2Q** | 0.90 | 0.96 | 0.98 | 0.93 | 0.68 | 0.89 | 0.93 | 0.94 | 1.02 | 0.99 | 0.87 | 0.97 | 1.00 | 1.04 | 1.01 |
| | **1Y** | 0.89 | 0.93 | 0.95 | 0.93 | 0.81 | 0.88 | 0.91 | 0.93 | 1.01 | 0.95 | 0.86 | 0.95 | 0.97 | 1.04 | 0.93 |
| | **2Y** | 0.88 | 0.93 | 0.95 | 0.96 | 0.91 | 0.87 | 0.93 | 0.95 | 1.00 | 0.93 | 0.87 | 0.95 | 0.96 | 1.08 | 0.87 |

**Table 4.** Expected returns: Role of distress, mergers and IV measurement. "Drop last" column indicates how long window before delisting is dropped. If followed by "B", only "Bankruptcies" are truncated, if followed by "BM", then mergers are dropped as well. "HP" denotes the holding period. Expected returns for horizons longer than one month are converted to a monthly basis.

| | 1 month, daily | | | | | 1 year, monthly | | | | | 2 years, monthly | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Drop last** | **IV1** | **IV2** | **IV3** | **IV4** | **IV5** | **IV1** | **IV2** | **IV3** | **IV4** | **IV5** | **IV1** | **IV2** | **IV3** | **IV4** | **IV5** |
| **Panel A: CAPM** | | | | | | | | | | | | | | | |
| **0Y** | 0.11 | 0.06 | −0.02 | −0.42 | −0.98 | 0.10 | 0.04 | −0.04 | −0.11 | −0.42 | 0.08 | 0.07 | 0.00 | −0.15 | −0.34 |
| | (0.17) | (0.38) | (0.89) | (0.07) | (0.00) | (0.20) | (0.54) | (0.76) | (0.58) | (0.18) | (0.33) | (0.34) | (0.97) | (0.50) | (0.30) |
| **1YB** | 0.11 | 0.08 | 0.00 | −0.33 | −0.78 | 0.10 | 0.05 | 0.00 | −0.08 | −0.24 | 0.09 | 0.09 | 0.03 | −0.11 | −0.22 |
| | (0.16) | (0.26) | (0.97) | (0.14) | (0.02) | (0.19) | (0.50) | (1.00) | (0.67) | (0.42) | (0.31) | (0.25) | (0.79) | (0.60) | (0.49) |
| **3YB** | 0.12 | 0.09 | 0.03 | −0.21 | −0.61 | 0.11 | 0.05 | 0.02 | −0.02 | −0.11 | 0.09 | 0.11 | 0.07 | −0.07 | −0.07 |
| | (0.15) | (0.22) | (0.81) | (0.33) | (0.05) | (0.15) | (0.51) | (0.85) | (0.91) | (0.71) | (0.32) | (0.14) | (0.59) | (0.74) | (0.82) |
| **1YBM** | 0.09 | 0.04 | −0.07 | −0.43 | −0.99 | 0.09 | −0.00 | −0.07 | −0.18 | −0.41 | 0.07 | 0.03 | −0.02 | −0.25 | −0.34 |
| | (0.26) | (0.57) | (0.59) | (0.06) | (0.00) | (0.26) | (0.95) | (0.53) | (0.35) | (0.18) | (0.40) | (0.72) | (0.89) | (0.25) | (0.29) |
| **3YBM** | 0.10 | 0.06 | −0.03 | −0.30 | −0.78 | 0.09 | 0.02 | −0.04 | −0.08 | −0.29 | 0.07 | 0.05 | 0.02 | −0.13 | −0.21 |
| | (0.24) | (0.43) | (0.80) | (0.17) | (0.01) | (0.23) | (0.84) | (0.68) | (0.67) | (0.33) | (0.41) | (0.47) | (0.89) | (0.54) | (0.52) |
| **Panel B: Fama-French model** | | | | | | | | | | | | | | | |
| **0Y** | 0.07 | 0.04 | −0.02 | −0.43 | −1.08 | 0.08 | −0.00 | −0.05 | −0.09 | −0.46 | 0.04 | 0.05 | 0.00 | −0.14 | −0.39 |
| | (0.21) | (0.61) | (0.85) | (0.01) | (0.00) | (0.14) | (0.96) | (0.65) | (0.54) | (0.03) | (0.46) | (0.50) | (0.97) | (0.39) | (0.10) |
| **1YB** | 0.08 | 0.05 | −0.00 | −0.34 | −0.88 | 0.08 | −0.00 | −0.01 | −0.06 | −0.28 | 0.05 | 0.06 | 0.03 | −0.10 | −0.27 |
| | (0.20) | (0.47) | (0.98) | (0.04) | (0.00) | (0.13) | (0.98) | (0.89) | (0.68) | (0.18) | (0.41) | (0.41) | (0.80) | (0.52) | (0.25) |
| **3YB** | 0.08 | 0.06 | 0.01 | −0.21 | −0.71 | 0.09 | −0.01 | 0.00 | −0.00 | −0.14 | 0.05 | 0.08 | 0.07 | −0.07 | −0.11 |
| | (0.18) | (0.44) | (0.90) | (0.16) | (0.00) | (0.09) | (0.94) | (0.99) | (0.98) | (0.50) | (0.44) | (0.28) | (0.56) | (0.64) | (0.65) |
| **1YBM** | 0.06 | 0.02 | −0.07 | −0.44 | −1.10 | 0.07 | −0.05 | −0.09 | −0.16 | −0.46 | 0.03 | 0.00 | −0.02 | −0.24 | −0.40 |
| | (0.34) | (0.82) | (0.47) | (0.01) | (0.00) | (0.21) | (0.47) | (0.42) | (0.32) | (0.03) | (0.57) | (0.96) | (0.90) | (0.14) | (0.09) |
| **3YBM** | 0.07 | 0.03 | −0.04 | −0.31 | −0.88 | 0.08 | −0.03 | −0.05 | −0.06 | −0.32 | 0.03 | 0.04 | 0.03 | −0.12 | −0.26 |
| | (0.29) | (0.65) | (0.69) | (0.05) | (0.00) | (0.16) | (0.67) | (0.58) | (0.72) | (0.14) | (0.58) | (0.64) | (0.81) | (0.45) | (0.27) |
| **Panel C: Four-factor model** | | | | | | | | | | | | | | | |
| **0Y** | 0.05 | 0.08 | 0.06 | −0.24 | −0.76 | 0.07 | 0.04 | 0.04 | −0.03 | −0.36 | 0.04 | 0.10 | 0.06 | −0.03 | −0.32 |
| | (0.42) | (0.28) | (0.55) | (0.17) | (0.00) | (0.18) | (0.59) | (0.72) | (0.84) | (0.11) | (0.56) | (0.23) | (0.64) | (0.85) | (0.18) |
| **1YB** | 0.05 | 0.09 | 0.07 | −0.16 | −0.59 | 0.07 | 0.04 | 0.07 | −0.01 | −0.19 | 0.04 | 0.11 | 0.09 | −0.03 | −0.19 |
| | (0.42) | (0.21) | (0.45) | (0.34) | (0.01) | (0.18) | (0.57) | (0.55) | (0.96) | (0.39) | (0.54) | (0.17) | (0.47) | (0.87) | (0.42) |
| **3YB** | 0.06 | 0.09 | 0.09 | −0.06 | −0.43 | 0.08 | 0.04 | 0.07 | 0.06 | −0.05 | 0.04 | 0.12 | 0.13 | −0.01 | −0.02 |
| | (0.40) | (0.21) | (0.34) | (0.69) | (0.06) | (0.14) | (0.62) | (0.49) | (0.70) | (0.83) | (0.57) | (0.14) | (0.30) | (0.97) | (0.93) |
| **1YBM** | 0.03 | 0.06 | 0.00 | −0.27 | −0.80 | 0.06 | −0.01 | −0.00 | −0.11 | −0.36 | 0.03 | 0.04 | 0.05 | −0.16 | −0.31 |
| | (0.63) | (0.44) | (0.97) | (0.12) | (0.00) | (0.26) | (0.90) | (0.98) | (0.49) | (0.11) | (0.69) | (0.56) | (0.69) | (0.32) | (0.19) |
| **3YBM** | 0.04 | 0.07 | 0.04 | −0.16 | −0.61 | 0.07 | 0.01 | 0.02 | −0.01 | −0.22 | 0.03 | 0.07 | 0.10 | −0.05 | −0.16 |
| | (0.57) | (0.34) | (0.72) | (0.33) | (0.01) | (0.22) | (0.89) | (0.86) | (0.97) | (0.35) | (0.71) | (0.35) | (0.46) | (0.75) | (0.51) |
| **Panel D: Five-factor model** | | | | | | | | | | | | | | | |
| **0Y** | 0.00 | 0.09 | 0.12 | −0.11 | −0.55 | 0.03 | 0.01 | 0.09 | 0.12 | −0.16 | −0.02 | 0.07 | 0.13 | 0.11 | −0.14 |
| | (0.97) | (0.28) | (0.25) | (0.55) | (0.04) | (0.61) | (0.90) | (0.42) | (0.44) | (0.48) | (0.80) | (0.36) | (0.31) | (0.46) | (0.58) |
| **1YB** | 0.00 | 0.09 | 0.13 | −0.05 | −0.37 | 0.03 | 0.01 | 0.10 | 0.13 | 0.01 | −0.01 | 0.08 | 0.15 | 0.10 | −0.00 |
| | (0.97) | (0.25) | (0.19) | (0.79) | (0.12) | (0.60) | (0.91) | (0.34) | (0.37) | (0.97) | (0.83) | (0.30) | (0.22) | (0.50) | (1.00) |
| **3YB** | 0.00 | 0.08 | 0.15 | 0.04 | −0.23 | 0.04 | 0.00 | 0.11 | 0.19 | 0.15 | −0.02 | 0.09 | 0.19 | 0.12 | 0.17 |
| | (0.94) | (0.29) | (0.13) | (0.79) | (0.34) | (0.50) | (0.98) | (0.30) | (0.20) | (0.49) | (0.78) | (0.27) | (0.14) | (0.47) | (0.46) |
| **1YBM** | −0.02 | 0.06 | 0.07 | −0.15 | −0.59 | 0.02 | −0.04 | 0.04 | 0.04 | −0.16 | −0.03 | 0.02 | 0.11 | −0.03 | −0.12 |
| | (0.79) | (0.45) | (0.51) | (0.42) | (0.02) | (0.76) | (0.59) | (0.73) | (0.81) | (0.49) | (0.68) | (0.78) | (0.36) | (0.86) | (0.63) |
| **3YBM** | −0.01 | 0.07 | 0.10 | −0.06 | −0.39 | 0.03 | −0.02 | 0.06 | 0.13 | −0.02 | −0.03 | 0.05 | 0.16 | 0.07 | 0.05 |
| | (0.87) | (0.41) | (0.33) | (0.75) | (0.11) | (0.63) | (0.77) | (0.56) | (0.42) | (0.94) | (0.67) | (0.53) | (0.22) | (0.67) | (0.85) |

**Table 5.** Portfolio alphas: Role of distress and IV measurement. Presented numbers are $100\alpha$ estimates from the corresponding factor model based on monthly returns, with (Newey-West adjusted) p-values in the parentheses. "Drop last" column indicates how long window before delisting is dropped. If followed by "B", only "Bankruptcies" are truncated, if followed by "BM", then mergers are dropped as well. Column names indicate the lookback period and sampling frequency used in computation of IV. Panel A: Alphas from CAPM. Panel B: Alphas from three-factor model of Fama and French (1993) (FF). Panel C: FF factors with momentum factor of Jegadeesh and Titman (1993). Panel D: FF factors with momentum factor and betting against beta of Frazzini and Pedersen (2014).

# Figures

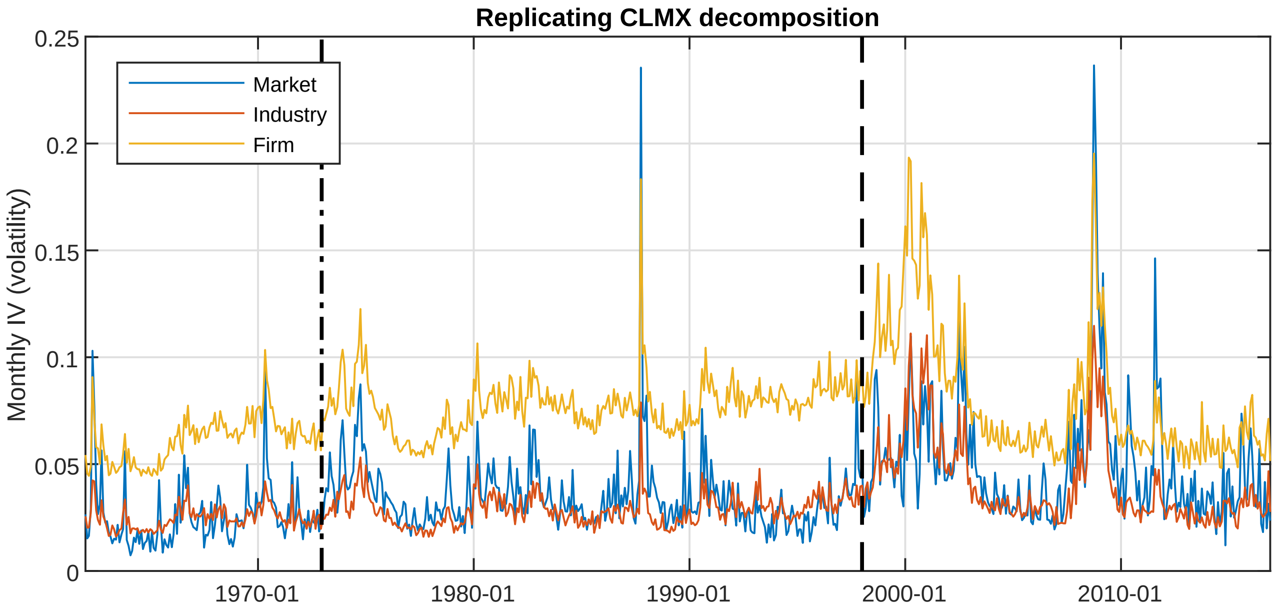



**Figure 1.** CLMX volatility decomposition using the initial capitalizations for both construction of the industry and market returns and for weighting the variances of industries and firms. Our sample spans the period from January 1962 to December 2016. Dot-dashed line, corresponding to December 14, 1972, marks the inclusion of NASDAQ stocks into the sample. The dashed line denotes the end of the CLMX sample (December 31, 1997).

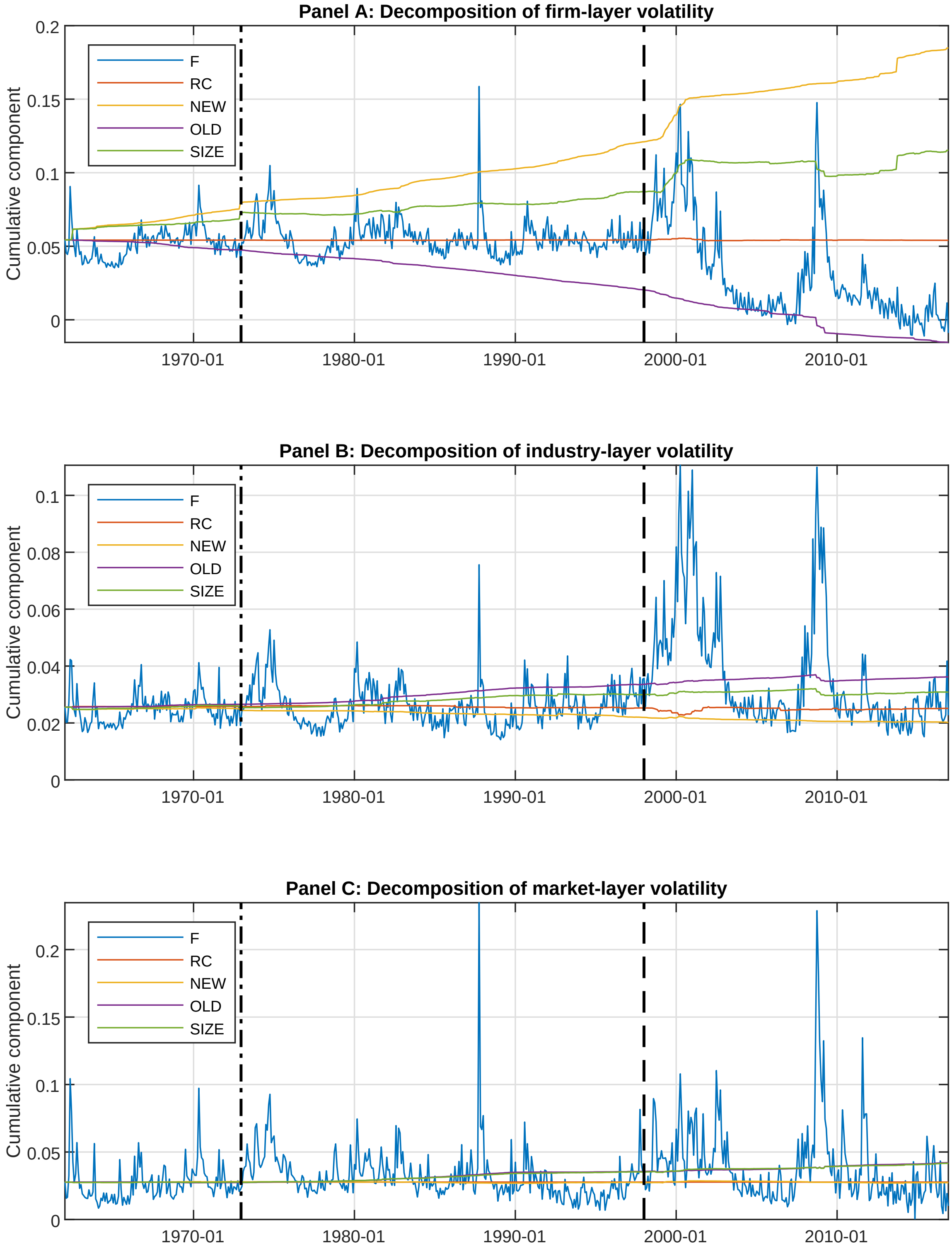



**Figure 2.** Decomposition of volatility layers to a fundamental (F), a reclassification (RC), a new listing (NEW), and a delisting (OLD) component; the size component (SIZE) equals a sum of the OLD and the NEW component. Panel A plots the firm layer, Panel B the industry layer, and Panel C the market layer. The sample spans the period from January 1962 to December 2016. Dot-dashed line, corresponding to December 14, 1972, marks the inclusion of NASDAQ stocks into the sample. The dashed line denotes the end of the CLMX sample (December 31, 1997).

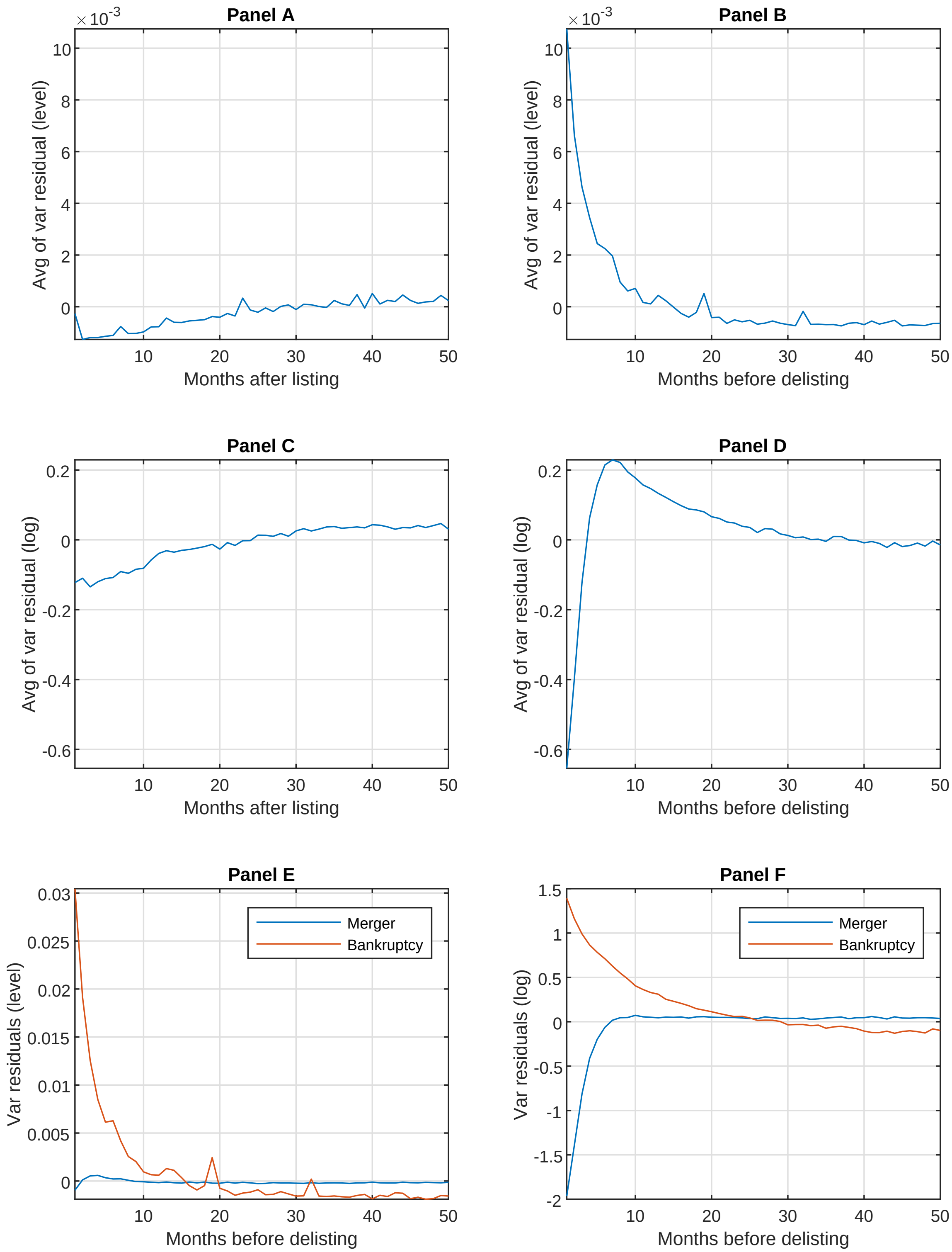



**Figure 3.** Panels A, B: Average of residuals from a fixed-effect regression on levels, months after listing (A) and months before delisting (B). Panels C, D: Average of residuals from a fixed-effect regression on log-levels (zero variances are excluded), months after listing (C) and months before delisting (D). Panel E: Average of residuals from a fixed-effect regression on levels by category. Panel F: Average of residuals from a fixed-effect regression on logarithm of variance (zero variances are excluded) by category.

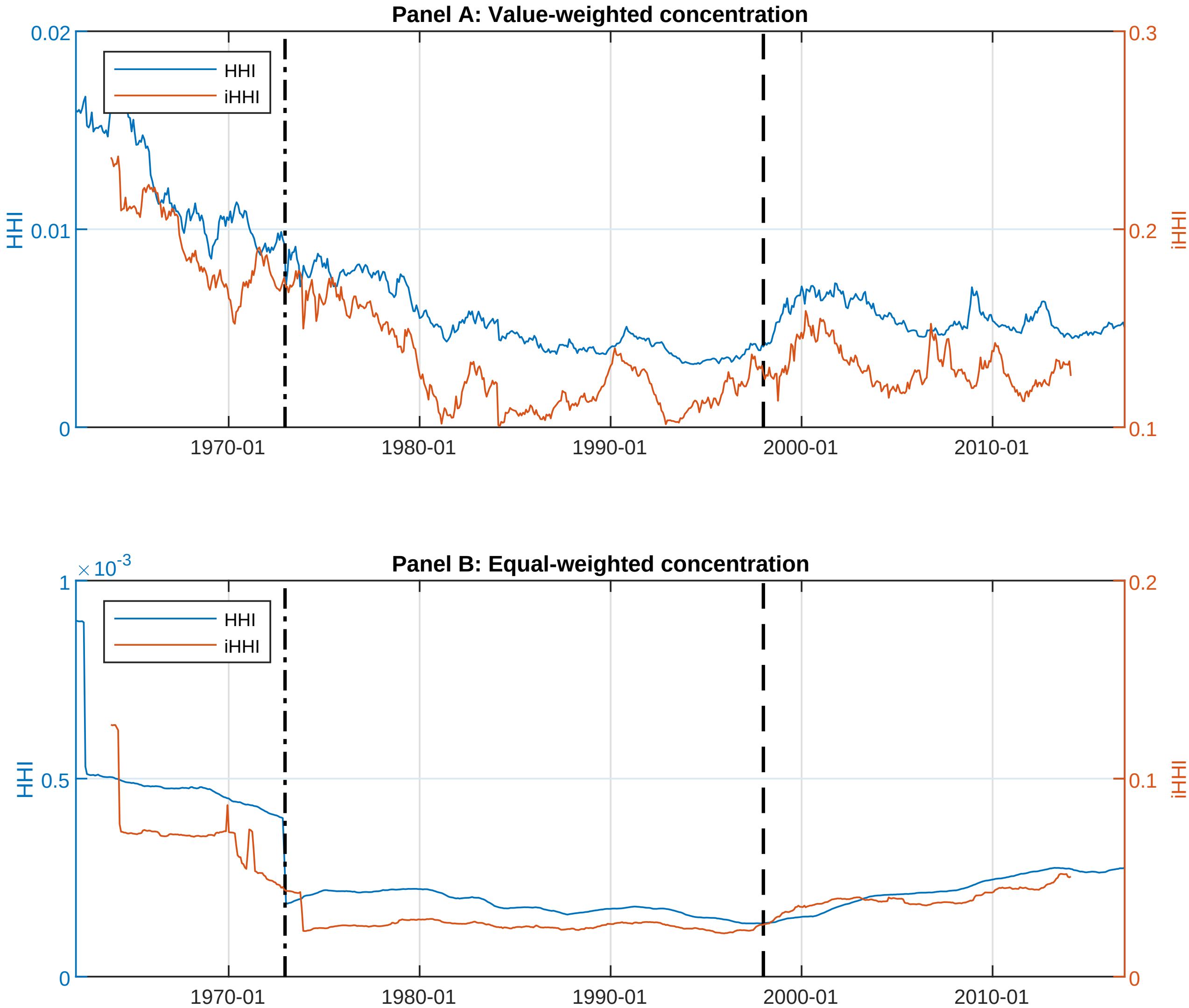



**Figure 4.** Panel A shows the time series of value-weighted market concentration measured by Herfindahl-Hirschman index (HHI), defined in equation (33), and a weighted average of industry concentrations (iHHI), defined in equation (34). In Panel B, we plot the market and weighted industry concentration based on equal weights, i.e., $w_i \equiv \frac{1}{F}$, where $F$ is the number of firms in a given period. All values are computed at the beginning of the month. The sample spans the period from January 1962 to December 2016. Dot-dashed line, corresponding to December 14, 1972, marks the inclusion of NASDAQ stocks into the sample. The dashed line denotes the end of the CLMX sample (December 31, 1997).

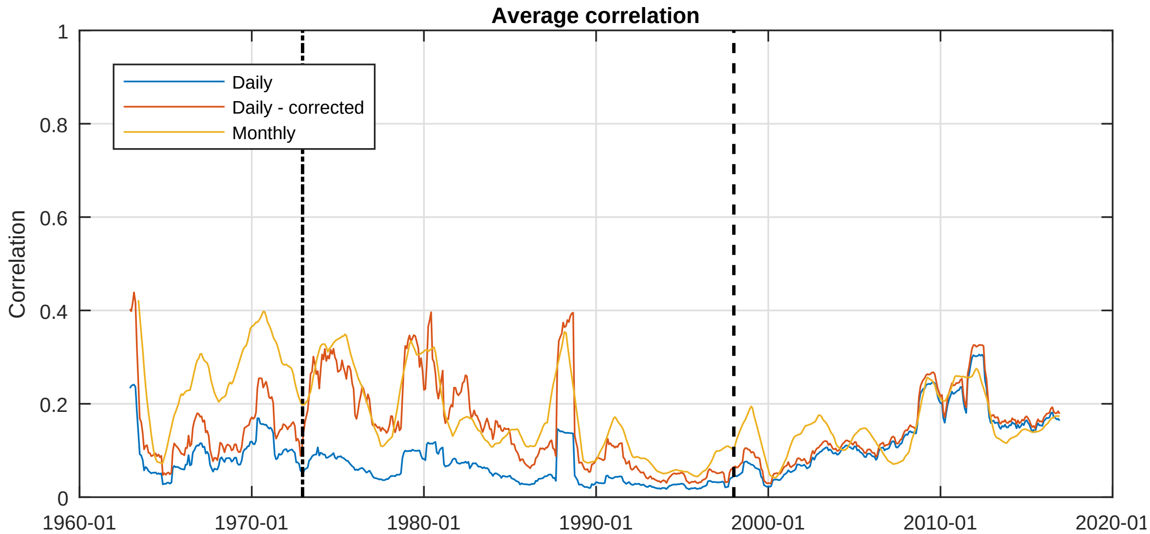
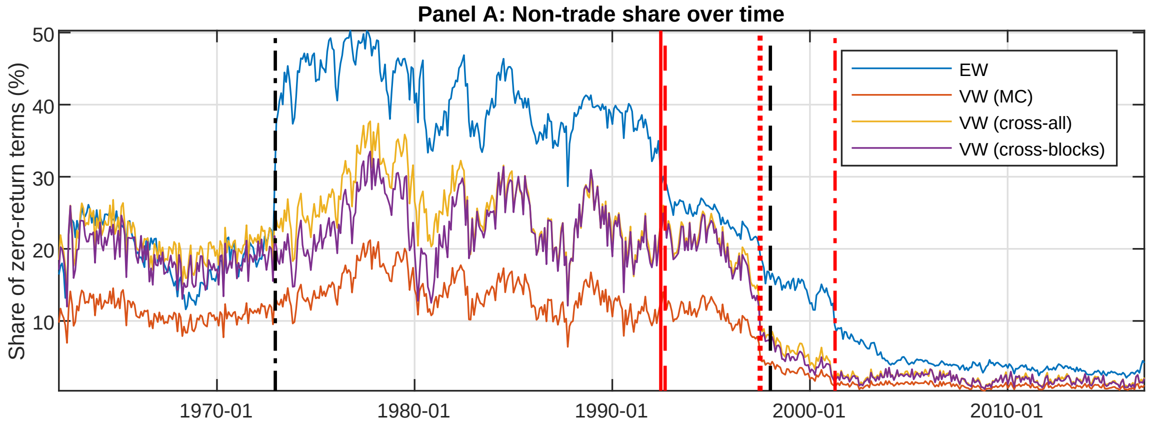



**Figure 5.** Equal-weighted correlations between all pairs of stocks, computed using one year of daily and monthly data. The daily - corrected series scales the average correlation computed using the daily data by $\bar{B}_{EW} = \frac{\sum_{i \neq j}^{N}(1-\lambda_i)(1-\lambda_j)}{N(N-1)/2}$, where $\lambda_i$ denotes the percentage of non-trades of asset $i$ in given month. $\bar{B}_{EW}$ reflects the correction for bias in sample correlations stemming from asynchronicity. Dot-dashed line, corresponding to December 14, 1972, marks the inclusion of NASDAQ stocks into the sample. The dashed line denotes the end of the CLMX sample (December 31, 1997).



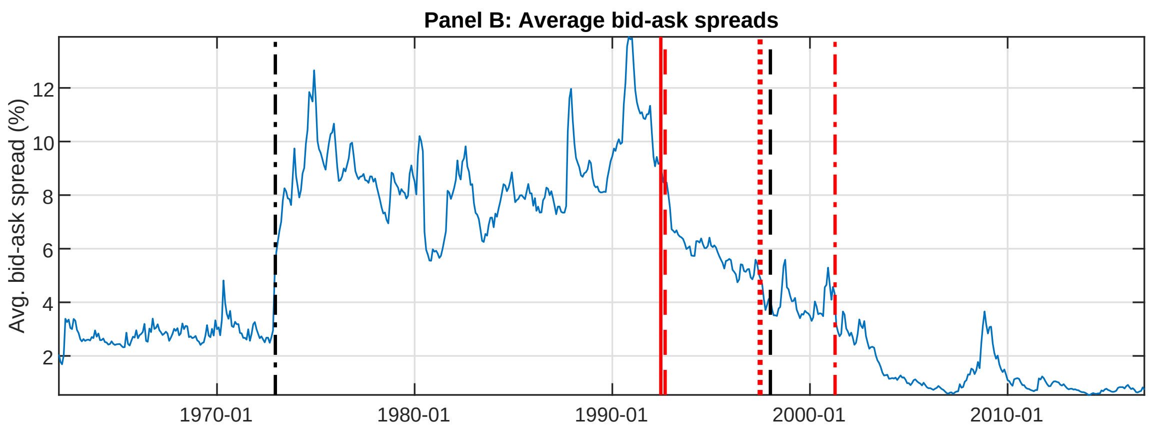




**Figure 6.** Panel A: The evolution of the non-trade share over time. EW series is computed as the proportion cross-firm average percentage of non-trades within month. VW weights each firm's non-trade share by its market capitalization. The series $(1 - \bar{B}_{EW})$ and $(1 - \bar{B}_{VW})$ correspond to the approximate percentage bias of equal-weighted and value-weighted (by $V_i V_j$ for correlation between assets $i$ and $j$) average of correlations; see Equations (36)-(40). Panel B: Average percentage bid-ask spread. The values presented are monthly averages from pooled sample of day-firm observations in given month. Details of the spread computation are provided in Appendix B. Dot-dashed line, corresponding to December 14, 1972, marks the inclusion of NASDAQ stocks into the sample. The dashed line denotes the end of the CLMX sample (December 31, 1997). Red vertical lines correspond to the dates of AMEX 1992 tick change (dashed), the 1997 tick change of NYSE, NASDAQ, and AMEX (dotted), the 2001 quote decimalization (dash-dotted), and the date (June 15, 1992) of available closing prices on NASDAQ SmallCap market (solid).

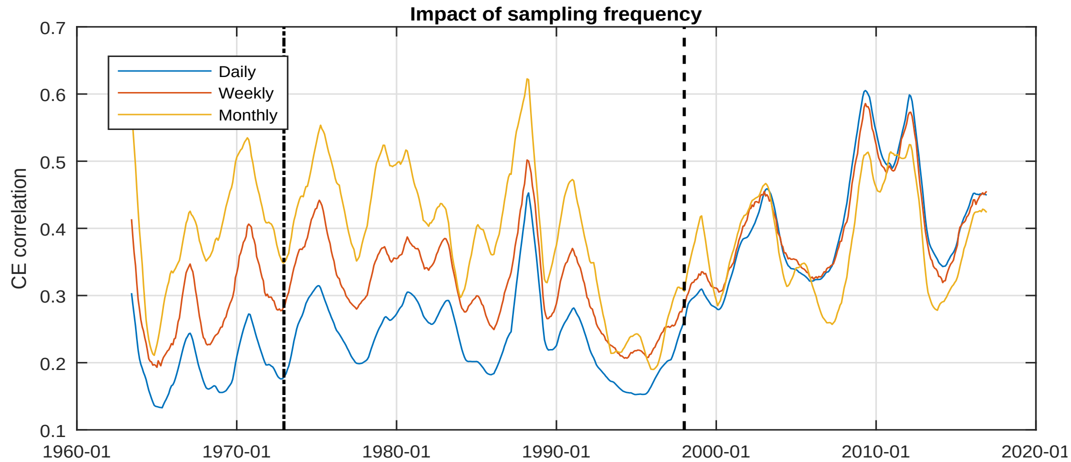



**Figure 7.** CE correlation estimates based on one year of data, using different sampling frequencies. Dot-dashed line, corresponding to December 14, 1972, marks the inclusion of NASDAQ stocks into the sample. The dashed line denotes the end of the CLMX sample (December 31, 1997).

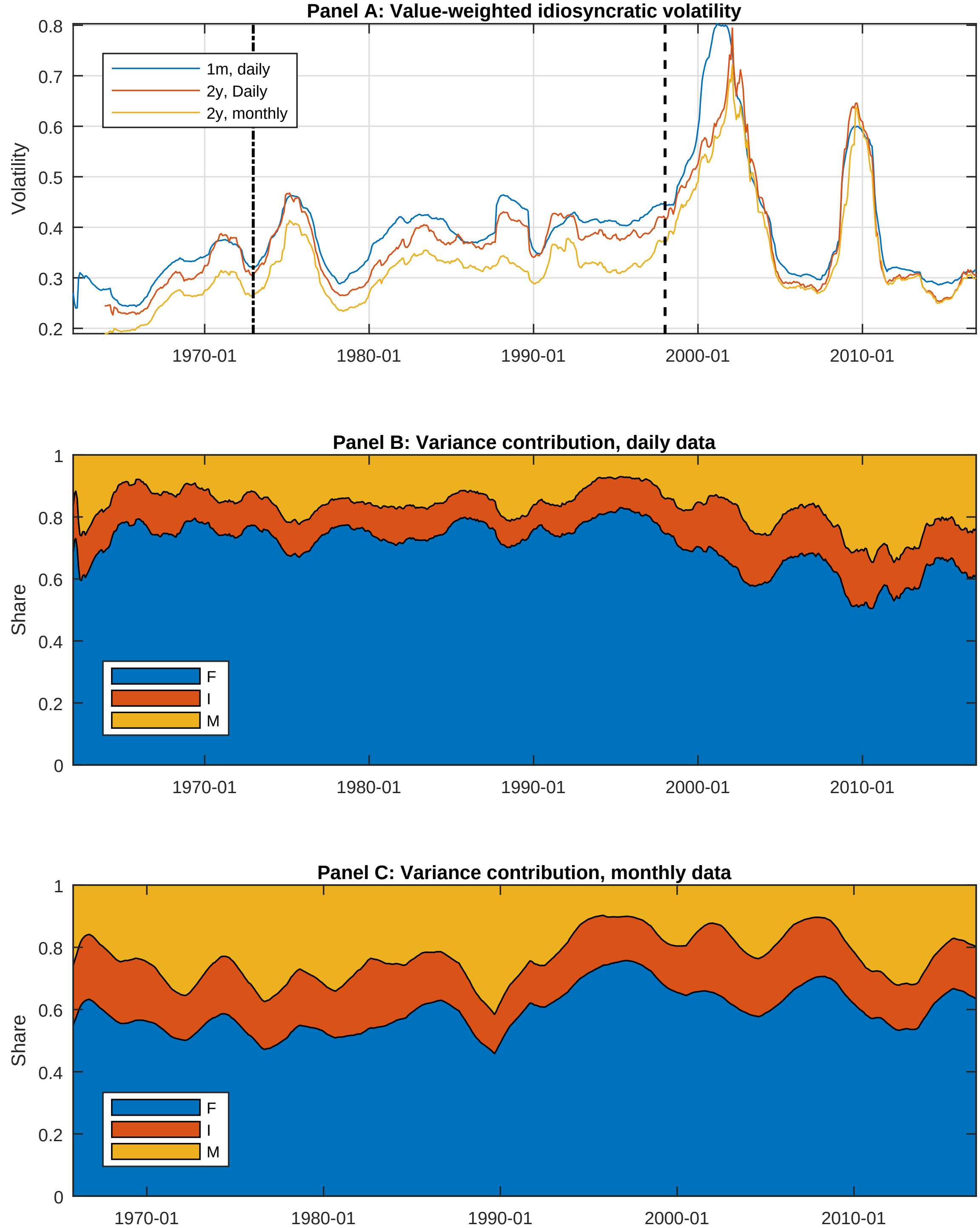



**Figure 8.** Value-weighted idiosyncratic volatility. Panel A: The series with "daily" and "monthly" in their label are computed on corresponding frequency. Labels "1m" and "2y" indicate whether one month or two years of data are used in computation. Dot-dashed line, corresponding to December 14, 1972, marks the inclusion of NASDAQ stocks into the sample. The dashed line denotes the end of the CLMX sample (December 31, 1997). Panel B: Contribution of variance components to total variance, using one month of daily data. Panel C: Contribution of variance components to total variance, using two years of monthly data. A 24-month moving-average filter is applied to the series based on one month.

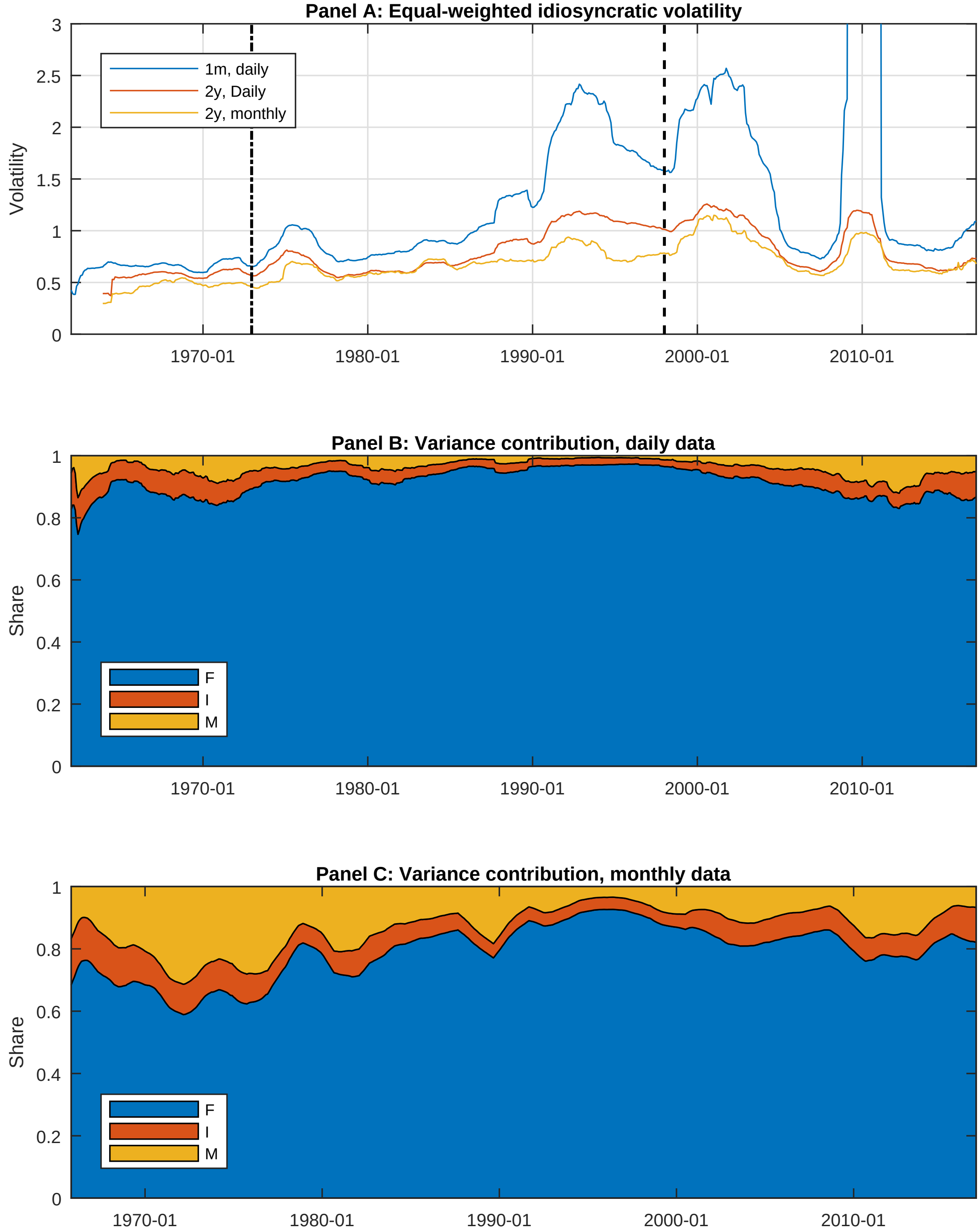



**Figure 9.** Equal-weighted idiosyncratic volatility. Panel A: The series with "daily" and "monthly" in their label are computed on corresponding frequency. Labels "1m" and "2y" indicate whether one month or two years of data are used in computation. Dot-dashed line, corresponding to December 14, 1972, marks the inclusion of NASDAQ stocks into the sample. The dashed line denotes the end of the CLMX sample (December 31, 1997). Panel B: Contribution of variance components to total variance, using one month of daily data. Panel C: Contribution of variance components to total variance, using two years of monthly data. A 24-month moving-average filter is applied to the series based on one month.

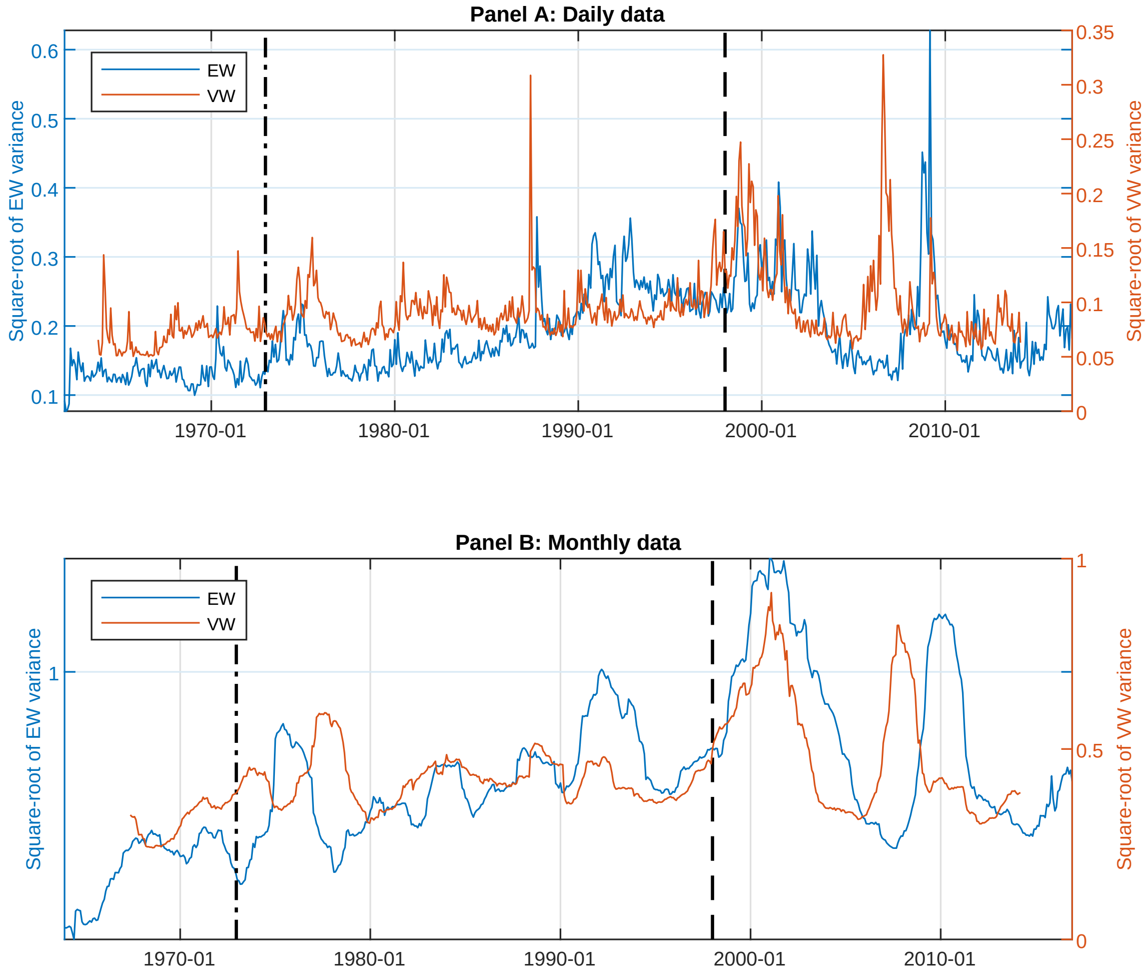



**Figure 10.** Comparison of square root of equal-weighted and value-weighted variances each month. Panel A: One month of daily data. Panel B: two years of monthly data. Dot-dashed line, corresponding to December 14, 1972, marks the inclusion of NASDAQ stocks into the sample. The dashed line denotes the end of the CLMX sample (December 31, 1997).

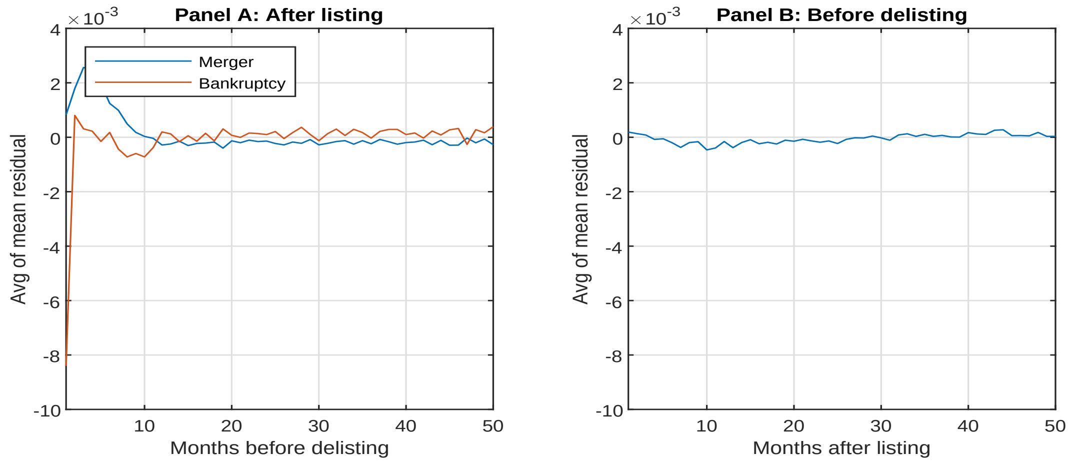



**Figure 11.** Panel A: Distribution of mean residuals after listing. Panel B: Distribution of mean residuals before delisting, by category.