Turning over Turnover

K. J. Martijn Cremers
Yale School of Management

Jianping Mei
Stern School of Business, New York University, and CKGSB

This article applies the methodology of Bai and Ng (2002, 2004) for decomposing panel data into systematic and idiosyncratic components to both stock returns and turnover panels. This approach works well for both returns and turnover, despite the presence of severe heteroscedasticity and nonstationarity of individual stocks' turnover. We test the mutual fund separation model of Lo and Wang (2000). Trading due to systematic risk in returns can account for 66% of systematic turnover. Thus, portfolio rebalancing due to systematic risk is a very important motive for stock trading. Finally, several common turnover measures may understate the impact of stock trading. (*JEL* G12)

There is increasing interest in improving the understanding of turnover, given its prominence in many behavioral finance, liquidity, and asymmetric information models.[1] Lo and Wang (LW hereafter, 2000, 2006) have developed a multifactor model for turnover based on asset-pricing models. They define turnover as the number of shares traded per period divided by the total number of shares outstanding. Their model introduces the decomposition of turnover into systematic and idiosyncratic components, similar to standard return decomposition. This model has several interesting applications for finance.

First, the decomposition of firm-level turnover into systematic and idiosyncratic components is useful for a large number of research questions. Most theoretical models of asset pricing and trading volume have systematic as well as firm-specific components;[2] so one needs separate

This article originally was titled "A New Approach to the Duo-factor model of Return and Volume". We would like to thank Andrew Ang, KC Chan, Kalok Chan, Lewis Chan, Robert Engel, Larry Harris, Campbell Harvey (the editor), Joel Hasbrouck, Anthony Lynch, Gideon Saar, Jeff Wurgler, Wei Xiong, Yexiao Xu and two anonymous referees, as well as seminar participants at NYU, HKUST, Rutgers University, USC, and AFA 2005 meetings for helpful discussions. We thank Dasol Kim and Haibo Tang for excellent research assistance. We also thank Andrew Lo and Jiang Wang, who graciously provided us with their Mini-CRSP data manual. Address correspondence to K. J. Martijn Cremers, International Center for Finance, Yale School of Management, 135 Prospect Street, New Haven, CT 6520, or e-mail: martijn.cremers@yale.edu.

[1] For theoretical models of turnover see, among others, Daniel, Hirshleifer, and Subrahmanyam (1998); Hong and Stein (1999, 2003); Scheinkman and Xiong (2003); Vayanos and Wang (2003), and Duffie, Garleanu, and Pedersen (2005). For empirical studies see Odean (1998); Chen, Hong, and Stein (2001); Ofek and Richardson (2003), and Mei, Scheinkman, and Xiong (2003).

[2] See, for example, Llorente et al. (2002) and Wang (1994).

doi:10.1093/rfs/hhm038 Advance Access publication August 27, 2007

measurements of these components to empirically test these models. So far, adjusting turnover for firm-fixed effects is typically dealt with by detrending total turnover [see, for example, Chen, Hong and Stein (2001)]. As we argue in this article, the systematic-idiosyncratic decomposition could be used as an alternative, and likely more effective, approach.

Furthermore, numerous studies have found common (i.e., systematic) components in liquidity [e.g., Chordia, Roll, and Subrahmanyam (2000) and Hasbrouck and Seppi (2001)]. The turnover decomposition directly measures how much trading is driven by systematic factors and how much is due to firm-specific causes. Using turnover as a proxy for informational trading, one can relate idiosyncratic turnover to firm-specific news and evaluate the impact of information asymmetry on asset pricing. In this article, we demonstrate that several commonly used turnover measures may understate the impact of stock trading.

Also, a better understanding of turnover can help provide more powerful asset-pricing tests. LW (2003) demonstrate that one can form a unique hedging portfolio using information from stock turnover. This would allow for additional tests of an asset-pricing model, since the hedging portfolio return together with the market return are the two risk factors determining the cross-section of asset returns.

Despite these interesting applications, studies of turnover have largely been limited to portfolios or to a small number of individual stocks. This may be partly due to the difficulty of implementing conventional multifactor estimation procedures, which results from severe heteroscedasticity and nonstationarity found in turnover data [see LW (2000)]. However, applying procedures developed by Bai and Ng (2002, 2004) [BN hereafter], we are able to consistently estimate the turnover factor model and to test for nonstationarity. In this way, we provide a close examination of turnover by ‘‘turning over’’ a large panel of individual stocks.

Our study makes a number of contributions to the turnover and factor model literature. First, we implement BN's methodology to decompose turnover into systematic and firm-specific components. We demonstrate that for estimating the required number of factors, the BN (2002) statistics work well for returns, but not for raw turnover. Instead, we propose a simple weighted least squares (WLS) procedure by standardizing turnover. We show how to use the BN (2004) method to test for nonstationarity in both systematic as well as firm-specific turnover components. The tests reveal that individual stock turnover consists of a systematic component that has a time trend and an idiosyncratic component that is essentially stationary. We find that three to five (depending on the period) systematic turnover factors together can capture 15.47–26.73% of the variation in individual stock turnover, quite similar to the percentages of systematic variation in individual stock returns.

Second, we provide a new test of the theoretical work by LW (2000). In particular, our empirical study uses data from a large panel of individual stocks rather than the beta-sorted portfolios that study used. As our empirical work shows, the number of systematic factors in return and turnover changes dramatically when individual stocks are used instead of beta-sorted portfolios.[3]

Third, our results indicate that trading due to systematic risk in returns can on average account for 66% of all systematic turnover variation in the weekly time series. Thus, portfolio rebalancing based on mutual fund separation is a very important motive for stock trading. We also find that there are more turnover factors than return factors in the data. This implies that the more restrictive hypothesis of the LW model, namely that returns and turnover should have the same number of systematic factors, is rejected by the data. Further, this suggests that further theoretical work is needed to unify stock price and trading volume under a multifactor asset-pricing-trading framework.

Fourth, our study complements recent market microstructure literature on common variation in liquidity and trading volume.[4] Chordia, Roll, and Subrahmanyam (2000) explore cross-sectional interactions in liquidity measures using quote data. Hasbrouck and Seppi (2001) use the Dow Jones Industrial Average of 30 actively traded firms. These studies all use high-frequency data rather than the weekly data used in our study. Using our new methodology, we find a *stronger* presence of commonality in turnover for most sample periods compared to these studies, and also provide an explicit test on the number of systematic factors in the turnover data.

Fifth and finally, in a direct application of the procedures advocated in this article, we provide average weekly liquidity estimates similar to Pastor and Stambaugh (2003). Using idiosyncratic turnover estimated from a multifactor turnover model, we demonstrate that several commonly used turnover measures may significantly underestimate the impact of stock trading.

The article is organized as follows. Section 1introduces an approximate multifactor model for turnover. We then present a consistent statistic developed by BN (2002) to determine the number of factors in the factor model and discuss how their framework can be employed in testing for nonstationarity in turnover data. In Section 2, we provide a description of the data set, followed by some evidence on the presence of severe

[3] Berk (2000) shows a significant drop in statistical power in asset-pricing tests using firm-characteristics sorted portfolios. Also, Brennan, Chordia, and Subrahmanyam (1998) report ". . . inferences are extremely sensitive to the sorting criteria used for portfolio formation, so that results based on regressions using portfolio returns should be interpreted with caution."

[4] This issue was highlighted by the Long Term Capital Management (LTCM) debacle, when there appeared to be a worldwide "flight-to quality" and a significant drop in trading volume across many assets.

heteroscedasticity and nonstationarity in turnover data. We then discuss several statistical procedures for dealing with these problems. Monte Carlo simulations are used to confirm our results. Section 3 briefly describes the LW model and provides an empirical examination of the importance of mutual fund separation in explaining individual stock turnover. Section 4 shows that several commonly used turnover measures may underestimate the price reaction to stock trading. Section 5 concludes.

1. Methodology for Decomposing Turnover

1.1 A multifactor turnover model and the BN (2002) statistic

LW (2000, 2006) provide a multifactor model for turnover:

$$\tau_{jt} = \tau_j + \delta_{j1} g_{1t} + \cdots + \delta_{jK} g_{Kt} + \xi_{jt} \tag{1}$$

where τ_{jt} is the turnover of stock j in period t (number of shares traded divided by the total number of shares outstanding), δ_{jk} is the exposure of firm j to economy wide trading shocks g_{kt}, and τ_j is a constant. Using terms common for discussing returns, we call δ_{jk} turnover betas. ξ_{jt} has mean zero and is assumed to be orthogonal to g_{kt}. In addition, we assume ξ_{jt} satisfies the regularity conditions as given in BN (2002, 2004).

We first estimate the common factors in Equation (1) using the asymptotic principal component method of [Connor and Korajczyk (1988), CK hereafter] based on the variance-covariance matrix of turnover over different periods rather than among different stocks. This is because the variance-covariance matrix for turnover among stocks or portfolios, *var* (τ_i, τ_j), is not well defined in the presence of time trends in turnover. We get around this problem by taking advantage of the large cross-section of individual stocks as in CK (1988) and using the variance–covariance matrix of turnover *over different time periods*. In other words, we apply a principal–component approach to *var* (τ_t, τ_s), where $Var\ (\tau_t, \tau_s) = N^{-1} \sum_{j=1}^{N} (\tau_{jt} - \overline{\tau}_t)(\tau_{js} - \overline{\tau}_s)$, and $\overline{\tau}_s = N^{-1} \sum_{j=1}^{N} \tau_{jt}$. $Var(\tau_t, \tau_s)$ is well defined for any given period t and s as long as the cross-sectional mean and variance for turnover exist for that period. Intuitively, *var* (τ_t, τ_s) depends on N-consistency rather than T-consistency. This implies τ_{jt} could have serial correlation as well as time-varying mean and volatility. The factor estimates could still be consistent as long as the data satisfy some necessary moment conditions [for details, see BN (2002)]. Throughout this article, we will use *var* (τ_t, τ_s) for factor extraction.

Because the true number of factors K is unknown, we start with an arbitrary number $k_{\max}$ ($k_{\max}$<min (N,T)). We estimate the k systematic

factors and factor loadings that solve the following optimization problem:

$$V(k) = \min_{D^k, G^k} T^{-1} N^{-1} \sum_{t=1}^{T} \sum_{j=1}^{N} (\tau_{jt} - D_j^k G_t^k)^2 \tag{2}$$

where G^k denotes the k-vector of systematic factors and D_j^k denotes k-vector of factor loadings for firm j.[5]

To determine the number of factors, BN propose the following selection criterion:

$$\hat{K} = argmin_{0<k<kmax} IC(k), \tag{3}$$

where *IC(k)* equals the measure of the goodness of fit $V(k)$ plus a second term that serves as an adjustment for the increase in the degrees of freedom that result from increasing k:

$$IC(k) = \log\{V(k, \hat{G}^k)\} + k \cdot \left(\frac{N+T}{NT}\right) \ln\left(\frac{NT}{N+T}\right) \tag{4}$$

BN show that $\hat{K}$, the value of k that minimizes the *IC(k)* statistic in (4), is a consistent estimate for the number of factors in the factor model.[6]

Intuitively, the estimation procedure treats the determination of the number of factors as a model selection problem. As a result, the selection criteria depend on the usual trade-off between the goodness of fit and parsimony (or model size). The difference here is that the *IC(k)* statistic not only takes the sample size in both the cross-section and the time-series dimensions into consideration, but also the fact that the factors are not observed.

There are several distinctive advantages of the BN statistic of (4) compared to the Δ statistic used in CK (1993). First, BN do not impose any restrictions between N and T, allowing for both large N and large T. Second, the results hold under heteroscedasticity in *both* the time and the cross-section dimensions. Third, the results also hold under *both* weak serial dependence and cross-section dependence. In addition, the model selection procedure is easy to implement. BN (2002) give the conditions under which the consistency of $\hat{K}$ holds.[7]

[5] Note that the minimization problem in (2) yields the same CK factor estimates, at least up to some normalization.

[6] They also propose two other asymptotically equivalent statistics. Our empirical study finds these give similar results in our balanced panel, but the information criteria (IC) has the best simulation results. Results are available on request.

[7] Jones (2001) introduces a heteroscedastic factor analysis (HFA) approach to extract factors but he does not provide a test on the number of factors, as CK (1993) do.

1.2 The BN (2004) PANIC test for nonstationarity

BN (2004) develop a PANIC (panel analysis of nonstationarity in idiosyncratic and common components) methodology to detect whether there is nonstationarity in either the systematic or idiosyncratic components, or in both. They examine the factor structure of a large panel data set, showing that the components can be consistently estimated using the panel, even in cases where individual (nonstationary) series would produce spurious regressions. In particular, they show that common stochastic trends can be consistently estimated by using the principal components method, *regardless* of whether the idiosyncratic series contain unit roots. Similarly, their proposed unit root test of the idiosyncratic series is valid whether any of the systematic factors contain a unit root. A great advantage of PANIC is that it directly tests the unobserved components of the data.

Under certain regularity conditions, BN (2004) show that the standard Dickey–Fuller (1979) statistics for testing for a unit root—with either only a constant or with a constant plus a linear time trend—in the estimated factors or in the idiosyncratic turnover have the same limiting distribution as the regular test statistics for observed data series, as derived in Fuller (1976). As a result, the 5% asymptotic critical value of the Dickey–Fuller unit root test of −2.86 applies. Therefore, we will begin by estimating the systematic factors using the principal components and then perform unit root and trend tests on estimated common and idiosyncratic components.

2. Dealing with Severe Heteroscedasticity and Nonstationarity in Turnover

2.1 Data description and a WLS solution to severe heteroscedasticity

Following LW (2000), we use the Center for Research in Security Prices (CRSP) daily file to construct weekly turnover series for individual NYSE and AMEX stocks from July 1967 to December 2001. The choice of a weekly horizon makes our results comparable to LW and is a compromise between maximizing sample size while minimizing the day-to-day volume and return fluctuations that have less direct economic relevance.[8]

Because our focus is on the implications of portfolio theory for trading behavior, we limit our attention to ordinary common shares on the NYSE and AMEX (CRSP share codes 10 and 11 only), omitting American Depositary Receipts (ADRs), Real Estate Investment Trusts (REITs), closed-end funds, and those others whose trading volume or turnover

[8] In addition to the weekly data, we conducted a parallel study of turnover by using monthly and intraday data. The monthly results, briefly discussed in the conclusion, are similar. For the intraday data, we replicated our results for the 1997–2001 period using the TAQ database, aggregating the number of traded shares in the TAQ database to weekly volume and turnover data. The average correlation of the weekly turnover from TAQ and CRSP is greater than 99%. For the standardized TAQ turnover data, the number of factors in turnover equals 4, the same as in the CRSP data. We conclude that the TAQ results confirm the robustness of the results using standardized turnover data.

may be difficult to interpret. We also omit NASDAQ stocks because of market structure differences relative to the NYSE and AMEX. Like LW, we discard firms that have either no or problematic turnover data.

Table 1 presents a brief summary of our data sample. The first column provides the number of securities traded on the two exchanges. The second column provides our unbalanced sample, which includes the number of securities with <50% missing observations in turnover and no problematic data. The third column provides our balanced sample, including the number of securities with neither missing observations in turnover nor problematic data.[9] As we find that the BN methodologies work equally well on either the balanced sample or the unbalanced sample (with <50% missing observations in turnover), for computational ease we will henceforth proceed using only the balanced sample.[10] The last column provides average weekly turnover estimates for the various sample periods, which are comparable to earlier studies such as LW (2000).

A close examination of the cross-sectional variation in turnover volatility indicates that the turnover distribution generally displays much larger skewness, as well as increased kurtosis, than returns. This provides the empirical justification for standardizing the turnover data. We also find that turnovers for many stocks display a strong presence of nonstationarity. In particular, there appears to be a strong trend component in turnover, which we will examine in great detail in Section 3. However, by performing several (augmented) Dickey–Fuller tests on individual stock turnover, we find no evidence of unit roots among any of the stocks in our sample for all of the seven periods. That is, for all of the firm-level individual turnover time series, we can reject a unit root with and without a time trend present. Further, unit root tests on decomposed turnover confirm that neither the systematic factors nor any of the idiosyncratic turnover components have a unit root in any period. As a result, we will not first-difference the data, as first-differencing the turnover data tends to increase noise in the idiosyncratic term when there is no unit root in turnover. This would result in significantly increased estimation error, leading to poor finite sample properties.[11]

The strong presence of both heteroscedasticity and time trends in turnover considerably affects the estimation of the number of systematic

[9] We consider two types of problematic data. The first type includes firms that have constant turnover in the period. The second are those firms that have likely data entry problems as evidenced by an unusually large standard deviation (specifically, 10 times the average standard deviation. See also the discussion on the so-called Z-flag in LW. As they argue, such large standard deviations probably indicate data errors).

[10] For example, simulation studies indicate that the factor extraction for the unbalanced turnover panels is quite as reliable as for the balanced panels (as shown in Table 4). Moreover, for the unbalanced turnover panels type I and type II errors are found similar to those in Table 7. Not surprisingly, we find evidence of one additional turnover factor in the unbalanced panel. The number of return factors is indifferent between the balanced and unbalanced panels. These results are available upon request.

[11] These results are presented in an earlier version of the article and are available upon request.

Table 1
Summary statistics

Dates	Number of firms in NYSE & AMEX	Firms with >50% turnover data and no problem data	Firms with no missing and problem data	Average weekly turnover (%)
1967–1971	2510	2159	1586	0.89
1972–1976	2527	2413	1912	0.53
1977–1981	2288	2134	1753	0.72
1982–1986	2141	1977	1514	1.00
1987–1991	1977	1757	1399	1.04
1992–1996	2249	2057	1528	1.06
1997–2001	2502	2104	1385	1.43

The table presents the number of stocks on NYSE & AMEX during each 5-year period, the number of stocks with at least half of the weekly turnover data available and without problem data in turnover (i.e., those with CRSP Z-flag), the number of firms with no missing observations in turnover (our balanced sample of firms used most widely in this article), and finally the average weekly turnover. Only common shares are considered (selected from CRSP share using codes 10 and 11).

turnover factors. Table 2 presents estimates of the number of factors for each sample period using raw and standardized turnover, as well as their detrended series. This standardization is conducted by first de-meaning, and then normalizing, each stock turnover series by its sample standard deviation over the relevant period. The number of factors reported here corresponds to the number that minimizes the IC statistic. Specifically, to determine the number of systematic factors in turnover, we compute a goodness of fit statistic, IC, conditional on a wide range of numbers of factors. For example, comparing $IC(k)$ for $k = 1, 2, \ldots, 20$ indicates that $k = 5$ provides the minimum $IC(k)$ for standardized turnover for 1967–71. This indicates that there may be five systematic factors for standardized turnover during the first sample period.

As can be seen from Table 2, the estimates of the number of factors for turnover are very sensitive to the standardization of the data, and somewhat sensitive to detrending. There seem to be an extraordinary large number of factors in raw turnover data. For example, there may be 16 systematic factors during 1967–71. While detrending should not reduce the number of factors by more than one factor, the results for raw turnover suggest otherwise.

In marked contrast, the estimates of the number of factors for excess returns are robust to standardization. Since standardization should not change the number of factors found, we conclude that, owing to the presence of severe heteroscedasticity in turnover, the BN statistics do not work well for raw turnover.

The main reason for the failure of the BN (2002) procedure in raw turnover is the presence of severe heteroscedasticity. Using raw turnover essentially gives the stocks with enormous swings in turnover a lot more weight in the sum-of-squared residuals in Equation (2). This could be mitigated by the standardization of raw turnover. This in effect amounts

Table 2
Impact of heteroscedasticity and nonstationarity on estimates of the number of factors for turnover

Time	Turnover						Excess returns	
period	Raw level	Raw detrended	Raw 5% cutoff	Raw 10% cutoff	Standardized level	Standardized + detrended	Raw	Standardized
1967–1971	16	16	10	9	5	4	2.0	2.0
1972–1976	16	15	8	7	4	4	2.0	2.0
1977–1981	11	10	7	6	5	4	2.0	2.0
1982–1986	8	8	5	5	4	3	2.0	2.0
1987–1991	11	8	5	5	3	3	2.0	2.0
1992–1996	12	10	6	5	4	3	2.0	2.0
1997–2001	12	11	7	6	4	4	3.0	3.0

The table presents the number of factors found in a large *balanced* panel of weekly turnover and returns, using the **BN** (2002) methodology. We use turnover for NYSE and AMEX ordinary common shares from January 1967 to December 2001. We use raw as well as standardized returns, and raw, detrended, standardized and standardized-and-detrended turnover. We also use a subset of raw turnover data by removing those with top 5 or 10% of highest kurtosis in turnover. We standardize each firm's turnover or return series by taking out the time series average and dividing by the time-series standard deviation.

to using WLS rather than OLS in the turnover regression of Equation (1). CK (1988) also propose using residual standard deviation to standardize returns for better return factor extraction.

As Table 2 shows, the use of standardized turnover leads to a large drop in the number of factors relative to the results for raw turnover. For example, for 1967–71 the estimated number of factors drops from 16 for raw turnover to 5 for standardized turnover. Detrending standardized turnover does not result in a similar sharp drop in the number of systematic factors. In Section 2.3, we will use Monte Carlo simulations to demonstrate that the BN (2002) statistic has good small-sample properties for standardized turnover, but not for raw turnover.

To understand what causes the sharp drop from 16 to 5 factors when we standardize, we look at what happens if we remove those stocks in the top 5 or 10% of highest kurtosis in turnover. We observe that the number of factors drops sharply by 6 or 7 when we use raw turnover (see 3rd and 4th columns of Table 2). On the other hand, the results for standardized data remain unchanged. Therefore, kurtosis is no problem at all for the standardized data, while the high kurtosis of a few stocks strongly affects the estimation for raw data. The reason for this, intuitively, is that the turnover of a few stocks with huge fluctuations is inordinately weighted in the IC computation. As a result, the BN selection criterion is forced to treat (some combination of the turnover of) these stocks as factors. However, these "factors" do not have pervasive impact across all stocks, which is why the number of factors drops sharply when those few stocks with high turnover kurtosis are removed. This also explains why our bootstrap simulation results (see Subsection 2.3) for raw turnover are so unstable—because they depend on the selection or exclusion of the few high kurtosis stocks. Standardization, however, downplays the importance of these stocks, and seems to pick up only those factors that are pervasive. As a result, after standardization, the number of factors chosen is no longer sensitive to the removal of those high kurtosis stocks.

2.2 The number of factors in turnover

The top panel of Table 3 provides the results of the test of the number of factors in standardized turnover. We report the incremental proportions of the explained variation (i.e., the mean R^2) from regressing the individual firm turnover series on 1 to 10 ordered turnover factors. The first principal component of turnover typically explains between 6.5 and 15.0% of the variation of the standardized turnover. Further examination of our results shows that the fourth and fifth components still explain a fair amount of turnover variation. For example, the fifth component explains 1.95% of variation for 1967–71.

The IC procedure selects a five-factor model for the first sample period, which is also reported in Table 3. It is reassuring to see that the number of

Table 3
Test of the number of factors in turnover and returns using the balanced panel

Standardized turnover

Period	$\theta 1$ (%)	$\theta 2$(%)	$\theta 3$(%)	$\theta 4$(%)	$\theta 5$(%)	$\theta 6$(%)	$\theta 7$(%)	$\theta 8$(%)	$\theta 9$(%)	$\theta 10$(%)	No. of factors	Average R^2(%)	Std R^2(%)
67–71	11.14	5.71	4.45	2.23	1.95	1.64	1.47	1.31	1.27	1.16	**5**	25.47	12.93
72–76	15.03	7.30	2.25	2.15	1.69	1.66	1.42	1.30	1.23	1.15	**4**	26.74	14.18
77–81	11.48	4.29	3.22	2.18	1.95	1.58	1.47	1.39	1.23	1.14	**5**	23.13	12.32
82–86	10.73	5.59	2.39	2.28	1.68	1.41	1.26	1.19	1.10	1.09	**4**	21.00	11.26
87–92	11.45	4.07	2.89	1.90	1.80	1.45	1.20	1.15	1.08	1.05	**3**	18.41	13.01
92–96	6.54	3.90	2.81	2.22	1.83	1.55	1.41	1.29	1.17	1.10	**4**	15.47	10.35
97–01	10.79	4.00	3.13	2.47	1.92	1.64	1.50	1.29	1.22	1.14	**4**	20.39	14.02

Excess returns

Period	$\theta 1$ (%)	$\theta 2$(%)	$\theta 3$(%)	$\theta 4$(%)	$\theta 5$(%)	$\theta 6$(%)	$\theta 7$(%)	$\theta 8$(%)	$\theta 9$(%)	$\theta 10$(%)	No. of factors	Average R^2(%)	Std R^2(%)
67–71	21.47	1.90	1.66	1.07	1.08	0.91	0.82	0.78	0.74	0.72	2	23.36	8.52
72–76	22.33	2.80	1.68	1.26	1.03	0.99	0.91	0.87	0.87	0.83	2	25.13	10.10
77–81	19.62	2.75	1.74	1.58	1.02	0.81	0.75	0.75	0.72	0.70	2	22.37	11.00
82–86	17.68	2.71	1.90	1.35	0.99	0.94	0.83	0.82	0.77	0.74	2	20.39	11.15
87–92	25.73	2.79	1.57	1.32	1.15	1.02	0.91	0.88	0.83	0.78	2	28.52	14.77
92–96	10.92	3.36	1.76	1.44	1.28	1.21	1.02	0.95	0.87	0.86	2	14.28	11.58
97–01	13.32	3.50	2.54	1.63	1.52	1.33	1.05	1.06	0.91	0.90	3	19.36	13.24

The table gives the incremental R^2 explained by subsequent ordered eigenvectors, θk, $k = 1, \ldots, 10$ of the covariance matrix of weekly turnover of NYSE and AMEX ordinary common shares for seven subperiods from July 1967 to December 2001. We also report the number of factors selected by the IC and cross-sectional average R^2 for the selected factor model for each sample periods.

factors identified by the IC statistic closely corresponds with the eigenvalues of the principal components. The eigenvalues of the statistically significant turnover factors typically exceed 1.95%.

Our result of four or five factors in standardized turnover (also reported in Table 2) is quite different from the results reported in LW, who find only one or two significant systematic factors, although they do not have a formal statistic testing for the number of factors. This difference can be seen as largely due to their use of factors extracted from 10 beta-sorted portfolios, while we use a large cross-section of individual stocks. As pointed out by Shukla and Trzcinka (1990), beta-sorted portfolios may lead us to find fewer factors in the model.[12]

While our procedure does not specifically identify what the factors are, it does provide some guidance for equilibrium model construction. Our results suggest that the two-factor model of LW (2003), which consists of a market factor and a hedging factor, has left out several systematic factors. This may help explain why their model, while providing a reasonable description of portfolio turnover, still does not fully capture the cross-section of average turnover.

Table 3 also reports the average R^2 of regressing individual stock turnover on the selected systematic turnover factors for each sample period. For example, for 1967–71 a five-factor model explains on average 25.5% of variation in turnover of individual stocks.[13] This is in sharp contrast to LW's finding that a two-factor model captures well over 90% of the time-series variation in turnover when using turnover-beta-sorted portfolios. Thus, there is much more heterogeneity in individual stock turnover than is represented in their studies.

To understand the relative importance of factor models in explaining returns vs. turnover, the bottom panel of Table 3 provides the estimates on the number of factors in excess returns. The same firms are used in both the return and the turnover sample. Similar to turnover, the first principal component of returns typically explains between 11 and 26% of the variation of excess returns, while the second and third components each explain about 2%. This is quite different from LW, who use returns from broadly diversified portfolios. Their first principal component typically explains over 70% of the variation in the portfolio returns. We also find that there are two or three pervasive return factors in the economy, depending on the period. Comparing the average R^2 results for turnover

[12] This can be shown by assuming the $N \times T$ matrix of turnover $\mathbf{\Gamma}$ is generated by Equation (9) in the Appendix: $\mathbf{\Gamma} = \mathbf{D\,G} + \mathbf{H}$, where $\mathbf{D}$ is the $N \times K$ matrix of turnover betas and $\mathbf{G}$ is the turnover factors. Now let us form L beta-sorted portfolios. This amounts to transforming the data by a $L \times N$ matrix $\mathbf{P}$. Thus, we have $\mathbf{P\Gamma} = \mathbf{PD\,G} + \mathbf{PH}$. Since rank ($\mathbf{PD}$) is no greater than rank ($\mathbf{D}$) and is generally smaller, this will likely lead us to find less than K factors in the transformed data.

[13] When we regress individual stock turnover on the five common factors, the cross-sectional standard deviation of the R^2s is 12.9%.

with those for returns, we find that turnover factors are just as important for explaining stock turnover as return factors are for explaining stock returns.

Comparing our findings with empirical results found in market microstructure studies by Chordia, Roll and Subrahmanyam (2000), our results suggests a *stronger* presence of commonality in turnover for most sample periods. Chordia, Roll and Subrahmanyam use high-frequency transaction data from a sample of 1169 stocks in 1992. They examine the common movement in market depth using value- and equal-weight indices, thus essentially using a one-factor model, and find the mean R^2 to be $<2\%$. Our results are more comparable to those of Hasbrouck and Seppi (2001), who use order flow data from the 30 Dow stocks during 1994 to study the common factors in stock prices and liquidity. They find the first three common factors explain about 20% of the variation in order flows. However, they do not provide a statistic for determining the number of factors in their factor model.

2.3 Monte Carlo simulations

While BN (2002) include a simulation study of the small-sample properties of their IC statistics, they use general data generating processes (DGP) that are not calibrated to typical stock return and turnover data. In this section, we will demonstrate that the IC estimates have reasonably good small-sample properties for standardized turnover.

We use two different DGPs in this article, both of which are designed to mimic the actual data as closely as possible. In particular, both approaches use bootstrapped samples of factor and beta estimates from the actual data similar to Jones (2001). However, we adapt the Jones (2001) bootstrap methodology by modeling factors through a Vector Autoregressive Regression (VAR) (1) process that preserves the persistence in turnover data. Further, both DGPs also preserve the correlation between firm-specific return and turnover components for each firm.

The two approaches differ only in how residuals are generated.[14] In the first 'full bootstrap' approach, we bootstrap the entire cross-section of (normalized) firm-specific components for both the return and the turnover series. Please note that this bootstrap process preserves the residual cross-sectional correlation of turnovers (and returns) among firms as well as heteroscedasticity, persistence, and the correlations between firm-specific return and turnover components for each firm.[15]

For the second "simulation" approach, we directly simulate return and turnover residuals so that there is no systematic factor in those residuals. However, we preserve that significant part of residual cross-sectional

[14] Please see the Appendix for a detailed description of the two DGPs.

[15] We are grateful to the referee for suggesting this bootstrap procedure.

correlation of both returns and turnover that exists within industries (see the Appendix for details). The main motivation of this alternative approach is to have a DGP that is transparent on the number of factors as well as residuals.

Panel A of Table 4 presents the frequency of the number of factors estimated for standardized turnover data over 1000 simulations using a balanced panel, with the results for the full bootstrap approach in panel A1 and for the simulation approach in panel A2. Conditional on the number of factors found in Table 3, each simulation involves the draw of a set of $N \times T$ individual turnover data for the corresponding sample period. For example, the 1997–2001 period involves 1000 draws of a 1385×252 panel of firm-level turnover. We find that the new procedure performs very well for returns in all periods and for both approaches.[16]

For turnover and using the full bootstrap approach, the procedure performs very well for 5 out of the 7 periods. For example, the first row of panel A1 of Table 4 shows that if the true number of factors is 5, the IC finds the right number of factors in 95% of the simulations, using parameters calibrated to resemble the data in the 1967–71 period. The mean of the estimated number of factors in the 1000 bootstrapped samples equals 4.95 with a standard deviation of 0.22, so the estimates are quite accurate. However, in two other periods the results are less accurate. In the 2nd period of 1972–1976, the procedure on average picks 0.83 more factors than the four factors used in generating the data, or choosing a model with one or two extra factors in 67.5% of cases. In the 7th period of 1997–2001, the procedure has chosen a model with one extra factor in 27.5% of cases. Thus, bootstrap simulations indicate that the accuracy of the BN procedure could be off by one or two factors in some cases, though they work well most of the time.

The simulation approach results of panel A2 also confirm this. The results are generally accurate in three out of the seven 5-year periods. However, in four periods, the criterion typically chooses one factor too few in 26.8–40% of cases. Therefore, in the presence of significant cross-sectional correlation, the criterion tends to underestimate the number of turnover factors by about 1, though still in a clear minority of simulations.

To understand the importance of standardizing turnover when estimating the number of required factors, we compare their small-sample properties to those for raw turnover. Using either bootstraps (panel B1 of Table 4) or simulations (panel B2), if the true number of factors in the first period is 16 for raw turnover, the IC finds the right number of factors only in 1% of the bootstraps or never in the simulations. The means of

[16] Results not reported to save space.

Table 4
Simulation test for the number of turnover factors extracted using the BN (2002) selection criterion

Panel A1: Full bootstrap approach.

	Standardized turnover–frequency (%) found in 1000 simulations									
	True K	1	2	3	4	5	6	7	Mean K	Std K
1967–1971	**5.0**	0	0	0	5	95	0	0	4.95	0.22
1972–1976	**4.0**	0	0	0	32.5	52	15.5	0	4.83	0.67
1977–1981	**5.0**	0	0	0	4.5	95.5	0	0	4.96	0.21
1982–1986	**4.0**	0	0	4	96	0	0	0	3.96	0.20
1987–1991	**3.0**	0	0	95	4	1	0	0	3.06	0.28
1992–1996	**4.0**	0	0	6.5	93.5	0	0	0	3.94	0.25
1997–2001	**4.0**	0	0	0	72.5	27.5	0	0	4.28	0.45

Panel A2: Simulation approach.

	Standardized turnover–frequency (%) found in 1000 simulation									
	True K	1	2	3	4	5	6	7	Mean K	Std K
1967–1971	**5.0**	0	0	0.7	38.9	60.4	0	0	4.60	0.50
1972–1976	**4.0**	0	0	15.3	84.7	0	0	0	3.85	0.36
1977–1981	**5.0**	0	0	0.5	38.2	61.3	0	0	4.61	0.50
1982–1986	**4.0**	0	0	26.8	73.2	0	0	0	3.73	0.44
1987–1991	**3.0**	0	0.4	99.6	0	0	0	0	3.00	0.06
1992–1996	**4.0**	0	0.3	40	59.7	0	0	0	3.60	0.50
1997–2001	**4.0**	0	0	10.4	89.6	0	0	0	3.90	0.31

Table 4
(Continued)

Panel B1: Full bootstrap approach based on RAW turnover.

	Raw turnover–frequency (%) found in 500 simulations													
	True K	5	6	7	8	9	10	11	12	13	14	15	16	Mean K	Std K
1967–1971	16	0	0	0	0	0	0	0.6	4.2	20.8	39	30	5.2	14.10	0.97
1972–1976	16	0	0	0	0	0	1.6	9.4	26.2	32.8	22.4	7.2	0.4	12.88	1.15
1977–1981	11	0	0	0.2	11.2	45.4	40.2	3	0	0	0	0	0	9.35	0.72
1982–1986	8	3.8	39.2	49	8	0	0	0	0	0	0	0	0	6.61	0.69
1987–1991	11	0.2	4.4	31	43.8	19.4	1.2	0	0	0	0	0	0	7.81	0.84
1992–1996	12	0	0	0.4	6	26.6	46.8	19.2	1	0	0	0	0	9.81	0.86
1997–2001	12	0	0	0	1	11	47	33.6	7.4	0	0	0	0	10.35	0.81

Panel B2: Simulation approach based on RAW turnover.

	Raw turnover–frequency (%) found in 500 simulations														
	True K	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean K	Std K
1967–1971	16	0	0	0	0	0	0.2	1.2	11.2	36.8	33.8	15.8	1	0	13.48	1.13
1972–1976	16	0	0	0	0	0	2.6	13.2	28.8	33.6	18.8	2.8	0.2	0	12.52	1.39
1977–1981	11	0	0	1	15.8	47	32.6	3.6	0	0	0	0	0	0	9.69	0.87
1982–1986	8	2	29.6	54.6	12.8	1	0	0	0	0	0	0	0	0	7.27	0.67
1987–1991	11	0.4	5.6	32.2	43.6	16.4	1.8	0	0	0	0	0	0	0	8.64	1.13
1992–1996	12	0	0	0	3.2	23.4	51.8	19.2	2.4	0	0	0	0	0	9.42	1.21
1997–2001	12	0	0	0.2	1	9.2	39.2	42.4	7.8	0.2	0	0	0	0	9.94	1.04

The table presents the frequency of the number of factors extracted using the BN (2002) selection criterion from the standardized turnover data over 1000 simulations. Each simulation involves the draw of a set of $N \times T$ individual turnover data with a VAR structure for the factors with bootstrapped betas. To draw residuals, we use two different approaches: a "full bootstrap" approach in panels A1 and B1 and a "simulation" approach in panels A2 and B2 (see the Appendix for details about the simulations). Panels A1 and A2 use the standardized turnover panel, whereas panels B1 and B2 use the raw, non standardized panel.

the estimated number of factors show a large downward bias compared to the number of factors found in the actual raw turnover panel. Thus, we conclude that, despite the presence of severe heteroscedasticity and the presence of time trends in turnover, the BN (2002) procedure has good small-sample properties for standardized turnover, though not for raw turnover.

2.4 Understanding the time-series properties of turnover components

An important question in the study of turnover is whether there is a systematic time trend. Evidence in Table 2 reports the number of systematic turnover factors for both the standardized data as well as for the detrended-and-standardized data. If there is indeed a time trend in systematic turnover, we expect the number of factors to be affected by detrending. Detrending could reduce the number of systematic factors by one if one of the systematic factors is a pure time trend. This is the case in four of seven periods; in the other three, the number of factors is not affected by detrending.

Next, we estimate whether there is a time trend in the factors extracted from the standardized (but not detrended) panel by regressing each factor on a constant, the lagged factor, and a time trend. Table 5 panel A presents the t-stats on the time coefficient of the Dicky–Fuller regression. We can see that, in all periods, most of the systematic factors have a statistically significant time trend, with different factors having either an up or down trend for the same period. These regressions include the lagged factor itself to ensure that the time trends are not merely an artifact of the large first-order autocorrelation. It is worth noting here that the signs of the trends are meaningless because the factors are only determined up to rotation.

In Table 5 panel B the pervasiveness of time trends in turnover is made clear. When regressing raw turnover on a constant and a time trend for each firm separately, we find a statistically significant time trend in 47% (for 1982–86) to 69% (for 1991–96) of firms.

Finally, taking out the systematic factors effectively removes most occurrences of time trends. When regressing each firm's idiosyncratic turnover on a constant and a time trend, in five of seven periods all firms have statistically insignificant trend coefficients. In the other two periods, only 10 and 7% of firms show evidence of a statistically significant time trend in idiosyncratic turnover.

Thus, the turnover decomposition allows us to break persistent turnover into two components: a persistent systematic component and a stationary idiosyncratic component. This not only gives us a better understanding of the dynamics of turnover at the firm level but also indicates that decomposition could be a more effective approach in removing persistence

Table 5
Time trend test for turnover components

Panel A: Test of trend in systematic turnover

Period	T-Statistics for the trend coefficients				
1967–1971	1.59	7.52	−5.97	−2.74	0.68
1972–1976	−1.87	−5.23	−6.43	5.89	—
1977–1981	3.08	3.58	−3.16	−3.63	4.73
1982–1986	1.68	4.32	8.28	−2.90	—
1987–1991	−3.17	−7.81	5.07	—	—
1992–1996	−10.99	−5.35	7.69	1.80	—
1997–2001	−9.30	2.05	−4.94	5.75	

Panel B: Test of trend in raw as well as idiosyncratic turnover

Period	Obs.	% with trend in raw τ_{it} (%)	% with trend in idio. τ_{it} (%)
1967–1971	1586	60.21	0.00
1972–1976	1912	57.74	10.46
1977–1981	1753	49.46	0.00
1982–1986	1514	47.42	0.00
1987–1991	1399	56.25	7.15
1992–1996	1528	68.78	0.00
1997–2001	1385	64.40	0.00

In the table the systematic factors are computed directly using the turnover data in levels. The computation is based on a balanced panel with standardized turnover. Panel A reports the t-Statistics of the trend coefficient for each systematic turnover factors. All those trend coefficients are estimated in a regression that *also* includes the once lagged systematic turnover factor series and a constant. Panel B reports the occurrence of statistically significant trend coefficients in both raw and idiosyncratic firm-level turnover, when regressed on a constant and a time trend only
"% with trend in raw τ_{it}" stands for percentage of the raw turnover series with statistically significant trend coefficients.
"% with trend in idio. τ_{it}" stands for percentage of the idiosyncratic turnover series with statistically significant trend coefficients.

in firm turnover than alternative approaches. The results are presented in Figure 1.

Figure 1 displays the average autocorrelations for the first 10 lags for raw and idiosyncratic turnover of individual stocks as well as for several alternative approaches to reduce the persistence. For all autocorrelations in Figure 1, we average first across stocks per period, and then across the seven 5-year periods. For comparison, we also report the autocorrelation of [raw and GRT-detrended based on the approach designed by Gallant, Rossi, and Tauchen (1992)] turnover of the equal-weighted index by LW. The average autocorrelations of raw turnover of individual stocks display persistence similar to that of the index, but are smaller in magnitude. Linear detrending removes some of the autocorrelation, but significant autocorrelation remains (10%) even after the 7th lag. Furthermore, two popular approaches of removing market turnover (using either "excess turnover" by taking the difference between individual turnover and market turnover or "residual turnover" computed by fitting a market model using value weighted (VW) turnover as the market turnover factor) help little and

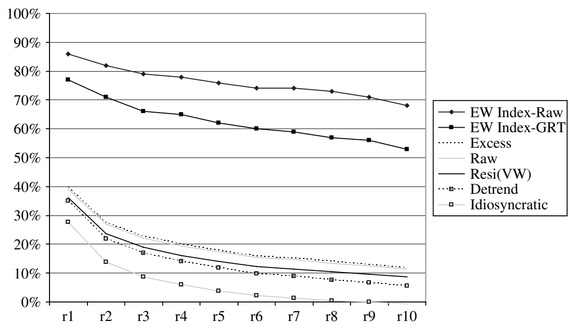

Figure 1
Average autocorrelations for equally weighted (EW) turnover index (raw and GRT-detrended) and individual turnovers (excess, raw, detrended, residual (VW market model), and idiosyncratic (multifactor model))
The figure plots the cross-sectional averages of the first ten lags of the autocorrelation (r1 is the first lag, r2 is the second lag, etc.) of seven different measures of firm-specific turnover (see their descriptions in the text), averaging across all sample periods.

may actually worsen the autocorrelation patterns of individual turnover series.[17]

Removing the systematic components using the BN decomposition method, however, significantly reduces autocorrelation in the idiosyncratic turnover series, leading to the lowest average correlations across all methodologies compared in Figure 1. For example, the autocorrelation drops to 5% after the 5th lag. Thus, shocks to firm-level (i.e., idiosyncratic) turnover die out in four weeks or so, much *faster* than what is suggested by the strong persistence for raw turnover or at the index level. This weak persistence is a desirable property for idiosyncratic turnover in time-series analysis.

Jones (2001) introduces a heteroscedastic factor analysis (HFA) for extracting factors, which allows for time-varying volatility in returns. While his simulation shows that an HFA may sometimes improve the accuracy of factor estimates, the methodology depends on the strong assumption that the idiosyncratic terms are uncorrelated over time. As shown in Figure 1, this assumption is seriously violated for turnover data. While the principal components approach of CK (1988) may not be as accurate as HFA in small samples, BN (2002) show it is nonetheless consistent in the presence of autocorrelation and heteroscedasticity. The simulation results presented in this article also show that the CK approach is quite accurate in estimating the number of factors in turnover. In our return application we also use this approach, as there is also strong evidence of return

[17] In section 5, we further show that these turnover measures tend to significantly understate the price reversals of stock trading. As a result, a multifactor model is needed in estimating "abnormal" trading volume as well as its impact on asset prices.

autocorrelation at the firm level, which has contributed to momentum trading strategies [see, e.g., Conrad, Hammed, and Niden (1994)].[18]

3. Testing the LW Model

Although turnover data has long been available, researchers in finance have concentrated on "asset pricing" while paying scant attention to "asset quantity". In their seminal paper, LW (2000) attempt to address this imbalance by deriving theoretically the relationship between return and turnover. They demonstrate that, in much the same way multifactor models such as the CAPM and ICAPM have guided empirical investigations of the time–series and cross-sectional properties of assets returns, the volume implications of these models provide similar guidelines for investigating the behavior of volume. LW (2003) further establish a tighter theoretical link between return and volume by modeling heterogeneous investors who hedge market risk and changing market conditions by trading a market portfolio and a hedging portfolio. Below, we provide a new test of LW (2000) and shed some further light on the results of LW (2005). All return and turnover factors used here are from the standardized balanced panels.

3.1 The LW (2000) model

Following LW, we assume that returns are generated by the following approximate K'-factor model:[19]

$$R_{jt} = E_t(R_{jt}) + f_{1t}\beta_{j1} + \ldots + f_{K't}\beta_{jK'} + e_{jt}\, j = 1, \ldots, N; t = 1, \ldots, T. \tag{5}$$

where $f_t' = (f_{1t}, \ldots, f_{K't})$ is a vector of unobservable pervasive shocks, $(\beta_{j1}, \ldots, \beta_{jK'})$ is a vector of factor loadings that are constant over the sample period, and e_{jt} represents an idiosyncratic risk specific to asset j at time t. We also assume e_{jt} has mean zero and is orthogonal to f_{kt}. As discussed in Chamberlain (1983), the above economy implies the following linear pricing relationship if there exist K well-diversified portfolios:[20]

$$E_t(R_{jt}) = r_{ft} + \lambda_{1t}\beta_{j1} + \ldots + \lambda_{K't}\beta_{jK'} \tag{6}$$

[18] It would be interesting to compare the accuracy of extracted factors of the HFA and the CK approaches while allowing for serial correlation in idiosyncratic return and turnover. However, that is beyond the scope of this article.

[19] To avoid confusion with the K-factor turnover model, we will use K' to indicate the number of factors in the return model.

[20] Connor (1984) derived the same result under the condition that the supplies of the assets are well diversified. To derive the consistency result of the BN statistic for the number of factors in the return model, some additional regularity conditions are imposed [see BN (2004)].

where $(\lambda_{1t}, \ldots, \lambda_{K't})$ is a vector of risk premiums corresponding to the pervasive shocks $(f_{1t}, \ldots, f_{K't})$, and r_{ft} is the return on a riskless asset.

Under the presence of K' well-diversified portfolios, Chamberlain (1983) shows that the above asset-pricing model satisfies K'-fund separation. Under the assumptions that these K' portfolios (or return factor-mimicking portfolios) are constant over time and the amount of trading in them is small for all investors, LW point out that investors would hold the K' portfolios for investment and would trade these portfolios for portfolio rebalancing or hedging. Thus, they derive the proposition that the turnover of each stock has an approximate K' factor just like Equation (5). Their insight is that in equilibrium, well-diversified investors hold the K' funds and just trade them to hedge the systematic risk mimicked by the K' factor portfolios. This implies that 100% of the variance of the common turnover factors can be explained by turnovers on return factor portfolios. Simply put, the turnover model is linked with the return model, as the turnover factors are nothing but the turnover on the K' return factors. Therefore, turnover also has a K'-factor structure, just like excess returns, that is, $K = K'$.

Although LW does not formally address the issue of idiosyncratic turnover, one way to justify its existence is the presence of noise traders in the economy who hold non-diversified portfolios. Trading on either information or speculation, their trades affect neither asset pricing nor the systematic turnover. In this case, by identifying idiosyncratic turnover at the firm level, one may learn about trading related to firm-specific information as well as firm-specific speculation [see, e.g., Michaely and Vila (1996) on trading volume in the presence of private valuation].

3.2 The importance of mutual fund separation for turnover

As we noted in the above subsection, mutual fund separation implies that the turnover factors are nothing but the turnover on the return factors. This gives us a new way of measuring the performance of asset-pricing models by the ability of turnover on their return factors to explain individual stock turnover. In this subsection, we compare the performance of three asset-pricing models: (a) CAPM, (b) the Fama and French (1993) three-factor model, and (c) the LW (2000) model of (5).

Panel A of Table 6 reports the average R^2 from regressing individual stock turnover on the CAPM market turnover. The CAPM market turnover is the average turnover across all individual firms, using value weights for all firms; that is, it is turnover in the VW) market portfolio.[21]

[21] An earlier version of the article reported results for equally weighted (EQ) portfolios and the results are quite similar to VW portfolios.

Table 6
Explaining stock turnover by turnovers on return factors

Time-series regression of stock turnover on turnovers of return factors

	CAPM		Fama & French		Lo and Wang		Turnover factors
Time	Average stock R^2 (%)	Performance (%)	Average stock R^2 (%)	Performance (%)	Average stock R^2 (%)	Performance (%)	Average stock R^2 (%)
67–71	3.4	13.3	12.1	47.5	15.6	61.3	25.5
72–76	8.4	31.3	19.8	73.9	21.0	78.4	26.7
77–81	9.0	38.9	12.9	55.7	13.7	59.0	23.1
82–86	6.8	32.1	9.4	44.8	14.8	70.6	21.0
87–92	8.3	44.9	10.4	56.6	12.2	66.2	18.4
92–96	4.0	25.8	5.7	37.0	6.8	44.0	15.5
97–01	5.2	25.4	7.3	36.0	14.5	71.3	20.4
Median		31.3		47.5		66.2	

This table provides the average R^2 from time-series regression of individual stock turnover on turnovers of return factors derived from three asset-pricing models: (a) CAPM, (b) Fama and French (1993), (c) Lo and Wang (2000). Performance is computed as the average R^2 of the model divided by the average R^2 obtained from regressing stock turnover on extracted turnover factors

The second column reports that the CAPM market turnover typically explains 3.4–9% of individual stock turnover over the seven sample periods. In comparison, the principal components of turnover (turnover factors) explain on average 15.5–26.7% of individual stock turnover. Table 6 also reports the relative performance of the CAPM market turnover. We measure performance by computing the average R^2 in the second column divided by the average R^2 in the last column, which is obtained from regressing individual stock turnover on their principal components.[22] This is because, if LW holds, we expect the variation of turnover on the market portfolio to capture all variation due to common turnover factors. We can see that the CAPM market turnover here captures on average 31.3% of all systematic turnover in individual stock trading.

Table 6 then reports the average R^2 from regressing individual stock turnover on turnovers of the three Fama and French (1993, FF hereafter) return factors. The three FF portfolios are a VW market portfolio plus 2 VW FF portfolios: that is, "small minus big" (SMB) and "high minus low" (HML). We define portfolio turnover as in LW (2000) by computing a value-weighted average of individual stock turnover, with the weights being the absolute value of the portfolio weights. We can see that the FF factor turnovers typically explain 5.7–19.8% of individual stock turnover. We can also see that the FF factor turnovers capture about 36.0–73.9% of all systematic turnover in individual stock trading. Therefore, the two additional FF factor turnovers increase the median performance by 16% in capturing systematic turnover.

Table 6 also reports the average R^2 from regressing individual stock turnover on turnovers of return factors extracted from the LW model of (5). The return factor portfolios here are determined by the principal component factors extracted from the return data, with portfolio weights given by the scaled eigenvector. The number of return factors used is reported in the bottom panel of Table 3. Thus, the corresponding turnover for each factor portfolio is simply the weighted average of individual turnover, using again the absolute value of scaled eigenvectors as weights. We can see that the LW turnovers typically explain 6.8–21% of individual stock turnover, while they capture on average 66.2% of all systematic turnover in individual stock trading. Therefore, the LW turnovers on return factor portfolios outperform those of FF in capturing systematic turnover by another 19%. It is worth noting that this outperformance is achieved with one less factor (except the last period when the number of

[22] An alternative measure of performance is the average R^2 of regressing systematic turnover on the turnover of the CAPM market portfolio. We find that the turnover of the market portfolio on average captures about 14.0–46.3% of systematic turnover in individual stock trading over various periods. Therefore, the two performance measures are similar. Parallel results also exist for the turnover of the FF and LW return factors.

factors is the same for both).[23,24] Ideally, if the LW mutual fund separation holds, we would expect the LW turnovers to capture 100% of individual stock turnover. Thus, 38.8% (= 100 − 66.2) can be thought of as the distance to the true LW return–turnover model of (1) and (5).[25]

The results in Table 6 establish for the first time the importance of portfolio rebalancing due to systematic risk in explaining stock turnover. Our results indicate that on average, it accounts for 66% of all systematic turnover (time) variation. Thus, portfolio rebalancing is clearly an essential motive for stock trading. Our results complement those of LW (2003) and earlier empirical studies, which emphasize the role of turnover in helping to understand asset pricing, whereas here we have examined the role of asset pricing in determining turnover.

As noted by LW, mutual fund separation has become the workhorse of modern investment management. While the assumptions of mutual fund separation are known to be violated in practice, this class of asset-pricing models, such as CAPM and Arbitrage Pricing Theory (APT), are still widely used for quantifying the trade-off between risk and expected return in financial markets. Our study of individual stock trading demonstrates that mutual fund separation has yielded a useful approximation for quantifying the time and cross-sectional variation of trading volume. Our finding that systematic risk and systematic trading are strongly related is quite good news for existing asset-pricing theory. The alternative would suggest the presence of important weaknesses in the theory of modern investment management, implying that existing asset-pricing models have little relevance for trading activity.

Moreover, while mutual fund separation captures only simple trading motives such as portfolio rebalancing and diversification, accurately accounting for them can help us better understand *other* motives for trading such as asymmetric information, speculation, and manipulation. For example, in studying the market reaction to price manipulation,

[23] It is worth noting that the turnovers of the return factors of the CAPM, FF, and LW models also provide an intuitive explanation of the principal components of turnover. For example, we find that the turnover of the CAPM market portfolio can explain between 11.7 and 62.6% of the time variation of the first principal component, suggesting that trading in the market portfolio explains a major component of systematic turnover. Not surprisingly, the turnover of the LW factors generally outperforms others in explaining the various principal components of turnover. We would also like to note that, while the principal components methods explain the biggest fraction of systematic turnover, the downside is that the factors are less transparent and less intuitive than the FF factors.

[24] Additionally, these $R^{2'}$s are not driven by any statistical artifact. For example, if we bootstrap return and turnover data independently of each other, the resulting R^2 of the turnover of the LW factors is approximately zero.

[25] In an earlier version of the article, we also examine to what extent mutual fund separation can explain the cross-sectional variation of trading volume. We discover that sensitivity to turnovers of return factors again helps explain the cross-sectional variation of trading volume (On average, it accounts for 64% of all systematic turnover variation in the cross-section). Thus, the evidence further suggests a close link between return factors and turnover factors as implied by LW.

Aggarwal and Wu (2006) use a simple market model to define abnormal trading activity that is associated with these events [also see Jiang et al. (2005)]. Nagel (2005) also examines the impact of trading style, such as firm size or momentum, on the cross-section of trading volume.

3.3 Test on the equality of return and turnover factors

Since our study in section 2 2.3 documents some apparent differences in the number of return and turnover factors, we now formally examine the Lo-Wang duo-factor model's hypothesis that the number of return and turnover factors should be equal. Table 7 presents the Type I and Type II error estimates of a formal test. The error estimates are based on 1000 simulations for each period, where each simulation involves the draw of a set of $N \times T$ individual return and turnover data. We use standardized turnover and balanced panels as before, though now we only use the simulation approach because we need to condition on the "right" number of factors (see Section 2.3 and the Appendix for details).

For the type I error estimates (see panel A), we set the true numbers of return and turnover factors equal to either the number of return factors (K) found each period or the number of turnover factors (K') found each period. In both cases, the specified difference between the number of factors for returns and turnover in the simulations, $K - K'$, equals zero. For the type II error estimates of panel C we set the true numbers of return and turnover factors equal to those found in the data as reported in Table 2.

Overall, our simulation study indicates a clear and unambiguous rejection of the null hypothesis in all periods, which states that the number of systematic factors in returns and turnover is the same. In all seven periods, if the simulated numbers of factors are the same for return and turnover and set equal to the number of factors as found in the data for returns, the type I error is essentially zero. If the simulated numbers of factors are set equal to the number of factors as found for turnover, then the probability that the IC finds a difference equal to those as estimated in the actual data is at most 4.7% (see the right side of panel A in Table 7, fifth period). Panel B of Table 7 shows that the IC has almost no Type II errors conditional on the actual number of factors found in the data. The probability of accepting the null of same factors, while not true, is at most 3.5% for the fourth period.

The rejection of the "same number of factors" restriction is not surprising, as the turnover factor model was derived based on K-fund separation, implying common mimicking factor portfolios held by all investors. As the previous subsection shows, turnovers on return factor portfolios explain at most 73% of all systematic turnover. The remainder could come from other motives of trading such as asymmetric information, speculation, and manipulation. If a large number of investors use private information to speculate on small or internet stocks, for example, this

Table 7
Simulation results on the test of equal number of factors

Panel A: Type I error estimates

Time period	Specified $K-K'$	No. of factors equals K (returns)					No. of factors equals K' (turnover)					
		−3	−2	−1	0	1	−3	−2	−1	0	1	2
67–71	0	**0**	0	0	100	0	**0**	6.3	46.8	39.2	7.5	0.2
72–76	0	0	**0**	0	100	0	0	**0.6**	20	66.2	13	0.2
77–81	0	**0**	0	0	100	0	**0**	0.1	54.9	40.9	4.1	0
82–86	0	0	**0**	0	100	0	0	**0**	11.2	67.2	21.4	0.2
87–92	0	0	0	**0**	100	0	0	0	**4.7**	94.6	0.7	0
92–96	0	0	**0**	0	100	0	0	**3.9**	37.7	43.8	14.5	0.1
97–01	0	0	0	**0**	99.5	0.5	0	0	**2.1**	88.2	9.7	0

Panel B: Type II error estimates

Time period	Specified $K-K'$	−4	−3	−2	−1	0	1
67–71	−3	0	95	5	0	0	0
72–76	−2	15	47.5	36.5	1	0	0
77–81	−3	0	95.5	4.5	0	0	0
82–86	−2	0	0	9.5	87	3.5	0
87–92	−1	0	1	4	95	0	0
92–96	−2	0	0	93.5	6.5	0	0
97–01	−1	0	0	27.5	72.5	0	0

The table presents the Type I (panel A) and Type II (panel B) error estimates for testing the difference between the number of return factors and the number of turnover factors based on 1000 simulations for each period. Each simulation involves the draw of a set of $N \times T$ individual return and turnover data with a VAR structure for the factors with bootstrapped betas and uses the "simulation" approach for drawing residual (see the Appendix for details about the simulations). The bold-faced numbers give the probability percentage of the error. The computation is based on a balanced panel with standardized turnover. K equals the number of factors found each period for returns using the actual data, and K' equals the number of factors for turnover (see Table 2). For panel A, the number of factors of returns and turnover is set equal, either to K or to K'

could lead to a violation of K-fund separation and the rejection of the duo-factor model. For example, Llorente et al. (2002) find that small firms tend to have high trading volume associated with asymmetric information. Thus, our result suggests that further theoretical work on turnover may be needed to unify stock price and trading volume under a multifactor assetpricing-trading framework.

Another possible explanation could be sample selection. Since our sample excludes bonds and NASDAQ stocks, our return sample may not be able to reflect all systematic risks in the economy. For example, FF (1993) find that, with stocks, only three factors seem sufficient to explain their cross-section, but five are needed when bonds are included in asset-pricing studies. To the extent that new technology and changing interest rates may have a systematic impact on the return of assets outside our sample, investors may need to rebalance their position on all assets. As a result, we may observe systematic changes in turnover but fail to detect their impact on stock returns in our sample.

4. Measuring Price Reversal as a Result of Stock Trading

In this section, a direct application of our turnover decomposition methodology demonstrates that failure to fully decompose turnover may lead to an underestimation of the price reversal due to stock trading. Following Pastor and Stambaugh (2003), we measure price reversal by running the following regression,[26]

$$r^e_{i,t+1} = \theta_i + \phi_i r^e_{i,t} + \gamma_i sign(r^e_{i,t})\tau^e_{i,t} + \varepsilon_{i,t+1}, \tag{7}$$

where $r^e_{i,t}$ is the (excess or otherwise) return on stock i and $\tau^e_{i,t}$ is a measure of the firm-specific turnover for stock i. Here, γ_i measures the price reversal due to the impact of order flow for stock i, constructed by using volume signed by the contemporaneous return.

Regression (7) estimates the average effect that week t trading has on the stock return in week $t+1$. Campbell, Grossman, and Wang (1993) show that a less liquid market would have a more negative γ_i due to the larger return reversal resulting from the larger price reaction of trading. Intuitively, the measure of trading in (7) can be interpreted as signed order flow, and greater liquidity is interpreted as a weaker tendency of trading in the direction of returns in t to be followed by opposite price changes in $t+1$.

Llorente et al. (2002) show that this volume–return relation may also reflect firm-level information asymmetry. While γ_i generally tends to be negative, they demonstrate it could be positive if the price impact of information trading dominates the liquidity effect. Therefore, market participants often pay close attention to the volume of trading to help distinguish portfolio-rebalancing trades from speculative trades based on private information. Following Pastor and Stambaugh (2003), we simply call γ_i a measure of liquidity, where a larger (or less negative) γ_i means less liquidity or a smaller price reversal due to the impact of trading.

To gauge the impact of different return and turnover measures on γ_i estimates, we set $r^e_{i,t}$ equal to either the excess return over the market, $r^e_{i,t} = r_{i,t} - r_{m,t}$, or to the idiosyncratic return in the multifactor model of (5), e_{jt}. We compare five different measures of firm turnover as in Section 3.5:

- (A) $\tau^e_{i,t}$ equal to the demeaned raw turnover for stock i during week t,
- (B) $\tau^e_{i,t}$ equal to $\tau_{i,t}$—τ^{vw}_{mt}, or the turnover in excess of the value-weighted market turnover,
- (C) $\tau^e_{i,t}$ equal to detrended turnover for stock i,
- (D) $\tau^e_{i,t}$ (resi(VW)) equal to the residual turnover in the one-factor market model of LW (2000), with the value-weighted market turnover used as the sole factor

[26] Using $r_{i,t}$ rather than $r^e_{i,t}$ as the first regressor in Equation (10), as in Pastor and Stambaugh (2003), gives similar results.

(E) $\tau^e_{i,t}$ equal to $\zeta_{i,t}$, the idiosyncratic turnover in the multifactor model of (1).

Table 8 Panel A presents the cross-sectional average of the estimates of γ_i for the seven sample periods using excess return over the market, $r^e_{i,t} = r_{i,t} - r_{m,t}$, and the five different measures of turnover. It also provides the t-tests for the hypotheses that the mean of γ_i under specifications A, B, C, or D equals the mean of γ_i under E.

Table 8
Average estimates of price reversal using five different measures of turnover

Panel A: $r^e_{i,t+1} = r_{i,t+1} - r_{m,t+1}$

	A	B	C	D	E
Turnover	Raw	Excess	Detrended	Resi.(VW)	Idio
67–71	−0.023	−0.105	−0.006	−0.014	−0.012
72–76	−0.739	−0.591	−0.792	−0.760	−0.743
77–81	−0.232	−0.215	−0.271	−0.235	−0.415
82–86	−0.146	−0.165	−0.137	−0.155	−0.173
87–91	−0.139	−0.180	−0.145	−0.141	−0.197
92–96	−0.073	−0.092	−0.078	−0.075	−0.100
97–01	−0.082	−0.116	−0.081	−0.085	−0.138
Mean Diff.	0.049	0.045	0.038	0.045	
t-Stat.	2.027	1.201	1.727	1.811	

Panel B: $r^e_{i,t+1} = e_{i,t+1} = r_{i,t+1} - \sum B_{ik}(\lambda_{kt+1} + f_{kt+1})$

	A	B	C	D	E
Turnover	Raw	Excess	Detrended	Resi.(VW)	Idio.
67–71	−0.023	−0.076	−0.022	−0.035	−0.052
72–76	−0.217	−0.231	−0.234	−0.226	−0.309
77–81	−0.087	−0.123	−0.100	−0.087	−0.164
82–86	−0.059	−0.098	−0.062	−0.066	−0.092
87–91	−0.044	−0.071	−0.058	−0.047	−0.081
92–96	−0.018	−0.024	−0.027	−0.020	−0.039
97–01	−0.030	−0.056	−0.036	−0.032	−0.064
Mean Diff.	0.046	0.017	0.037	0.041	
t-Stat.	4.516	1.386	4.307	3.986	

This table provides the cross-sectional averages of the price reversal (γ_i) estimates for the seven 5-year sample periods using two different measures of returns (firm return in excess of the market return in panel A and idiosyncratic returns in panel B) and five different measures of turnover (see description below panel B). Further, we report the average across periods of the difference between the mean of γ_i under specification A, B, C, D and the mean of γ_i under specification E, plus its t-statistic under the assumption that this difference is independent across periods.

$$r^e_{i,t+1} = \theta_i + \phi_i r^e_{i,t} + \gamma_i sign(r^e_{i,t})\tau^e_{i,t} + \varepsilon_{i,t+1}$$

The five different measures of firm turnover as defined as follows:
(A) $\tau^e_{i,t}$ equal to the demeaned raw turnover for stock i during week t
(B) $\tau^e_{i,t}$ equal to $\tau_{i,t} - \tau^{vw}_{mt}$ turnover in excess of the value-weighted market turnover
(C) $\tau^e_{i,t}$ equal to detrended turnover for stock i
(D) $\tau^e_{i,t}$ (Resi.(VW)) equal to the residual turnover in one-factor model of LW(2002). We use the value-weighted market turnover as the sole factor.
(E) $\tau^e_{i,t}$ equal to $\xi_{i,t}$, the idiosyncratic turnover in the multifactor model of (1).

For idiosyncratic turnover (E), we can see that the average price reversal was the smallest during 1967–71, then liquidity dropped sharply in 1972–76, but has been improving ever since. For example, a 10% increase in weekly signed turnover would, on average, cause a 7.43% reversal in excess return $r_{i,t}^{e}$ over 1972–76 but only 1.38% over 1997–2001. If we compare the average estimates of γ_i based on E with A, B, C, and D, we find that all four alternative measures tend to provide less negative estimates (or those smaller in absolute value) of γ_i, suggesting a smaller price reversal (the exception being the first period). For example, the average γ_i was -0.591 under excess turnover B compared to -0.743 for idiosyncratic turnover (E) for 1971–76. This suggests that, using market turnover to compute excess turnover may understate the price reversal by as much as 20–40% if one is interested in *firm-specific* trading volume. The same holds for using the residual from a market turnover factor model (D), as well as for raw (A) and detrended (C) turnover. The reason for the smaller price reversal is that all four of these turnover measures still contain portions of systematic turnover. It is not surprising that the price reaction would be smaller for systematic trading, presumably because this type of trading is not based on private information, but instead is due to risk sharing.

In panel B of Table 8, we report the cross-sectional averages of γ_i using idiosyncratic (rather than excess) returns, again for the five different measures of turnover. We can see that a 10% increase in weekly signed turnover would on average cause a 3.09% reversal in idiosyncratic return e_{jt} over 1972–76, but only 0.64% over 1997–2001. The estimates of the mean γ_i are generally smaller in absolute value, but the results are otherwise similar to those using excess returns.

Next, for each of the first four turnover measures (A, B, C, and D), we formally test whether the respective estimate of the mean of γ_i, averaged across periods, is different from the average estimate using idiosyncratic turnover (E). We assume that the differences are independent across the seven 5-year periods, and compute their standard errors and t-stats. The mean differences and t-stats are provided in the bottom of panel A. The mean differences are positive for all four turnover measures. For raw, detrended, and residual turnover (A, C, and D), this difference is highly statistically significant using idiosyncratic returns (panel B), and more marginally statistically significant using excess returns (panel A). Only the results for excess turnover show no significant difference from the results using idiosyncratic turnover.

Therefore, we conclude that several commonly used turnover measures may significantly understate the price reversals of stock trading. As a result, a multifactor model is needed in estimating “abnormal” trading volume as well as its impact on asset prices.

5. Conclusion

This article employs two statistical procedures developed by BN (2002, 2004) to estimate an approximate factor model for turnover and test for nonstationarity. We document the presence of severe heteroscedasticity and nonstationarity in turnover data of individual stocks. We find the BN (2002) information criterion works well for raw returns but not for raw turnover in estimating the required number of systematic factors. However, a WLS-type approach of standardizing turnover is very effective in dealing with the problems in turnover data, such as the presence of correlation and heteroscedasticity at both time and cross-section dimensions.

Using this approach, we provide a new test of the return–turnover model developed by LW (2002). An important element of our methodology is the use of data from individual stocks rather than from beta-sorted portfolios. In particular, by exploiting the advantage of a large cross-section of individual stocks, we are able to break the persistent turnover into two components: a persistent systematic component and a stationary idiosyncratic component. This not only gives us a better understanding of the dynamics of turnover at the firm level but also indicates that decomposition could be a more effective approach in removing persistence in firm turnover compared to alternative approaches.

On the basis of a balanced panel of return and turnover data from NYSE and AMEX stocks, we find several interesting results. First, systematic turnover factors are quite useful in explaining the variation of turnover for a large panel data set. There are four or five systematic factors driving stock turnover, depending on the period. These common factors explain 15 to 26% of trading volume, similar to the proportion of systematic variation in individual stock returns. Second, our results indicate that portfolio rebalancing due to systematic risk is really a very important motive for stock trading. Mutual fund separation yields a useful approximation for quantifying the time and cross-sectional variation of trading volume. Third, we show that several commonly used turnover measures may significantly underestimate the effect of stock trading.

In addition to weekly data, we have also examined return and turnover data using monthly time series for NYSE and AMEX stocks. The results from using monthly data confirm our analysis using the weekly time series. The BN statistics are quite robust and consistent in estimating the number of factors in monthly data for the balanced panel using the standardized turnover. We find four or five systematic factors driving firm turnover and, on average, 36.5% of firm turnover is determined by common turnover factors. Further, we discover that there are two or three systematic factors driving excess returns. Finally, idiosyncratic risk, on average, explains 32% of idiosyncratic turnover (detailed results are available on request). This suggests that there is indeed an "inextricable link" between trading activity and return volatility at the firm level.

There are several issues that remain to be examined. If the duo-factor model provides a parsimonious description of weekly data, it would be interesting to know whether it would work equally well on higher-frequency data. Second, our decomposition can provide firm-specific parts of turnover related to price momentum, and therefore could potentially be used to identify different firm-specific stages of momentum-value cycles, as in Lee and Swaminathan (2000). Finally, if firm-level asymmetric information drives idiosyncratic volume and risk, then by using the return and turnover decomposition developed in this article we may obtain a proxy for measuring the degree of information asymmetry across stocks and thus be able to evaluate the impact of price discovery risk on asset pricing [see O'Hara, 2003)].

Appendix A:

This Appendix briefly discusses the two bootstrap/simulation procedures used to test LW (2000). The two approaches differ only in how residuals are generated. In the first "full bootstrap" approach, we bootstrap the entire cross-section of the (normalized) firm-specific components for both the return and the turnover series. For the second "simulation" approach, we simulate residuals according to a one-industry factor model. Below, we give the details for the "simulation" approach only.

Given estimates of the $T \times K'$ matrix $\mathbf{F}$ of factor realization for returns, we sample (with replacement) T rows of $\mathbf{F}$ to use as the true factors in the simulations. Let $\mathbf{F}_i$ denote the ith bootstrap draw of the factor matrix. The factor betas assumed in the DGP are bootstrap samples of the least squares estimates of the betas from the actual data and we assume them to be constant over time. Denoting $\mathbf{B}$ to be the $N \times K'$ matrix of OLS estimates of the factor betas from real data, we follow Jones (2001) by drawing with replacement N rows of the $\mathbf{B}$ matrix to use as the true betas in the simulations. We then draw the corresponding elements of the $N \times N$ diagonal matrix $\mathbf{\Omega}$, whose (j, j) element is the unconditional sample variance of the residual of stock j. We denote $\mathbf{B}_i$ to be the ith bootstrap draw of the beta matrix and $\mathbf{\Omega}_i$ the corresponding draw of $\mathbf{\Omega}$. As a result, the $N \times T$ matrix of simulated excess returns $\mathbf{R}_i$ will then be generated by the equation

$$\mathbf{R}_i = \mathbf{B}_i \mathbf{F}_i + \mathbf{\Psi}_i * \mathbf{E}_i \tag{A1}$$

where $\mathbf{\Psi}_i$ is the Cholesky-decomposition factor of $\mathbf{\Omega}_i$ and $\mathbf{E}_i$ is an $N \times T$ matrix of return residuals. Here, we assume all alphas to be zero.

Similarly, given estimates of the $T \times K$ matrix $\mathbf{G}$ of factor realizations for normalized turnover, we first estimate a first-order VAR on the estimated factor realizations and use the VAR transition matrix and simulated VAR errors to obtain $T \times K$-factor innovations, V. We then draw T rows of V to use as the true factor innovations to form factor matrix G in the simulations, maintaining the same order as returns. Let $\mathbf{G}_i$ denote the ith bootstrap draw of the factor matrix. The factor betas assumed in the DGP are the bootstrap samples of the least squares estimates of the turnover betas from the actual data, which are assumed to be constant over time. Denoting $\mathbf{D}$ to be the $N \times K$ matrix of OLS estimates of the turnover betas from real data, we draw the same N rows as returns of the $\mathbf{D}$ matrix that we use as the true betas in the simulations. The $N \times T$ matrix of simulated turnover $\mathbf{\Gamma}_i$ will then be generated by the equation

$$\mathbf{\Gamma}_i = \mathbf{D}_i\mathbf{G}_i + \mathrm{H}_i \tag{A2}$$

where H_i is an $N \times T$ matrix of turnover residuals.

For simulating residuals in $\mathbf{E}$ and $\mathbf{H}$ using the "simulation" approach, we first assign each firm to one of 48 industry groups according to the FF industry classifications.[27] Next, we calculate equally weighted portfolios of the residual returns and turnover (using the actual number of factors as found in the data to find the residuals), and regress each firm's residual return (turnover) series on its appropriate, corresponding industry residual return (turnover) factor, resulting in an estimate of an "industry beta" for each firm's return and turnover.[28] The industry return and turnover factors are then simulated independently using a AR(3) process estimated from the data (i.e., the equally weighted industry return and turnover portfolios). Finally, the residuals are calculated as the product of a bootstrapped industry beta (preserving the link between return and turnover betas for the same stock) times the appropriate industry factor plus an independent, normally distributed error. This DGP preserves that significant part of residual cross-sectional correlation of both returns and turnover that exists within industries. The main motivation of this alternative approach is to have a DGP that is transparent on the number of factors as well as residuals.

References

Aggarwal, R. K., and G. Wu. 2006. Stock Market Manipulations *Journal of Business* 79:1915–53.

Bai, J., and S. Ng. 2002. Determining the Number of Factors in Approximate Factor Models. *Econometrica* 70:191–222.

Bai, J., and S. Ng. 2004. A PANIC Attack on Unit Roots and Cointegration. *Econometrica* 72:1127–77.

Berk, J. 2000. Sorting Out Sorts. *Journal of Finance* 55:407–27.

Brennan, M., T. Chordia, and A. Subrahmanyam. 1998. Alternative Factor Specifications, Security Characteristics and the Cross-Section of Expected Stock Returns. *Journal of Financial Economics* 49(3):345–73.

Campbell, J. Y., S. J. Grossman, and J. Wang. 1993. Trading Volume and Serial Correlation in Stock Returns. *Quarterly Journal of Economics* 108:905–39.

Chordia, T., R. Roll, and A. Subrahmanyam. 2000. Commonality in Liquidity. *Journal of Financial Economics* 56:3–28.

Chamberlain, G. 1983. Funds, Factors, and Diversification in Arbitrage Pricing Models. *Econometrica* 51:1305–23.

Chen, J., and H. Hong, and J. Stein. 2001. Forecasting Crashes: Trading Volume, Past Returns and Conditional Skewness in Stock Prices. *Journal of Financial Economics* 61(3):345–81.

Connor, G. 1984. A Unified Beta Pricing Theory. *Journal of Economic Theory* 34:13–31.

Connor, G., and R. Korajczyk. 1988. Risk and Return in Equilibrium APT: Application of A new Methodology. *Journal of Financial Economics* 21:255–89.

[27] We thank Kenneth French for making this data available on his web site.

[28] The residual industry betas for both returns and turnover equal unity on average, by construction. The average R^2 of the regression of residual returns (turnover) on the residual industry portfolio return (turnover) is around 25% (15%) across the seven 5-year periods.

Connor, G., and R. Korajczyk. 1993. A Test for the Number of Factors in an Approximate Factor Model. *Journal of Finance* 48(4):1263–91.

Conrad, J., A. Hammed, and C. Niden. 1994. Volume and Autocovariance in Short-horizon and individual stock returns. *Journal of Finance* 49:1305–29.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam. 1998. Investor Psychology and Security Market Under- and Over-reactions. *Journal of Finance* 53:1839–86.

Dickey, D. A., and W. A. Fuller. 1979. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association* 74:427–31.

Duffie, D., N. Garleanu, and L. Pedersen. 2005. Over-the-Counter Markets. *Econometrica* 73:1815–47.

Fama, E., and K. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of financial economics* 1:1–33.

Fuller, W. A. 1976. Introduction to Statistical Time Series. New York: Wiley.

Gallant, R., P. Rossi, and G. Tauchen. 1992. Stock Prices and Volume. *Review of Financial Studies* 5:199–242.

Hasbrouck, J., and D. Seppi. 2001. Common Factors in Prices, Order Flows and Liquidity. *Journal of Financial Economics* 59(2):383–411.

Hong, H., and J. C. Stein. 1999. A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets. *Journal of Finance* 54:2143–84.

Hong, H., and J. Stein. 2003. Differences of Opinion, Short-Sales Constraints and Market Crashes. *Review of Financial Studies* 16(2):487–526.

Jiang, G., P. Mahoney, and J. Mei. 2005. Market Manipulation: A Comprehensive Study of Stock Pools. *Journal of Financial Economics* 77:147–70.

Jones, C. 2001. Extracting Factors from Heteroskedastic Asset Returns. *Journal of Financial Economics* 62(2):293–325.

Lee, C. M. C., and B. Swaminathan. 2000. Price Momentum and Trading Volume. *Journal of Finance* 55:2017–69.

Llorente, G., R. Michaely, G. Saar, and J. Wang. 2002. Dynamic Volume-Return Relation of Individual Stocks. *Review of Financial Studies* 15(4):1005–49.

Lo, A., and J. Wang. 2000. Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory. *Review of Financial Studies* 13:257–300.

Lo, A., and J. Wang. 2006. Trading Volume: Implications of an Intertemporal Capital Asset-Pricing Model. *Journal of Finance* 61(6):2805–40.

Mei, J., J. Scheinkman, and W. Xiong. 2003. Speculative Trading and Stock Prices. Working Paper, Princeton University.

Michaely, R., and J. Vila. 1996. Trading Volume with Private Valuation: Evidence from the Ex-Dividend Day. *Review of Financial Studies* 9:471–509.

Nagel, S. 2005. Trading Styles and Trading Volume. Working Paper, Stanford University.

Odean, T. 1998. Volume, Volatility, Price, and Profit When All Traders Are Above Average. *Journal of Finance* 53:1887–934.

Ofek, E., and M. Richardson. 2003. DotCom Mania: The rise and fall of internet stock prices. *Journal of Finance* 58:1113–38.

O'Hara, M. 2003. Presidential Address: Liquidity and Price Discovery. *Journal of Finance* 58:1335–55.

Pastor, L., and R. Stambaugh. 2003. Liquidity Risk and Expected Stock Returns. *Journal of Political Economy* 111:642–85.

Scheinkman, J., and W. Xiong. 2003. Overconfidence and Speculative Bubbles. *Journal of Political Economy* 111:1183–219.

Shukla, R., and C. Trzcinka. 1990. Sequential Tests of the Arbitrage Pricing Model: A Comparison of Factor Analysis and Principal Components. *Journal of Finance* 45:1541–64.

Vayanos, D., and T. Wang. 2003. Search and Endogenous Concentration of Liquidity in Asset Markets. *Journal of Economic Theory* Forthcoming.

Wang, J. 1994. A Model of Competitive Stock Trading Volume. *Journal of Political Economy* 102:127–68.