CHE2: Forecasting Chinese Equity Risk

Nathan Sosner, Xiaowei Li, and Guy Miller

August 2005

CHE2 forecasts risk in portfolios of mainland Chinese equities —i.e., in portfolios composed of A- and B-shares. Its predictions utilize daily returns data, so that the model responds quickly to changes in the dynamic Chinese risk environment.

1 Introduction

China's rapidly developing equity market attracts both individual and institutional investors keen to share in China's economic growth. Like many emerging markets, in addition to potential profits it offers demonstrated risks. Returns volatility in Chinese equities is typically high.

The problem of characterizing and quantifying Chinese equity risk is a difficult one. Not only do risk levels tend to be high, they are also highly volatile. Regulatory changes and sustained expansion continue to transform the market environment. Many new issues appear every year. Some 1400 issues trade on the Shanghai and Shenzhen exchanges in 2005, approximately triple the number present a decade earlier. Unrelenting growth and change make it difficult to uncover patterns of risk in the past and apply them to forecasts.

One way to mitigate the difficulties posed by rapid change is to base risk predictions on high-frequency data. Daily data offer roughly 20 times more information than monthly data, and permit quicker recognition of changing conditions.

MSCI Barra's new model for Chinese equity risk, CHE2, uses daily returns data throughout. This represents a significant advance over its predecessor, CHE1. CHE1 uses monthly returns data and consequently recognizes change more slowly.

Another key difference between CHE2 and the earlier model is coverage. It is hard to define the Chinese market without exciting some controversy. A key question is whether to combine some or all Hong Kong-traded equity with the mainland in a Chinese risk model. Although some Chinese shares representing companies that operate on the mainland trade in Hong Kong (H-shares and Red Chips), most mainland Chinese equities trade in the domestic stock markets of Shanghai and Shenzhen. Returns to Chinese shares resemble those of the market in which they trade rather than the markets where their firms operate, so from a modeling perspective it is best to devote separate models to shares that trade in Hong Kong and to those traded in mainland markets. Thus, although CHE1 includes some Hong Kong-traded shares, CHE2 confines itself to

Chinese A- and B-shares — equity issues that trade on the Shanghai or Shenzhen exchanges. It aims to be exhaustive within this scope.

The remaining parts of this paper explore the construction of CHE2 and compare it closely with CHE1. Section 2 briefly reviews the ideas behind factor models of equity risk, setting CHE2 in context and making subsequent material more accessible for readers with little previous exposure to factor models. Section 3 addresses CHE2 asset coverage in more detail, and describes the data that CHE2 applies to its forecasts. The CHE2 model factors represent sources of return that are common across assets. The factors are of two kinds. *Industry factors* describe the influence of industry membership on risk. The model's industry classification system and the rationale behind it are discussed in Section 4. *Style factors* describe non-industry sources of common risk. Section 5 reviews the complement of style factors chosen for CHE2. Section 6 tells how CHE2 uses daily factor returns data to forecast risk. Risk not associated with common factors is specific to individual assets. It is diversifiable. CHE2's specific risk forecast is the topic of Section 7. A detailed review of model performance occupies Section 8. Section 9 advises current users of CHE1 of changes to expect on moving to CHE2. Conclusions and a summary appear in Section 10.

2 CHE2 as a Factor Model

2.1 The Need for Factor Models

The risk of an equity portfolio depends on the relationships between its constituent assets. Quantitatively, if the portfolio risk is defined as the (annualized) standard deviation σ_p of the portfolio return, the asset returns covariance matrix V contains all the information required to evaluate its risk. That risk is given by the variance,

$$\sigma_p^2 = w^t V w.$$

Here w is a column vector of asset weights. If the fifth asset represents 2% of the portfolio's wealth, the fifth component of w is $w_5 = 0.02$. If an asset is not present in the portfolio, its weight is zero.

One of the greatest sources of difficulty in forecasting risk is the very large number of relationships between assets. In a market with N issues, the covariance matrix V is a symmetric $N \times N$ array of numbers. To predict risk in a market with 1,000 assets, one must forecast an independent covariance for each pair of assets — roughly half a million numbers. If these forecasts depend on daily historical returns, then the risk model must accumulate over 500 days of data with 1,000 returns each even to begin to have enough information for the task. Ideally, a span of time much longer than this would enter the forecasts, perhaps 10 years of data. This is undesirably long in developed markets and is completely unfeasible in rapidly evolving markets such as China. To forecast risk well, it is essential to use data more efficiently.

Worse than being inefficient, uncritical use of the historical record is dangerous. Accidents can readily distort a naïve historical appraisal of risk. For example, if one company had happened to report good earnings and return strongly while another found its shares suffering because of a corporate scandal, the historical coincidence would mistakenly suggest that the two assets hedge one another. In reality, they would not.

2.2 Factor Models

Factor models of risk apply data efficiently and reduce the impact of historical accidents on forecasts. They do this by asserting that assets respond to a relatively small number of sources of common return, called common factors. Exposures X link the factor returns f to individual asset returns. The factor model decomposes the asset returns via the attribution equation

$$r = Xf + u.$$

The $N \times 1$ column vector r of asset returns resolves into a common factor component — an $N \times M$ matrix of asset exposures times an $M \times 1$ column vector of factor returns — and a specific return component u. If the factor model performs well, the specific returns of one firm will be statistically independent from those of another. Consequently, the associated specific risk will diversify in a portfolio. Increasing the number of names in the portfolio decreases its specific risk. Factor risk must be hedged by controlling portfolio exposures to the individual risk factors.

The factor model's forecast of risk is

$$\begin{aligned}\sigma_p^2 &= w^t V w \\ &= w^t X F X^t w + w^t \Delta w.\end{aligned}$$

The factor model has re-expressed the asset covariance matrix V as a sum of common factor and specific components. These involve F, the common factor returns covariance forecast, and Δ, the specific returns covariance forecast. The common factor covariance matrix is comparatively small ($M \times M$) and requires forecasting far fewer quantities than the asset-level covariance matrix V. The specific returns matrix Δ is diagonal (if one ignores the relations between separate issues from the same firm; see the comments in Section 7 on linking assets from a single-issuer in the specific risk forecast) and so has only N non-zero entries. The model's factor structure drastically reduces the number of quantities to be forecast, and so can apply the available returns data efficiently to evaluating important cross-asset returns relationships.

2.3 Risk Model Factors

The model factors represent characteristics that cause similar assets to experience similar returns. One of the most basic characteristics is industry membership. For example, one expects the returns of information technology companies to resemble one another more closely than they resemble the returns of assets selected from other industries. To reflect this, each information technology company has an exposure $X_{IT} = 1$ to the Information Technology industry factor. A company in another industry will have an exposure $X_{IT} = 0$, but will have an exposure of 1 to its own industry. Relationships between the industries appear in the factor covariance matrix F.

Characteristics such as Value (as expressed by ratios such as book-to-price or earnings to price) or Size (as expressed by the logarithms of market capitalization and of assets) also help place stocks into groups that experience some returns commonality. These non-industry style characteristics are represented by factors called *risk indices*.

2.4 The Factor Regression

To forecast the factor covariance matrix F, a history of factor returns is necessary. Statistical relations between the factor returns are evaluated from the history. To produce the history, a multivariate factor regression extracts factor returns from the asset returns.

The regression uses a group of assets called the *estimation universe* or ESTU. In CHE2 the ESTU consists of all A-shares that have traded for at least one full month and have neither been designated as ”Particular Treatment” nor “Special Treatment'' stocks by the exchanges within the past 12 months. The history of the CHE2 ESTU appears in Table 1.

The regression identifies the factor returns that best explain the asset returns. Quantitatively, it chooses factor returns to minimize the weighted sum of squared residuals,

$$WSSR = \sum_j \rho_j \left(r_j - \sum_k X_{jk} f_k \right)^2 .$$

The sum is over all assets. The regression weights ρ_j are proportional to the square root of market float capitalization. (As a refinement, the largest weights in the model are reduced to equal the weight at the 95th percentile.) This choice ensures that the assets of greatest interest, the high capitalization assets, influence the factor returns most strongly. In the case of the risk indices, it also ensures that a large number of assets contribute to the factor return. This is necessary to prevent the estimated factor return from being contaminated by the specific returns of a few very influential assets.

A note on capitalizations: In this report, the term “capitalization” always refers to the capitalization of traded shares in a firm, the float capitalization. The large proportion of government ownership in China makes it difficult to define a compelling market valuation for the entire firm.

2.5 Thin Industry Corrections

Despite the advantages of the square root of capitalization weighting scheme, it may still happen that in a sparsely populated industry or in an industry with a few very high capitalization assets, undue weight will be placed on a few assets. This means that, within an industry, the effective number of assets contributing to the factor return may be very small. A *thin industry correction* employs a simple Bayesian prior to ameliorate this problem.

When necessary, CHE2 applies the thin industry correction. The "effective" number of issues[1] in industry i is

$$n_i = \frac{W_i^2}{\sum_{j \in i} \rho_j^2}.$$

The sum is over issues in the industry; $W_i \equiv \sum_{j \in i} \rho_j$ is the total regression weight of assets in the industry. If an industry contains fewer than five effective assets, then a proxy asset is added to it. The proxy asset has a return equal to the market return and a weight equal to

$$(5 - n_i)\frac{W_i}{n_i}.$$

The proxy has no non-zero risk index exposures. During a time when an industry contains no assets, the industry factor return is set equal to the market return. The thin-industry correction reduces the influence of asset-specific behavior on industry factor returns.

2.6 B-shares and the B-share Secondary Regression

The CHE2 estimation universe contains only A-shares. It therefore cannot assign factor returns to the two CHE2 risk indices that are particular to B-shares. These indices are BUNIV, which captures the return that stems from being a B-share rather than an A-share, and BSZSE, which captures the return peculiar to Shenzhen-traded B-shares.

---

[1] Readers may recognize this measure as the reciprocal of the Herfindahl index.

The BUNIV and BSZSE returns are estimated in a secondary regression. First, returns attributed to the other model factors are subtracted from all assets to form residuals. The residuals to the A-shares become their specific returns. The secondary regression runs over the B-share residuals. It is weighted by the square root of capitalization in the same way as in the primary regression. The B-share secondary estimation universe consists of all B-shares that, like shares in the primary estimation universe, have traded for at least one month and have not been designated as “SP/PT” stocks over the past twelve months. The returns attributed to BUNIV and BSZSE are subtracted from the B-share residuals and the remaining residual returns become the B-share specific returns.

3 The CHE2 Universe and CHE2 Data

3.1 The CHE2 Universe

In deciding whether to include Hong Kong-traded equity in CHE2, the need to maintain a certain degree of market homogeneity was an important consideration. Factor models of equity risk such as CHE2 perform best when they describe a relatively homogeneous group of assets. To build a model covering two disparate markets or asset classes, the best approach is often to build two distinct sub-models and then devise an interface to unite them. It would be a mistake to try to force assets from very different markets into the same factor structure.

A comparison of the returns to four portfolios clarifies the issue of market membership. They include a portfolio of Hong Kong shares (less Red Chips and H-shares), a portfolio of Chinese A-shares, a portfolio of Red Chips, and a portfolio of H-shares. Each portfolio weights its constituents by the square root of float capitalization. This weighting allows a broader representation of assets than weighting by float capitalization, but still emphasizes large companies.

The comparison is implemented through multiple regressions of the H-share portfolio and Red Chip portfolio returns against returns to the Hong Kong and A-share portfolios. Table 2 summarizes the results over three historical periods: January 1998 through June 2001, July 2001 through June 2005, and the combined period. The results indicate that the Hong Kong portfolio largely explains the Red Chip portfolio returns — the beta to the Hong Kong portfolio is statistically very significant and no r-squared is less than 70%. At the same time, the beta to the portfolio of A-shares is always insignificantly small. One concludes that historically the Red Chip portfolio resembles the Hong Kong market in which Red Chips trade more closely than the Chinese domestic market, although Red Chips represent companies that operate and do most of their business in China's mainland. Matters are not quite as clear-cut for the H-shares, although the conclusion is the same. H-shares apparently combine aspects of both the Hong Kong and mainland markets, but are dominated by Hong Kong.

CHE1 includes A-shares, Red Chips, and H-shares in its estimation universe. Rather than describing A-share returns well or Hong Kong share returns well, it attempts a balance between them that results in a poorer overall description. Specifically, the average adjusted r-squared of CHE1 monthly factor regressions over the period extending from January 2000 to June 2005 is 22%. The adjusted r-squared tells what fraction of the squared asset returns could be ascribed to the model's common factors. The larger the fraction, the more successful the model is in capturing sources of return characterizing risk. The average adjusted r-squared over the same period of a model that treats only A-shares and that only invokes a single factor — the market factor, to which all assets are given an exposure of 1 — is 33%. Even this very simple model outperforms CHE1 when allowed to concentrate its efforts on a more tractable universe of stocks. The performance of CHE1 suffers because its set of assets is too heterogeneous for its factors to accommodate. CHE2 addresses only A-shares and B-shares; its estimation universe includes only A-shares. The average adjusted r-squared of CHE2 over 2000/01–2005/06 is 43%. This compares favorably with successful models for developed markets such as the US or Japan.

3.2 CHE2 Data

The CHE2 model history begins with forecasts for January 1998. These forecasts are based on market and financial reporting data that extend back to 1991. The market data include daily share prices, dividends, and volumes. They cover all equities traded on the Shanghai and Shenzhen exchanges.

All data have been screened carefully for coverage gaps and errors. Financial report data in particular have been checked for consistency between accounting items and year-to-year consistency; year-to-year changes are checked for abnormally large jumps both in absolute and in percentage terms. Sporadic manual comparisons were also made between the raw digital input data and copies of the financial reports available independently. The final data history for CHE2 has been made as complete and error-free as possible.

4 Industry Factors

CHE2's industry structure is based on the standard GICS industry classification scheme. This consists of ten broad sectors that are further subdivided into industry groups, industries, and sub-industries. The 24 CHE2 industry factors emphasize sectors that are important in China, and include some provisions for developing industries.

4.1 Considerations Underlying the CHE2 Industry Structure

Development of the CHE2 industry scheme began at the level of GICS industry groups. The final industry factors are the result either of aggregating groups or of disaggregating them into finer categories. Industry group aggregation occurs only *within* GICS sectors; industry groups are never aggregated across sectors.

CHE2 industries are generally required to be well populated (specifically, the effective stock number n_i should not be too small) and to be economically important in the sense that they occupy a nontrivial fraction of the total market capitalization. From the standpoint of risk modeling, the requirement that an industry be well populated is essential. The estimated industry factor return is basically a weighted average of the contributing asset returns and so it necessarily contains some contamination from specific returns. If the effective number of assets in the industry is too small, the specific returns will not diversify well; the factor return will be ill determined and loaded with specific return. An industry that is unpopulated or lightly populated early in the model's history can be accommodated by means of a thin industry correction to the industry factor return. The correction essentially blends the raw industry factor return with the market return. The logic behind the correction is that in the absence of statistically reliable information about the industry factor return, the market return provides a simple and reasonable proxy that is not loaded with specific risk and estimation error.

Another general requirement for CHE2 industries is that their returns behave distinctly from one another. Thus, if the difference between two candidate industry factor returns is smaller than the accuracy with which they can be determined, the two candidates usually are aggregated. Nevertheless, the Chinese equity market is still young and its industry composition changes constantly as new companies list their shares. If a candidate industry seems to represent an important emerging area but is still so thinly populated that its factor identity is only weakly established, the requirement of statistical distinction may be weakened or waived. Examples of such CHE2 industries are Commercial Services & Supplies and Telecommunications Services.

4.2 An Illustration of Industry Factor Selection

The CHE2 Steel industry illustrates the selection of industry factors. At the end of October 2004, the GICS Metals & Mining industry included 83 ESTU assets and accounted for about 7% of Chinese market capitalization. The number of assets and the proportion of aggregate capitalization were both large enough to suggest that some sub-industries might constitute a separate industry.

Steel is a natural candidate industry to be tested for spawning from Metals & Mining. In October 2004 there were 39 assets in the GICS Steel sub-industry. They accounted for about half of the market capitalization of the Metals & Mining industry and represented 3.24% of the aggregate market capitalization. It was therefore clear that Steel companies comprise a prominent component of the Chinese market.

Statistical testing followed the identification of Steel as a candidate industry factor. In the tests both Steel and non-Steel assets within the Metals & Mining category were exposed to the Metals & Mining factor. In addition, Steel assets were given an exposure of 1 to a Steel "difference" factor. Factor regressions on monthly ESTU asset returns (without risk index exposures; at this point in model development no risk indices have been defined) produce an industry factor returns history for the test. This peculiar arrangement causes the Steel factor return to carry only the difference between Steel and non-Steel Metals & Mining returns. If this return differs significantly from zero within the history, then it is established that Steel behaves distinctly from the rest of the Metals & Mining industry and should be separated from it.

Two test statistics determine the significance of the Steel industry difference return. The first is the F-statistic, which is essentially the squared t-statistic of the Steel return averaged over the test history. (The t-statistic is the ratio of the measured factor return to the estimated error in measurement. A high t-statistic indicates a well-determined factor return.) The second test statistic is the proportion of significant t-statistics (t-statistics greater than 2 or less than -2) in the history. The F-statistic shows how strong the industry distinction is overall, while the significance proportion discriminates between persistent distinction and apparent distinction generated by a few rare returns. F-statistics above 2 and significance proportions above 20% are good indications that the candidate industry should be separated from its parent.

The Steel industry was tested in three alternative samples: the full period from January 1998 to October 2004, the bull market from January 1998 to June 2001, and the weak market from July 2001 to October 2004. Table 3 shows the results of statistical tests for these periods. For the full sample period, Steel's F-statistic is well above 2. This result is not persistent from month to month, but rather derives from 14 very significant months — this is indicated by the fact that the significance proportion is near the threshold for acceptance. Nevertheless, distinct returns occur frequently enough to identify Steel as an independent industry.

4.3 The CHE2 Industry Structure

The final CHE2 industry structure emphasizes the Materials, Industrials, and Consumer Discretionary GICS sectors, making numerous distinctions within them. There are 24 CHE2 industries in total. Table 4 shows the breakdown of the ten GICS sectors into CHE2 industries. It also shows the CHE1 industry scheme. Some important differences are apparent between the two industry structures. First, there is no catchall industry such as CHE1's Miscellaneous industry in CHE2. Second, new industries such as Steel, Commercial Services, and Media have been introduced in CHE2, and Information

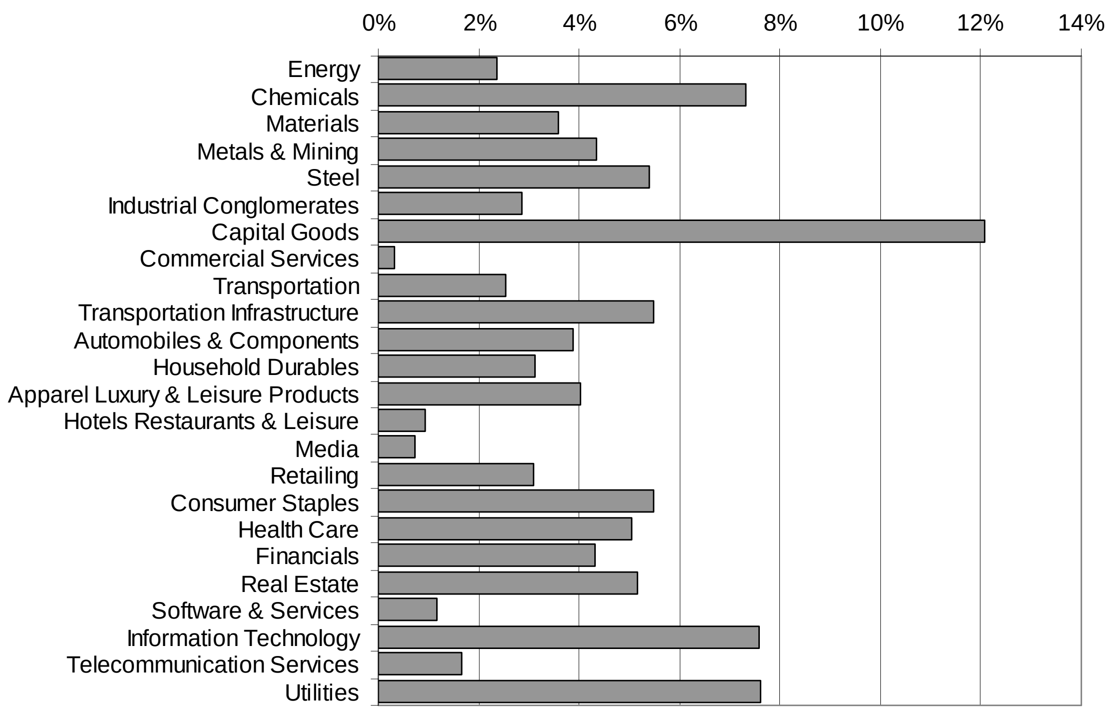
Technology has been divided into the Software & Services and IT (roughly, Hardware & Equipment) industries. However, many CHE1 industries survive in CHE2. Examples include Chemicals, Energy, Finance, Health Care, and Real Estate. That the independently derived CHE2 industry scheme preserves many features of CHE1's industry structure suggests that even in China's rapidly growing markets there are enduring distinctions that have maintained their usefulness.

Figure 1 shows market shares for the CHE2 industries in October 2004. Most industries have large market shares. The few industries with relatively small shares are Commercial Services; Hotels, Restaurants, & Leisure; Media; Software & Services; and Telecommunication Services. These industries exist independently to allow CHE2 to accommodate development in the Chinese economy: we expect that their importance will grow as the Chinese economy becomes more services-oriented.

How well estimated are the CHE2 industry factors? The performance of the raw industry factors is good but uninformative; all industries are exposed to the market and bear similar amounts of market return. (Here the market is defined as a float capitalization weighted portfolio of ESTU A-shares.) To gauge the performance of individual industry factors it is better to examine their returns in excess of the market and see what additional information they contain. Table 5 shows the results for the market factor and the market-excess industry returns. Two thirds of the industries have F-statistics above 2 and significance proportions exceeding 20%. In addition, the table shows that the market content of many industries is similar — the betas of many market-excess industry returns are close to zero. However, there are some industries that have significantly higher betas: IT, Software & Services, and Industrial Conglomerates. There are also a few industries that have significantly lower betas: Steel, Transportation Infrastructure, and Materials.

A final indicator of overall industry structure performance is the increase in the adjusted r-squared that accompanies the introduction of the industry factors in regressions. For the sample period from January 2000 to June 2005, the average adjusted r-squared of a model with all assets uniformly exposed to a single market factor is 32.6%. Introducing the industry factors increases the average adjusted r-squared to 36.5%, a very significant improvement. The industry structure of CHE2 was completed in November 2004. In the first six month of 2005, the industry structure increases the adjusted r-squared over the single-market factor from 39.0% to 43.7%; the value of the industries is being maintained out-of-sample.

Figure 1: Fraction of the total market float capitalization in each CHE2 industry.



5 Risk Indices

The CHE2 risk indices describe sources of return that go beyond industry membership. They include technical factors, fundamental financial factors, and exchange factors. Some of these factors are known to be important in other markets around the world, and some are unique to the Chinese market.

A major feature of CHE2 is its use of daily data. Not only do daily returns play an important role in rapidly assessing risk levels, they also permit fine distinctions in the returns-based characteristics of different assets. Characteristics such as beta can be estimated more reliably from daily data than from monthly, even given a shorter data history. For example, the period used for beta estimations can be shortened from five years to just one year. This makes returns-based risk factors more responsive, while still ensuring that they are stable enough for institutional investing.

Each risk index exposure quantifies a general property recognized and priced by investors. Examples include Size, Momentum, and Value. For most risk indices, the exposure is a scale that runs roughly from -3 to 3. Outlying exposures are truncated to avoid extremely large values. Exposures are normalized so that their standard deviation in the estimation universe is equal to 1. Also, the capitalization-weighted average over estimation universe stocks is zero. This normalization makes the capitalization-weighted market style-neutral, and gives the scale of the style exposures a persistent meaning even as the number of assets in the model changes over time.

Risk index exposures are weighted averages of one or more descriptor values. The descriptors represent elemental characteristics that contribute to the more general property represented by a risk index. Table 6 compares CHE2 and CHE1 risk indices and descriptors. Appendix B supplies a list of descriptor definitions for the CHE2 risk indices.

5.1 Technical Risk Indices

Technical risk indices are based on trading characteristics, particularly on returns and volumes. The technical risk indices in CHE2 are Volatility, Momentum, Trading Activity, and Downside Momentum.

The Volatility risk index is a strong source of risk. It includes descriptors known to be important in many markets, such as historical beta and sigma. These respectively characterize the amount of market and non-market risk the asset bears. The index also features descriptors such as the median log of the daily high-to-low price ratio. This descriptor captures the tendency of a stock to experience large price movements within a day. Another descriptor, the daily variance ratio, captures the tendency of unusually high or low returns to cluster in time, thus contributing to large excursions in cumulative return.

Momentum is a widely recognized characteristic that describes an asset's sustained returns performance relative to other stocks. The Momentum risk index quantifies historical returns success. Like the Volatility factor, it is a strong risk source.

The Trading Activity risk index shows how actively a stock has traded and how strongly trading volume affects its price. Stocks with high Trading Activity exposures have high turnover and low price impact. They are particularly liquid.

Downside Volatility exposures are high for stocks that experience increased volatility and correlation with the market when the market declines. From 2001 through 2005 the Downside Volatility factor has shown a very significant and positive average alpha (c.f. Table 7 for a list of historical factor alphas).

5.2 Fundamental Risk Indices

The CHE2 fundamental risk indices benefit from accounting data with excellent coverage and thorough quality control (see Section 3). These data are used to produce three risk indices: Value, Growth, and Leverage. Of the three, only Value has a counterpart in CHE1.

The Value risk index ranks stocks by fundamental ratios such as book-to-price, earnings-to-price, and cash-flow-to-price (dividend to price). The Dividend Yield risk index of CHE1 is not present in CHE2, but instead enters the model as the cash flow descriptor in the Value risk index. The reason for this demotion is that few Chinese stocks pay cash dividends; usually share dividends or rights distributions are preferred. Placing a dividend descriptor in the Value risk index follows financial tradition and avoids giving dividends undue emphasis. Since 2001, the Value risk index has shown a significant factor alpha.

The Growth risk index is based on the concept of historically demonstrated growth. It includes descriptors for historical earnings growth, as well as for return on assets and return on equity. An "unexpected earnings" descriptor identifies unusually large earnings in the most recent reporting year.

The Leverage risk index describes a company's capital structure. In particular, it combines book leverage, the debt-to-assets ratio, and market leverage to create an overall characterization of the firm's debt.

5.3 Exchange Risk Index

A single risk index captures returns that distinguish Shanghai and Shenzhen exchanges-traded issues. The Shenzhen risk index is equal to 1 if a stock trades on the Shenzhen exchange, and is equal to 0 otherwise.

5.4 B-Share Residual Factors

CHE2 model factors have been designed with the primary goal of describing A-share risk and return. Nevertheless, the model also makes an effort to achieve a good description of B-shares. How are B-share returns distinguished from those of A-shares? A regression of B-share residual returns (i.e., returns remaining after the model industry factors and other risk indices have been attributed) on two B-share factors addresses this question. One factor is a B-share market factor, equal to 1 for every B-share. This places a constant term in the B-share residual return regression, and captures the overall difference between the A-share market and the B-share market. The second factor is an indicator variable that identifies B-shares listed in the Shenzhen stock exchange. It is equal to 1 for B-shares listed in Shenzhen and 0 for B-shares listed in Shanghai. Both factors are significant in historical regressions, and help CHE2 to include B-shares in its sphere of applicability.

5.5 Explanatory Power of the Factors

Table 7 summarizes the statistical properties of the CHE2 risk indices. The table presents historical results in two consecutive four-year periods. The first period is the bull market ending in June 2001, and the second is the weaker market that started in July 2001. Except for Leverage, all of the style factor returns are strongly significant in statistical tests. They are much larger overall than the returns one would expect from a factor with randomly assigned exposures (i.e., their returns F-statistics are much larger than 1). Furthermore, their strengths do not derive from peculiarly large individual returns but rather emerge repeatedly in a good fraction of months (the proportion of significant t-statistics is greater than 20% in all cases). The table also reveals that some factors have been appreciable sources of alpha, although performances vary across periods. (The measured factor alphas are considered statistically significant when their t-statistics are greater than 2 or less than -2.)

Although Leverage is weaker than the other risk indices, it remains in the model because it is statistically significant in the second period. It also carries a premium. Although its alpha is small — 1.32% annualized — it is very persistent and has a t-statistic of 3.15. This means that, over the past few years at least, the Leverage factor has figured importantly in performance attribution.

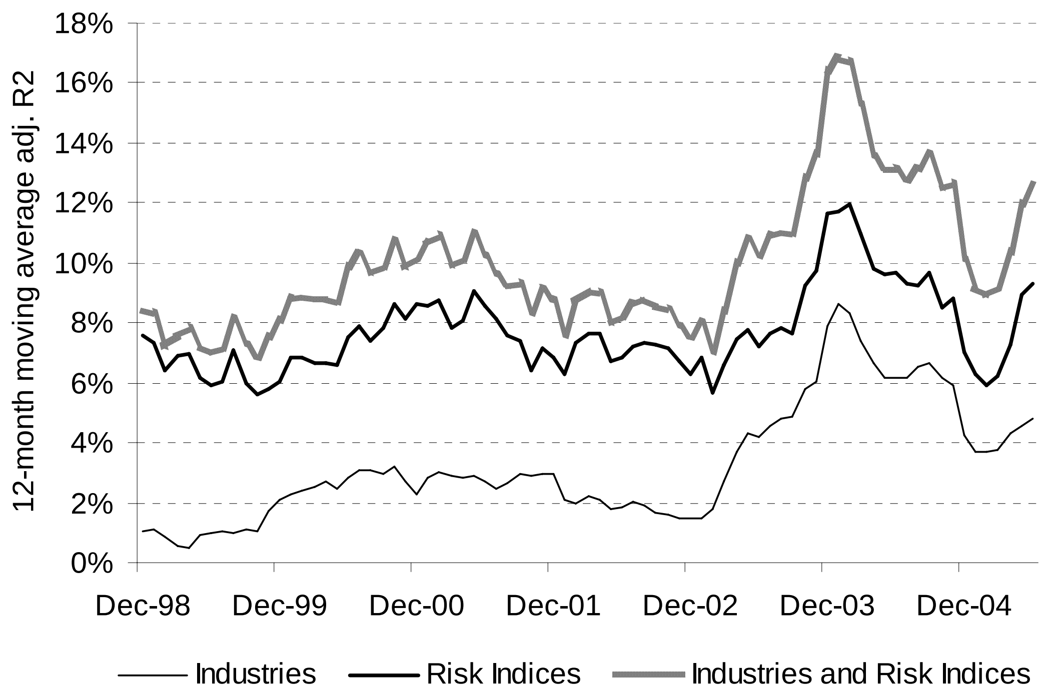
Risk index factors are designed so that the market has no exposure to them. This suggests — although it does not guarantee — that most risk indices describe non-market risk. The last two columns of Table 7 show that the market betas of all the risk indices are indeed economically small in both sample periods. The amount of non-market risk borne by each risk index is given by its CAPM sigma. The annualized sigmas in Table 7 lie in the approximate range 2–7% in the earlier period and 1–4.5% in the later, more quiescent period.

The historical performance of the B-share indices is summarized in Table 8. It shows that both B-share style factors capture strongly significant sources of risk. B-shares were very volatile in the first period, and experienced extremely large returns when they became available to domestic investors in 2001. Although their volatility diminishes in the second period, the B-share indices continue to represent important independent returns sources for B-shares.

Table 9 displays historical correlations between the risk factor returns. The correlations below the diagonal are for the overall period 1998/01–2005/06. The correlations above the diagonal are for the recent sub-period 2001/07–2005/06. While a few pairs show larger historical correlations (e.g., Momentum and Growth), the correlations are small for most pairs in both periods. This suggests that the risk factors respond to influences that are largely independent.

How much do the CHE2 risk factors contribute to the overall explanatory power of the model? A useful statistic for this inquiry is the adjusted r-squared. It tells what fraction of the asset-level returns variance can be attributed to common factors. Figure 2 starts with a simple single-factor model in which all assets are given exposures of 1 to the market. It then shows the improvement in adjusted r-squared that occurs when the CHE2 industry factors are introduced, and when the risk indices are introduced. The average adjusted r-squared of the single factor model for the period 1998/01–2005/06 is 30.9%. When the market factor is replaced by industry exposures, the average adjusted r-squared increases to 34.2%. Adding risk indices to the single market factor model increases the r-squared to 38.8%. If both the industry factors and the risk indices are put in place, the average adjusted r-squared becomes 41.1%, an improvement in r-squared of 10% over the single factor model. The CHE2 factors perform well by any standard.

Figure 2: Explanatory power of the CHE2 factors.



6 Factor Covariance Forecasts

The pace of change in the Chinese equity market is fast, and the market itself does not have a long history. To help meet these two challenges, factor covariance forecasts in CHE2 are based on daily returns data. This represents a substantial improvement on

CHE1, which used monthly factor returns to estimate factor covariances. The higher density of daily data makes it possible for CHE2 to recognize changes in risk levels more promptly and accurately than was possible for CHE1.

The CHE2 common factor risk forecast is based essentially on an average historical risk level. The model achieves sensitivity to change by emphasizing recently realized risk in calculating the average, and down-weighting risk that occurred in the more distant past. More precisely, historical averages are exponentially weighted. The most recent period T is weighted most heavily, and the weight of each period is larger than the one immediately preceding it by a fixed factor. The *half-life* t_{half} of the exponential weighting scheme tells how rapidly weights decline moving into the past: the weight of data at $T - t_{\text{half}}$ in the average is half of the weight of data at time T. The half-life for factor volatility levels, produced from historical factor variances, is 90 trading days in CHE2. This short half-life makes the model very responsive to variations in factor risk, a necessary trait in a market that changes as rapidly as China's. For comparison, the half-life of CHE1's covariance forecast is 48 months.

To see quantitatively how the CHE2 factor risk forecast works, start with the statistical properties of daily returns. The average daily returns variance is

$$\sigma_d^2 = \text{avg}\left(\left[f_d - \text{avg}\left(f_d\right)\right]^2\right).$$

From it one can calculate the implied daily risk, σ_d. If the returns from one day were uncorrelated with returns from any other day, the monthly risk σ_m would be related very simply to the daily risk: $\sigma_m^2 = N_d \sigma_d^2$. Here $N_d \approx 21$ is the number of trading days in a month.

Real returns do exhibit correlations across days. This changes the way daily risk aggregates to monthly levels. Cross-day correlations --- also called serial correlations --- can affect monthly levels of risk significantly. Because cross-month correlations are much weaker than cross-day correlations, monthly risk aggregates more simply to longer horizons.

As an example of how serial correlations affect monthly risk, imagine that a factor return exhibits a correlation of 0.1 across adjacent days, but that a return on one day is uncorrelated with the return two days later. The positive cross-day correlation implies that larger than average returns show a weak tendency to group together, as do returns below the average. Returns therefore do not diversify fully across days. Instead, movement away from the average is sustained across days, and multi-day departures from the average will be greater than if returns on adjacent days were unrelated. The same will be true over a month: monthly risk is greater than it would be without serial correlations. Quantitatively, the monthly variance in this example is $\sigma_m^2 \approx 1.2 \times N_d \sigma_d^2$, 20% larger than it would have been without the cross-day correlation. The risk, expressed as a standard deviation, is about 10% larger than in the absence of correlations.

Serial correlations can also reduce risk. Factor returns that are negatively correlated across days experience less risk over a month than similar but uncorrelated returns would. In this case, returns on adjacent days tend to cancel, reducing longer-term departures from the average.

The CHE2 factor covariance forecast uses cross-day correlations to adjust the way daily risk aggregates to monthly levels. It accounts for correlations between any single factor return and factor returns on 10 subsequent days. This requires estimates of many correlations, both across days and across factors. The large number of required estimates demands a longer history of data than is needed for daily risk levels. CHE2 uses a "correlation half-life" of 480 trading days, a little under 2 years. Significant correlations should have an economic basis or a connection with liquidity. As such, they can be expected to change more slowly than overall risk levels and to allow a longer estimation data length. At the same time, CHE2's 90 trading day volatility half-life ensures that its covariance forecasts reflect recent (rather than multi-year) risk levels.

Figures 3a–3f show realized and forecast levels of risk in several CHE2 industry and risk index factors. Realized risks are based on a 6-month forward-looking data window of daily factor returns. There is no correction for serial correlations in the realized risks.

The figures show a dramatically varying risk landscape, which the CHE2 forecasts attempt to predict. Although CHE2 cannot anticipate abrupt changes in risk, it quickly adjusts to new conditions.

The general success of risk predictions can be quantified using *bias statistics*. The bias statistic for returns r_t and risk forecasts σ_t is defined as the standard deviation of the "normalized return" $z_t \equiv r_t / \sigma_t$. Qualitatively, the bias statistic is an estimate of the ratio of historically realized risk to forecast risk. Ideally, this ratio would be equal to 1. However, because the bias statistic is an estimate based on a finite selection of returns, even a perfect model will produce bias statistics that are not equal to 1. Instead, they will tend to lie within a range called the no-bias confidence region. At the level of 95% confidence, the no-bias region for a series of T normally distributed returns is $\left(1-\sqrt{2/T}, 1+\sqrt{2/T}\right)$. If the returns distribution is fat-tailed rather than normal, the no-bias region is broader than this. In this paper we conservatively adopt the region for normally distributed returns in deciding whether or not forecasts are biased. Bias statistics above the no-bias region indicate *under-prediction*: the risk forecasts were too small. Correspondingly, biases below the no-bias region indicate *over-prediction*: the risk forecasts were too large.

Table 10 displays historical bias statistics for the CHE2 industry and risk factor forecasts. Biases are reported in two periods, 1998/01–2001/06 and 2001/07–2005/06. (The month of March 2001 is omitted from the B-share risk index bias calculation, since B-shares were first made available to Chinese investors at the end of February 2001, causing a very large one-time shock the following month.) The table indicates in boldface which bias statistics fall outside of the no-bias confidence interval. CHE2's forecasts of monthly factor risk are very successful overall, particularly in the second period.

Only two bias results deserve special comment: those for the Metals and Telecom residual factor risk forecasts in 2001/07–2005/06. The attributed residual returns are calculated by subtracting from the monthly factor return f its market component $\beta_f r_m$:

$$f_{\text{residual}} = f - \beta_f r_m .$$

The factor beta β_f itself is determined by the model's factor covariance forecast. Similarly, the forecast of residual factor variance is based on the forecast factor variance σ_f^2 and the forecast market variance σ_m^2,

$$\sigma_{f,\text{residual}}^2 = \sigma_f^2 - \beta_f^2 \sigma_m^2 .$$

The Telecom bias of 1.99 — an extreme under-prediction of risk — is an artifact of the thin industry correction in CHE2. No telecommunications companies exist in the CHE2 estimation universe until 2003. When they appear, the Telecom industry factor return suddenly acquires non-market risk characteristics that it did not possess before. This produces a jump in residual return that causes a strong transitory bias. The residual return is still relatively small compared to the market component of the factor, so the forecast for total Telecom factor risk is unbiased.

The residual risk bias for Metals in 2001/07–2005/06 is due entirely to a single outlying return in December 2003. If the outlying return is removed from the history, the bias statistic becomes 1.03, indicating an unbiased forecast.

Figure 3a: Forecast and realized risk in the IT industry factor.

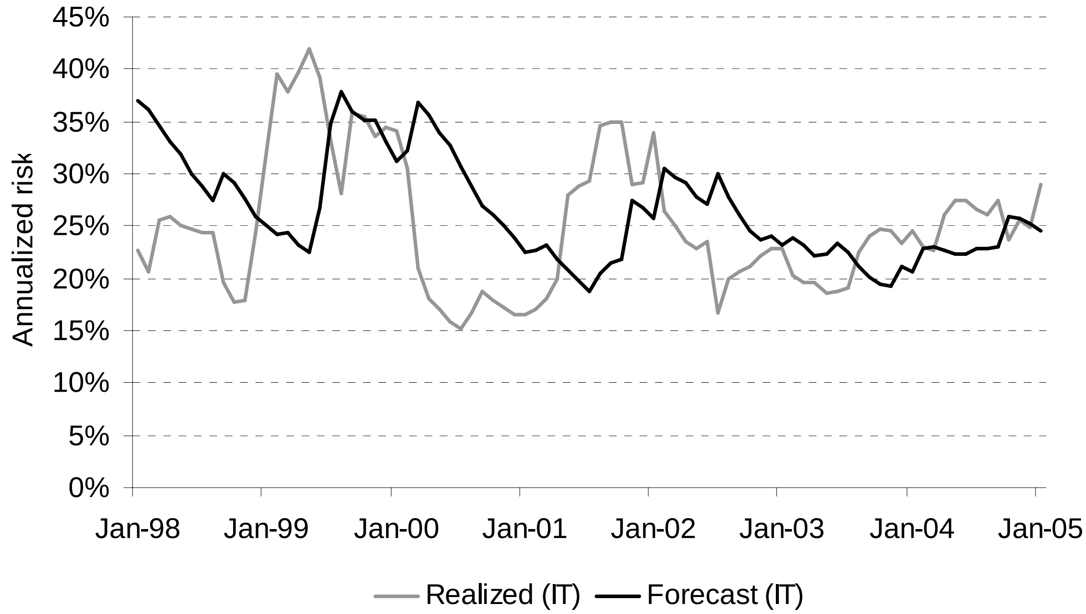


Figure 3b: Forecast and realized risk in the Infrastructure industry factor.

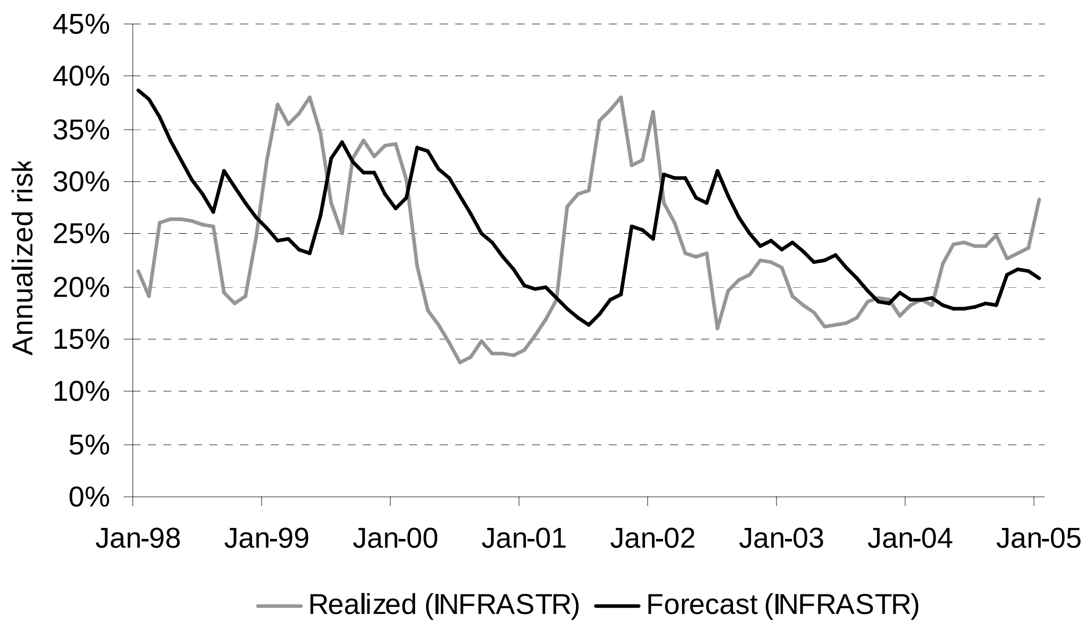


Figure 3c: Forecast and realized risk in the Size risk index factor.

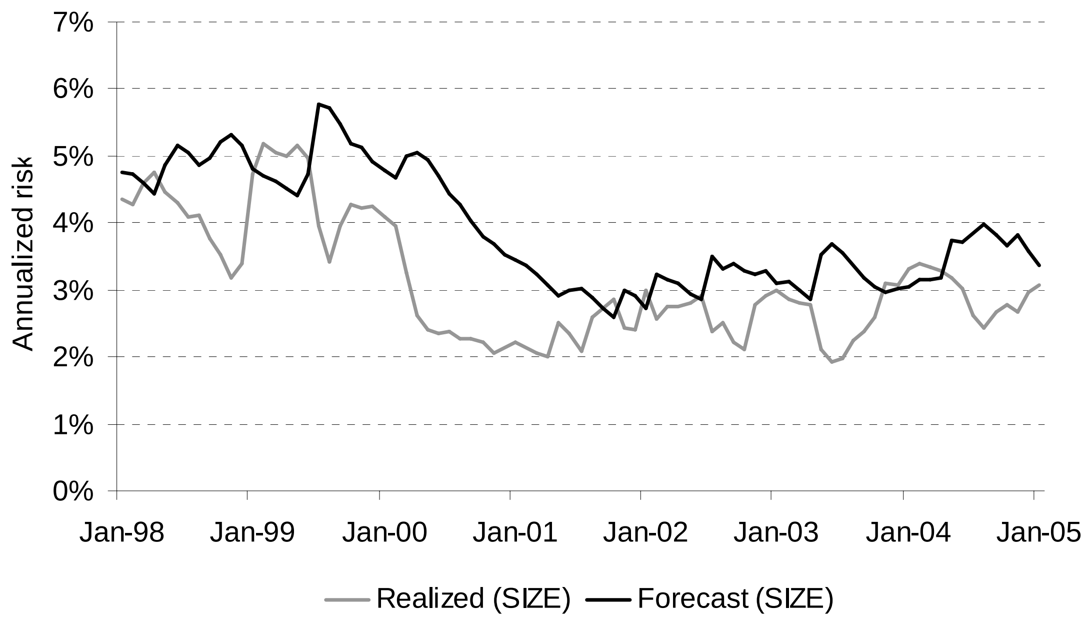


Figure 3d: Forecast and realized risk in the Momentum risk index factor.

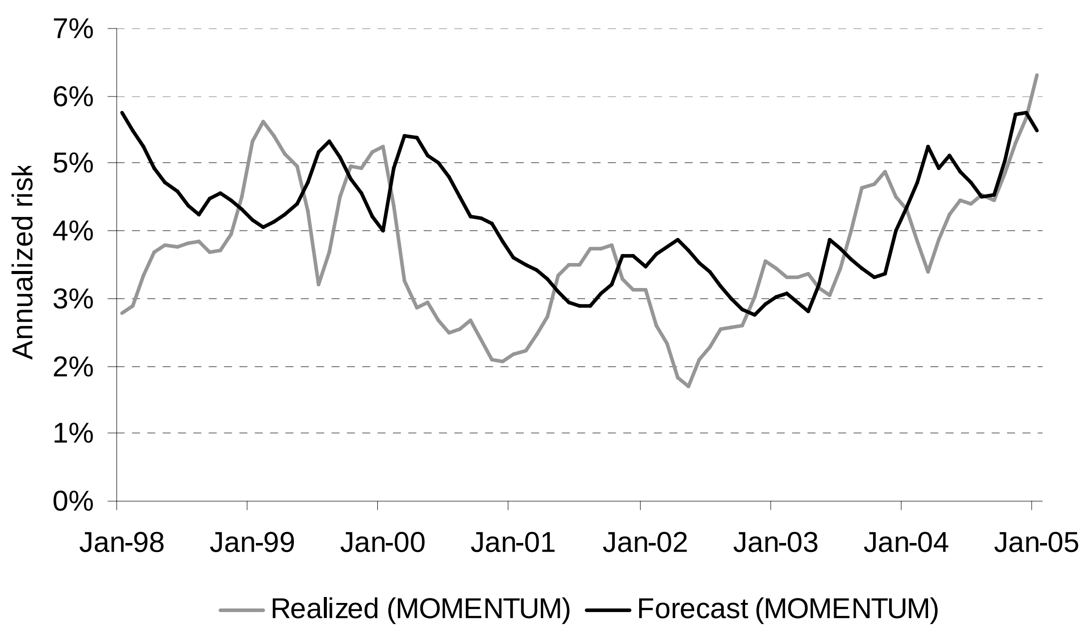


Figure 3e: Forecast and realized risk in the Trading Activity risk index factor.

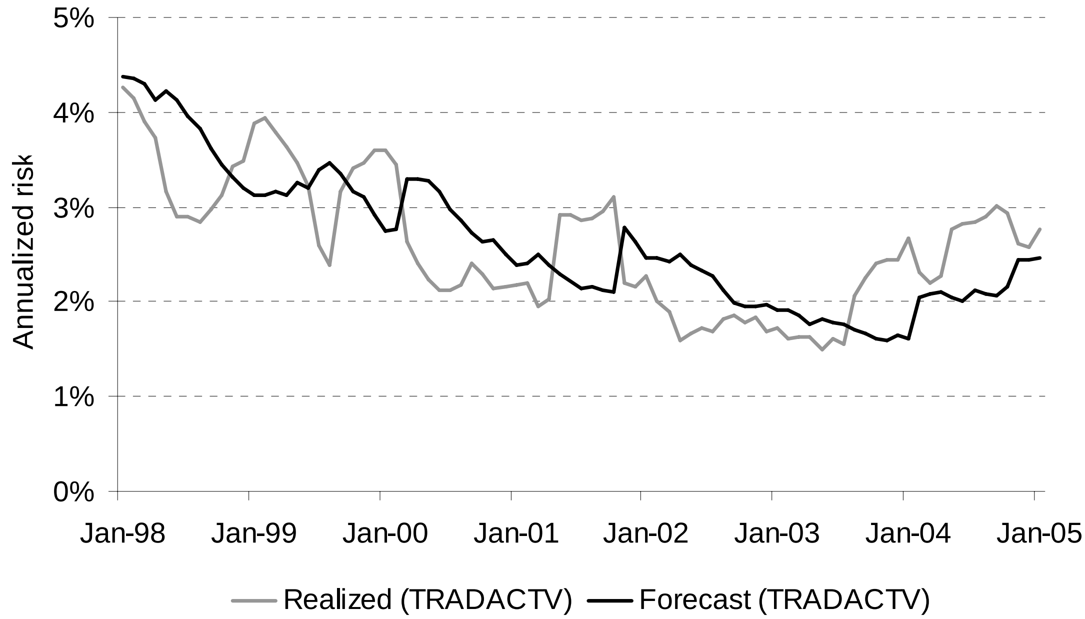
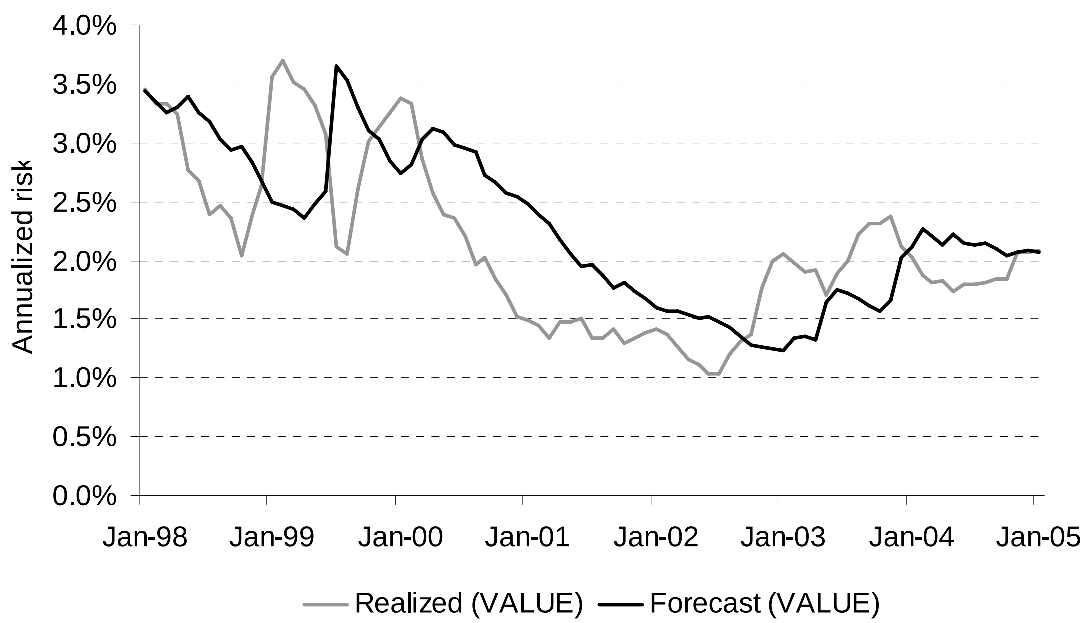


Figure 3f: Forecast and realized risk in the Value risk index factor.



7 Specific Risk

A structural model underlies CHE2's asset-specific risk forecasts. Structural models forecast similar levels of specific risk for companies that share similar characteristics: industry membership, size, leverage, and so on.

To appreciate the benefits of a structural specific risk model, it is helpful to contrast the structural approach with a commonly used alternative, forecasting based solely on historical risk. A historically based forecast simply says that an asset that has experienced low specific risk in the past will continue to do so. It may be that the company is in an industry that typically exhibits high specific risk, but the historical method does not increase its risk forecast until the company experiences large specific gains or losses. In contrast, a structural model would place the company's specific risk forecast closer to the average for its industry, and use other characteristics — including historically realized risk — to increase or decrease the forecast away from the level implied by industry membership. By doing so it does not have to wait until a disaster happens to increase the risk forecast. The structural technique is similar to that used by actuaries in the insurance industry. They do not have to see a tobacco user enter the hospital to know that he or she bears a higher health risk than a similar non-smoker.

7.1 The CHE2 Structural Specific Risk Model

The CHE2 specific risk model is based on methods developed at Barra and successfully deployed in models for several other markets. It expresses the forecast specific risk for an asset j as a product of three terms:

$$\sigma_{jt} = \hat{S}_t (1 + \hat{V}_{jt}) \kappa_{jt} .$$

The first term, $\hat{S}_t$, is a forecast for the average market level of specific risk. The second and third terms make adjustments that depend on the characteristics of the individual asset.

The model defines the average specific risk level S_t for month t as a capitalization-weighted average over the estimation universe of absolute specific returns $|u_{jt}|$. (The risk level is expressed in terms of absolute returns to make it less sensitive to unusually large individual returns than if squared returns were used.) This captures the typical size of specific returns across the market. The forecast for the average risk level is

$$\hat{S}_t = \alpha + \beta E_{t-1} .$$

Here E_{t-1} is an exponentially weighted average of S_{t-1}, S_{t-2}, and S_{t-3}, the realized market specific risk levels over the past three months. The model re-estimates the constant α and coefficient β every month with historical regressions to ensure that the forecast remains accurate. A history of forecast and realized average specific risk levels appears in Figure 4. The figure shows that the forecast level successfully tracks changes in the realized market-wide risk level, although the realized level can move quickly. The forecast generally accomplishes this smoothly and without unnecessary fluctuations.

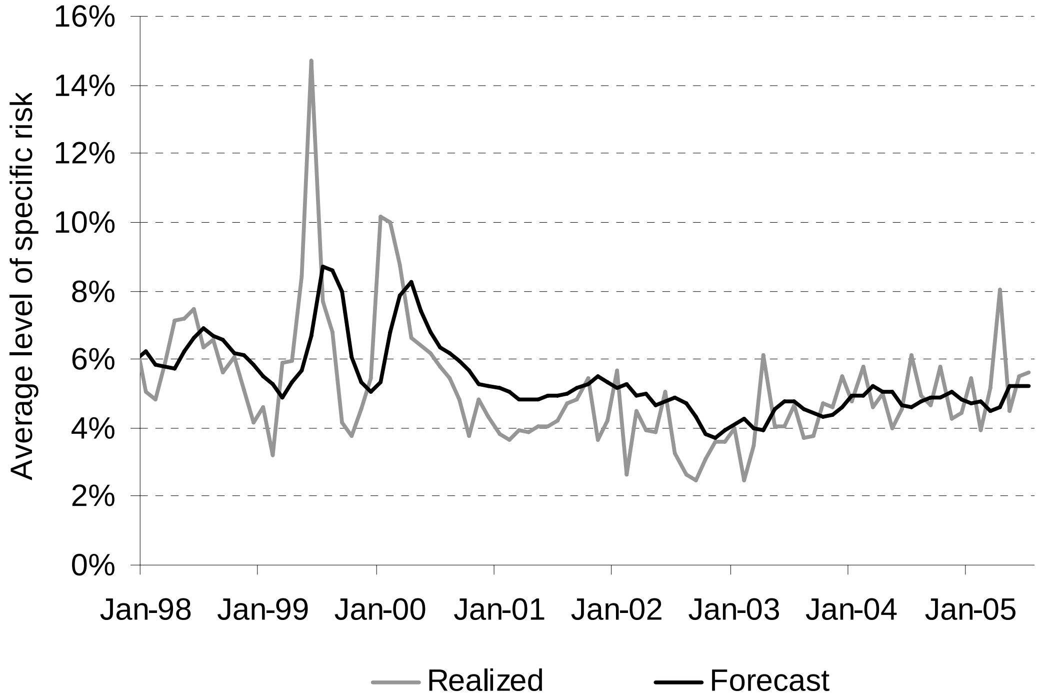
The adjustment $\hat{V}_{jt}$ makes the forecast for asset j higher or lower than the market average. The forecast adjustment is a simple sum of influences from a set $\{Z_k\}$ of characteristics that include the model factors, the asset's historical sigma (the standard deviation of its non-market risk), its exponentially weighted average absolute return over the last three months, and its most recent monthly return. The forecast is

$$\hat{V}_{jt} = \sum_k \gamma_k Z_{kjt} .$$

CHE2 re-estimates the sensitivities γ with regressions of $V_{jt} = \left(\left|u_{jt}\right| - S_t\right) / S_t$ against the sensitivities Z_{kjt} every month.

The scaling factor κ_j is termed the kurtosis correction. It transforms the forecast of expected absolute specific return to a forecast of specific return standard deviation. The kurtosis correction depends on the capitalization of the asset, and on whether it is an A-share or a B-share.

Figure 4: Forecast and realized market-averaged specific risk levels.



7.2 Linking the Specific Returns of Issues from the Same Firm

Equity issues from the same firm often are linked more closely than one would expect simply on the basis of their similarities in factor exposures. In China firms rarely (to date, never) have multiple issues, but sometimes a single firm will have both an A-share and a B-share issue. As a refinement, the CHE2 specific risk model "links" these issues. CHE2 specifies a non-zero covariance between the specific risks of issues from the same firm by assuming that there is a single "root" asset (the A-share issue). The root asset has a specific return u_r. The specific return u_j of each linked issue j is expressed as a historically estimated beta times the root specific return, plus a residual s_j that is not related to any other issue: $u_j = \beta_{j:r} u_r + s_j$. The cross-asset specific returns covariance $\text{cov}(u_j, u_k)$ is given by the specific return variance of the root asset times a product of the respective betas: $\text{cov}(u_j, u_k) = \sigma_r^2 \beta_{j:r} \beta_{k:r}$. The beta of the root is equal to 1. Thus, the covariances of same-firm A-share and B-share specific returns are equal to the A-share specific return variance times the B-share's specific return beta against the A-share.

The performance of the specific risk forecast at the portfolio level reflects both the accuracy of the forecast at the asset level and the success of the model factors in removing all non-diversifying factor return from the attributed specific returns. The next section explores the quality of CHE2's portfolio risk forecasts.

8 Portfolio Risk Forecast Performance

Section 6 described how CHE2 predicts factor return covariances. It then reviewed the accuracy of CHE2's factor risk forecasts. This section examines the performance of CHE2's portfolio-level risk forecasts. The tool used to quantify portfolio-level performance is the bias statistic, introduced in Section 6. Portfolio risk forecast bias results are reported in two periods, 1998/01–2001/06 and 2001/07–2005/06.

Several types of portfolios are tested. They belong to three broad classes. The first class has no strong dependence on the model factor structure. It includes the capitalization-weighted estimation universe (also referred to in this paper as the market), a capitalization-weighted portfolio of the 50 largest A-share issues, a fully invested portfolio chosen from the estimation universe to minimize risk, and random portfolios. Each random portfolio consists of 40 equally weighted issues, selected at random from the top 70% of issues by capitalization. The random portfolios tend to have negative Size exposures and random exposures to other factors. The second class of portfolios is based on CHE2's industry structure. These industry tilt portfolios are equally weighted collections of either the largest 20% of issues or the smallest 20% of the issues by capitalization within an industry. Risk index tilt portfolios comprise the third class. The top and bottom quintiles of each risk index by factor exposure make up these equally weighted portfolios. They explore the consequences of extreme risk index exposure.

Table 11 summarizes the portfolio bias test results. Because of the large number of tilt and random portfolios, only summary results are shown for them. Since some industry portfolios consist only of a few assets, especially in the first test period, not all industry portfolios are included in the summary results. To be included, an industry tilt portfolio must contain five or more names for at least 30 months during the sample period. The table is divided into three panels, reflecting performance in total, common factor, and specific return risk. Both the full portfolio return and active return risk forecasts are tested. The active return is the difference in return between the portfolio in question and a benchmark. For the purposes of these bias tests, the capitalization-weighted A-share estimation universe is adopted as the benchmark.

Panel A of Table 11 shows the bias statistics for the total risk forecasts. Overall, CHE2 forecasts the total risk of the test portfolios quite accurately. This is encouraging given the violent changes that have characterized market risk levels in China, especially in the earlier period. The risk of the estimation universe portfolio is correctly forecast in both sample periods. The same is true for the portfolio of the 50 largest A-shares and the random portfolios. Industry and risk index tilt portfolios generally perform well.

The minimum risk portfolio is conspicuously under-predicted in the first period — an outlying market return of 33.7% in June 1999 raises its bias statistic from 1.07 (without 1999/06) to 1.53. The effect of this large return is more pronounced in the minimum risk portfolio than in other portfolios by virtue of its construction; although it is fully invested it contains short positions that attempt to hedge as much risk as possible. The forecast risk of the portfolio in January 1999 is 5.3%, but the extraordinary returns of that month extended across all industries (industry factor returns ranged from 25% to 50%), breaking the hedge. The minimum risk portfolio's forecast is unbiased in the second period.

Panel B of Table 11 shows bias statistics for common factor risk in the test portfolios. Since the test portfolios are typically well diversified, specific return comprises only a small part of the total return, and the results for common factor risk follow those of the total risk forecasts.

Panel C shows the results of the bias tests for the specific risk. Because the specific return is such a minor component of the overall return in most of the portfolios (especially the risk index tilts, which are equally weighted over several hundred assets), specific risk biases represent extremely stringent tests of both the model's attribution capabilities and its forecast quality. Although somewhat weaker for the risk index tilts than for the other categories, overall performance is very good in both periods.

In summary, despite the considerable challenges posed by the rapid growth and volatile nature of the Chinese equity markets, CHE2 risk forecasts at all levels — total, common factor, and specific — acquit themselves well.

9 Changes to Expect in Moving from CHE1 to CHE2

Users moving from CHE1 to CHE2 can expect to enjoy more accurate risk forecasts. Because CHE2's risk forecasts use daily data, CHE2 responds more readily to changes in risk than CHE1. CHE2's improved sensitivity to changing conditions also means that its users will see risk forecasts change more from month to month than with CHE1. In CHE1, the specific risk forecasts change readily and are essentially as reactive as in CHE2, but the common factor risk forecasting system adjusts slowly to change. In CHE2, the forecasting sensitivities of the common factor risk system and the specific risk system are much more closely balanced.

To explore the differences between the models, we start with forecasts for the total return risk of a simple equally weighted portfolio chosen at the beginning of August 2005 to hold the 100 largest A-share issues. A comparison of the CHE1 and CHE2 forecasts for this portfolio appears in Table 12. Both models agree that industry factors are the dominant contributors to risk. Risk indices play a minor role and specific risk is negligibly small. The difference between the two forecasts — 29.0% annualized standard deviation for CHE1, and 22.3% for CHE2 — is striking. It arises because the CHE1 factor covariance forecast uses a historical average with a 48-month half-life, while the risk levels in CHE2 are dominated by risk within the past year. CHE1 still "remembers" very high risk levels that prevailed half a decade ago, while CHE2 recognizes that risk levels have been significantly lower for the past several years.

Other items in the risk report deserve attention. The specific risk ("asset selection risk") forecasts of the two models are much closer than the forecasts of overall portfolio risk. This is because both models have specific risk sub-models that react quickly to change. The fact that CHE2's specific risk forecast is smaller than CHE1's by about one part in ten probably reflects the greater effectiveness of CHE2's factor system and estimation universe. CHE2 recognizes a greater proportion of asset-level return as common factor than CHE1, and consequently sees a smaller proportion as asset-specific.

Forecasts of total portfolio return risk emphasize the market return. Active returns typically emphasize non-market sources of risk.

Our second example compares forecasts for the active return risk of the equally weighted 100 asset portfolio. As a benchmark, we choose the 200 largest A-share issues and weight them equally. This choice of benchmark suggests that the active return will exhibit very little market risk but will receive a large contribution from the Size risk index. In accord with this expectation, both models assert that the active risk is dominated by risk index contributions, with a secondary contribution from asset-specific risk; the industry factors (and hence the market) are of little importance. Somewhat surprisingly given their very different risk forecasting mechanisms, the models agree closely in their active risk forecasts. CHE1 forecasts 4.06% annualized tracking error, while CHE2 forecasts 4.09%.

In general, we do not expect such close agreement in active risk forecasts. Over the period 2001/07–2005/06, CHE1 over-predicted risk in all of its risk indices except Momentum. This happened because its long factor covariance half-life impressed high historical levels on its risk forecasts. In contrast, over the same period none of the CHE2 risk index forecasts shows evidence of bias.

10 Conclusions

CHE2 is an advanced equity risk model designed to meet the challenges of the rapidly evolving Chinese market. It makes extensive use of daily data, which enable it to capture changes in the risk environment promptly and accurately. Its industry structure, based on the GICS standard, reflects current and anticipated industry emphases in China. Its style factors have been chosen carefully to represent major sources of common return among A- and B-shares. To improve the performance of its factors, Hong Kong-traded shares are excluded from CHE2; they are represented in a dedicated model for the Hong Kong market, HKE1.

Its design enables CHE2 to achieve superior performance. Its factors capture returns commonality very successfully; its average adjusted r-squared compares favorably with recent models for developed equity markets, and is nearly twice that of its predecessor, CHE1. Its forecasts of common factor risk, based on daily returns data, adjust quickly and appropriately to the large variations that characterize the Chinese equity market. This responsiveness is matched by the model's forecasts of asset-specific risk. The result is that portfolio-level risk is generally well predicted, even though risk levels have changed by factors of two or three over the model history.

Appendix A: Forecasting Long-term Risk with Daily Returns

To a good approximation, the return $f_k[t,t+N]$ to factor k over days $\{t,t+1,\ldots,t+N\}$ is the sum of the daily factor returns:

$$f_k[t,t+N]=\sum_{d=t}^{t+N} f_k[d,d]. \tag{0.1}$$

The covariance of returns to factors j and k over the same interval is consequently

$$\mathrm{cov}(f_j[t,t+N],f_k[t,t+N])=\sum_{d_1=t}^{t+N}\sum_{d_2=t}^{t+N}\mathrm{cov}(f_j[d_1,d_1],f_k[d_2,d_2]). \tag{0.2}$$

Now suppose that all covariances between factor j on day d and factor k on day $d+\Delta$ depend only on the separation in days Δ, and not on the particular days d and $d+\Delta$. Then it is possible to rewrite expression (0.2) in the more compact and convenient form,

$$\mathrm{cov}_{N+1}(j,k)=\sum_{\Delta=-N}^{N}(N+1-|\Delta|)\,\mathrm{cov}(f_j[0,0],f_k[\Delta,\Delta]). \tag{0.3}$$

Note that this accounts for the $N+1$ covariances between factor returns that occur on the same day, the N covariances between a return to factor j and a return to factor k one day later, etc.

If the daily factor returns were independent, their cross-day covariances would be equal to zero and only the same-day ($\Delta=0$) covariances would contribute to the sum. The result would be the familiar aggregation relation, $\mathrm{cov}_{N+1}(j,k)=(N+1)\mathrm{cov}_1(j,k)$.

But the cross-day covariances are not equal to zero, and forecasts are needed for the additional terms in equation (0.3). These can be prepared as exponentially weighted moving averages from a long history of daily factor returns. The weight assigned to day t is

$$w_t = \frac{2^{t/L}}{\sum_{\tau} 2^{\tau/L}}. \tag{0.4}$$

The sum in the denominator is over days in the factor return history, so that the sum of weights over the history is equal to 1. The quantity L is the *half-life* of the exponentially weighted average. (If the half-life is increased to infinity, the weights become equal.) The individual cross-day covariance estimates are

$$\mathrm{cov}(j,k;\Delta) \equiv \mathrm{cov}(f_j[0,0], f_k[\Delta,\Delta]) = \sum_t w_t \left(f_j[t,t] - \overline{f}_j\right)\left(f_k[t+\Delta, t+\Delta] - \overline{f}_k\right). \tag{0.5}$$

Here $\overline{f}$ denotes an average historical factor return, calculated with the weights w_t. The sum is once again over all days in the history. If factor returns occur in the sum that do not exist in the history they are set equal to zero. The cross-day covariances (0.5) enter the $N+1$ day covariance formula (0.3) to produce a multi-day covariance matrix, $\mathrm{cov}_{L,N+1}$. This matrix does not represent the final covariance forecast, however. It represents an intermediate stage in the production of the forecast.

One disadvantage of working with cross-day covariances is that since they link blocks of $N+1$ days, a historical sample of T days roughly contains only $T/(N+1)$ independent blocks of data. A rather long history may therefore be required to estimate multi-day covariances, especially if $N+1$ is large.

To address this difficulty, it is important to identify the smallest value of $N+1$ such that, to a sufficient approximation, two successive $N+1$ day returns are uncorrelated. In CHE2 a value $N+1=11$ has been chosen. Then the returns covariance matrix over a much longer horizon H may be written as $H/(N+1)$ times the $N+1$ day covariance matrix:

$$\mathrm{cov}_{L,H} = \frac{H}{N+1}\mathrm{cov}_{L,N+1}.$$

This avoids incorporating many cross-day covariances that are so small that they cannot be reliably estimated in the estimation of $\text{cov}_{L,H}$. Nevertheless it still fails to fully exploit the high density of daily data, since with *N*+1=11 each month contains only two independent *N*+1 day intervals.

Several years of data may be needed to accurately model the effects of serial correlations among factor returns. Over such a long time, factor volatilities may have changed substantially. Evidence suggests that although daily factor return variances can change significantly and rapidly, factor *correlations* vary more slowly. The correlations are just the covariances in expression (0.5), divided by the standard deviations of the daily returns:

$$\begin{aligned}\text{corr}(j,k;\Delta) &= \frac{1}{\sqrt{\text{cov}(j,j;0)}}\text{cov}(\text{j,k};\Delta)\frac{1}{\sqrt{\text{cov}(k,k;0)}} \\ &\equiv \frac{1}{\sigma_{L,j}}\text{cov}(\text{j,k};\Delta)\frac{1}{\sigma_{L,k}}.\end{aligned} \tag{0.6}$$

The standard deviations $\sigma_{L,j}$ and $\sigma_{L,k}$ come from variances that use the same long half-life L and weights w_t employed in the covariance estimates (0.5). To recover covariances that agree with *recently* realized levels of daily factor volatility, all that is necessary is to multiply the correlations by the appropriate standard deviations — standard deviations $\sigma_{S,j}$ and $\sigma_{S,k}$ that depend on a short half-life, S, also called the volatility half-life. In CHE2 the volatility half-life S is equal to 90 trading days, or a little over four months.

The end effect of these manipulations is to take the "long half-life" covariances and replace them with updated covariances,

$$\text{cov}_H(\text{j,k}) = \frac{\sigma_{S,j}}{\sigma_{L,j}}\text{cov}_{L,H}(\text{j,k})\frac{\sigma_{S,k}}{\sigma_{L,k}}. \tag{0.7}$$

The new covariances agree with long-term factor correlations and yet also fully reflect recent risk levels.

There is a last technical refinement in the construction of the factor covariance forecast. It is not hard to show that when the weights w_t are all equal, formula (0.3) will produce a positive semi-definite covariance matrix without fail; i.e., the method always will forecast returns variances that are equal to or greater than zero (see Newey and West 1987: *Econometrica* **55**, pp. 703-708). This is a necessary property for any covariance matrix. Nevertheless, equal weights are not generally desirable. Weights that gradually become smaller for days farther in the past are preferable, so that very old data are treated as less relevant than more recent data. Exponential weighting achieves this. When the exponential weighting half-life L is large but finite, it is possible (although highly unlikely) that the covariance matrix cov_H will forecast one or more small, negative variances. As a precaution, if a negative eigenvalue (the cause of negative variance forecasts) appears in the matrix cov_H, it is replaced by a very small positive number. This positive semi-definite matrix is the factor covariance forecast in its final form.

Appendix B: Descriptor Definitions

Size

ASSI Natural logarithm of total assets.

$$ASSI = \ln(\text{total assets})$$

Total assets are from the balance sheet of the most recent available annual report.

LCAP Natural logarithm of total market capitalization.

$$LCAP = \ln\left(\text{price per share} \times \text{total shares outstanding}\right)$$

Price per share is the closing price on the last day of the previous month, and *total shares outstanding* is the total number of tradable and non-tradable shares at the end of the previous month.

LFLO Natural logarithm of float market capitalization.

$$LFLO = \ln\left(\text{price per share} \times \text{tradable shares}\right)$$

Price per share is the closing price on the last day of the previous month, and *tradable shares* is the number of shares available through the open market at the end of the previous month.

Volatility

CMRA Monthly cumulative return range over the past 24 months.

$$CMRA = \ln(\text{max cumret}) - \ln(\text{min cumret})$$

Max cumret is the maximum of the cumulative return and *min cumret* is the minimum of the cumulative return during the last 24 months. The cumulative return is defined as follows:

$$cumret_{i,t} = \prod_{s=1}^{t}(1 + r_{i,s}) - \prod_{s=1}^{t}(1 + r_{rf,s})$$

Here $r_{i,s}$ is the monthly stock return and $r_{rf,s}$ is the monthly risk-free rate.

DHILO Median difference between the log of the high price and the log of the low price for each day during the past three months.

$$DHILO = \text{median}\{\ln(\text{high price}) - \ln(\text{low price})\}$$

High price and *low price* refer to daily intraday prices.

DVRAT Daily returns variance ratio – serial dependence in daily returns.

$$DVRAT = \frac{\sigma_q^2}{\sigma^2} - 1$$

Let $r_{i,t}$ be the daily stock return, $r_{rf,t}$ be the daily risk free rate, and $e_{i,t} = r_{i,t} - r_{rf,t}$ be the excess stock return. Then:

$\sigma^2 = \frac{1}{T-1}\sum_{t=1}^{T} e_{i,t}^2$ - the variance of daily excess returns,

$\sigma_q^2 = \frac{1}{m}\sum_{t=q}^{T}\left(\sum_{j=1}^{q} e_{i,t+1-j}\right)^2$, where $m = q\left(T-q+1\right)\left(1-\frac{q}{T}\right)$ - the variance of the cumulative q day excess return divided by q. The sample size T is the number of days in the last 24 months and q is set to 10. An interpretation and a discussion of the properties of this statistic appear in Lo and MacKinlay (1999).

HBETA Historical daily beta.

Beta with respect to the market portfolio (CAPM beta) calculated using daily returns for the last 12 months and corrected for three lags in autocorrelation.

HSIGMA Historical daily sigma.

Sigma of the CAPM regression calculated using daily returns for the last 12 months and corrected for the three lags in autocorrelation.

Downside Volatility

DDNBT Downside beta.

Beta with respect to the market portfolio, calculated using only days from the previous twelve months when the market return was negative.

DDNCR Downside correlation.

Correlation of the stock return with the market portfolio return, calculated using only days from the previous twelve months when the market return was negative.

DDNSR Ratio of downside standard deviations.

Ratio of the stock return standard deviation to the market return standard deviation, both calculated using only days from the previous twelve months when the market return was negative.

Momentum

RSTR12 Relative strength for the last 12 months.

$$RSTR12 = \sum_{t=1}^{12} \ln\left(1 + r_{i,t}\right) - \sum_{t=1}^{12} \ln\left(1 + r_{rf,t}\right)$$

Here $r_{i,t}$ is the monthly stock return and $r_{rf,t}$ is the monthly risk-free rate.

RSTR24 Relative strength for the last 24 months.

$$RSTR24 = \sum_{t=1}^{24} \ln\left(1 + r_{i,t}\right) - \sum_{t=1}^{24} \ln\left(1 + r_{rf,t}\right)$$

Here $r_{i,t}$ is the monthly stock return and $r_{rf,t}$ is the monthly risk-free rate.

Trading Activity

MNDTO3 Mean daily turnover during the last 3 months.

$$MNDTO = mean\left\{\frac{\text{trading volume}}{\text{tradable shares}}\right\}$$

Trading volume is the number of shares traded on a given day, *tradable shares* is the number of tradable shares on that day.

MNDTO6 Mean daily turnover during the last 6 months.
See description of MNDTO3.

MNDTO12 Mean daily turnover during the last 6 months.
See description of MNDTO3.

TOBT Liquidity – turnover beta.
Let $r_{i,t}$ and $r_{M,t}$ be daily stock and market returns, respectively. Let $r_{rf,t}$ be the daily risk free rate, and $e_{i,t} = r_{i,t} - r_{rf,t}$ and $e_{M,t} = r_{M,t} - r_{rf,t}$ be the excess stock and market returns, respectively. Let $TO_{i,t}$ be the stock's daily turnover (see MNDTO3 for definition). The following regression uses daily data for the last 24 months:

$$\left|e_{i,t}\right| = \alpha_i + \beta_i TO_{i,t} + \sum_{j=1}^{5} \rho_j \left|e_{i,t-j}\right| + \sum_{j=1}^{5} \lambda_j \left|e_{M,t-j}\right| + \varepsilon_{i,t}$$

The turnover coefficient $\hat{\beta}_i$ is a measure of liquidity.

Value

BTOP Book to price.

$$BTOP = \frac{\text{book equity}}{\text{market capitalization}}$$

Book equity is the book value of equity from the most recent annual report. *Market capitalization* is the market value of tradable shares at the end of the last month. If a company issues both A and B shares, the market value is aggregated across the two types of shares.

CTOP Cash flow to price.

$$CTOP = \frac{\text{cash flow}}{\text{market capitalization}}$$

Cash flow is the common stock dividend from the most recent annual report. *Market capitalization* is the market value of tradable shares at the end of the last month. If a company issues both A and B shares, the market value is aggregated across the two types of shares.

CTP5 Five year average cash flow to price.

$$CTP5 = \frac{\frac{1}{5}\sum_{t=1}^{5} \text{cash flow}_t}{\frac{1}{5}\sum_{t=1}^{5} \text{market capitalization}_t}$$

Cash flow is the common stock dividend from the annual report of year *t*. *Market capitalization* is the market value of tradable shares at the end of year *t*. If a company issues both A and B shares, the market value is aggregated across the two types of shares.

ETOP Earnings to price.

$$ETOP = \frac{\text{earnings}}{\text{market capitalization}}$$

Earnings is the net profit from the most recent annual report. *Market capitalization* is the market value of tradable shares at the end of the last month. If a company issues both A and B shares, the market value is aggregated across the two types of shares.

ETP5 Five year average cash flow to price.

$$ETP5 = \frac{\frac{1}{5}\sum_{t=1}^{5} \text{earnings}_t}{\frac{1}{5}\sum_{t=1}^{5} \text{market capitalization}_t}$$

Earnings is the net profit from the annual report of year *t*. *Market* *capitalization* is the market value of tradable shares at the end of year *t*. If a company issues both A and B shares, the market value is aggregated across the two types of shares.

Growth

EGRO Five-year earnings growth.

$$EGRO = \frac{\hat{\beta}}{\frac{1}{5}\sum_{t=1}^{5} \text{earnings}_t}$$

Here $\hat{\beta}$ is the time trend coefficient from the following regression:

$$\text{earnings}_t = \hat{\alpha} + \hat{\beta} \cdot \text{time}_t$$

The regression is estimated using earnings for the past 5 years; a minimum of 3 years of data are required.

ROA Return on assets.

$$ROA = \frac{\text{earnings}}{\text{total assets}}$$

Earnings is the net profit from the most recent annual report. *Total assets* come from the previous annual report.

ROA5 Five-year average return on assets.

$$ROA5 = \frac{\frac{1}{5}\sum_{t=1}^{5} \text{earnings}_t}{\frac{1}{5}\sum_{t=1}^{5} \text{total assets}_{t\text{-}1}}$$

Earnings is the net profit from the annual report of year *t*. The *total assets* applied to the earnings of year *t* come from the annual report of year *t*-1.

ROE Return on assets.

$$ROE = \frac{\text{earnings}}{\text{book equity}}$$

Earnings is the net profit from the most recent annual report. *Book equity* is the book value of equity from the previous annual report.

ROE5 Five-year average return on assets.

$$ROE5 = \frac{\frac{1}{5}\sum_{t=1}^{5}\text{earnings}_t}{\frac{1}{5}\sum_{t=1}^{5}\text{book equity}_{t\text{-}1}}$$

Earnings is the net profit from the annual report of year *t*. The *book equity* applied to the earnings of year *t* comes from the annual report of year *t*-1.

SUE Standardized unexpected earnings.

$$SUE = \frac{\text{earnings-mean(earnings)}}{\text{std(earnings)}}$$

Earnings is the net profit from the most recent annual report, year *t*. The mean and standard deviation are calculated using earnings for the preceding 4 years, *t*-1 to *t*-5, with a required minimum of 2 years of data.

Leverage

BLEV Book leverage.

$$BLEV = \frac{\text{long term debt}}{\text{long term debt + book equity}}$$

The accounting items are from the most recent annual report.

DTOA Debt to total assets.

$$DTOA = \frac{\text{long term debt + debt in current liabilities}}{\text{total assets}}$$

The accounting items are from the most recent annual report.

MLEV Market leverage.

$$MLEV = \frac{\text{long term debt}}{\text{long term debt + market capitalization}}$$

The accounting items are from the most recent annual report. *Market capitalization* is the market value of tradable shares at the end of the last month. If the company issues both A and B shares, the market value is aggregated across the two types of shares.

Table 1: Number of listed issues and the size of the estimation universe.

Month	A-shares	ESTU	ESTU as percent of A-shares	B-shares	B-shares used in B-share factors	Estimation B-shares as percent of all B-shares
Dec-98	826	792	95.9%	106	102	96.2%
Dec-99	923	859	93.1%	108	97	89.8%
Dec-00	1,060	958	90.4%	114	98	86.0%
Dec-01	1,136	1,052	92.6%	112	100	89.3%
Dec-02	1,200	1,095	91.3%	111	97	87.4%
Dec-03	1,263	1,122	88.8%	111	88	79.3%
Dec-04	1,353	1,201	88.8%	110	86	78.2%
Jun-05	1,367	1,197	87.6%	110	90	81.8%

Table 2: Red Chip and H-share portfolio returns regressed against HK and A-share returns. (Standard errors are in parentheses.)

Portfolio	A-shares	Hong Kong	R-squared
	Sample period: 1998/01 - 2005/06		
Red Chips	0.13	1.43	74.2%
	(0.11)	(0.09)	
H-shares	0.46	1.12	50.2%
	(0.15)	(0.14)	
	Sample period: 1998/01 - 2001/06		
Red Chips	0.17	1.50	75.1%
	(0.20)	(0.14)	
H-shares	0.66	1.17	52.7%
	(0.28)	(0.20)	
	Sample period: 2001/07 - 2005/06		
Red Chips	0.12	1.18	74.0%
	(0.10)	(0.11)	
H-shares	0.25	0.91	46.5%
	(0.14)	(0.17)	

Table 3: Tests of statistical significance for the Steel industry.

Start Date	End Date	Num Obs	F-stat	Proportion |t-stat|>2
1998/01	2004/10	82	3.21	18.3%
1998/01	2001/06	42	1.69	19.1%
2001/07	2004/10	40	4.82	17.5%

Table 4: The CHE2 and CHE1 industry structures.

Sect. No	GICS Sector	Ind. No	CHE2 Industry
1	Energy	1	Energy
2	Materials	2	Chemicals
		3	Steel
		4	Metals & Mining
		5	Other Materials
3	Industrials	6	Industrial Conglomerates
		7	Capital Goods ex Conglomerates
		8	Commercial Services
		9	Transportation Infrastructure
		10	Other Transportation
4	Consumer Discretionary	11	Automobiles & Components
		12	Household Durables
		13	Apparel, Luxury & Leisure Products
		14	Media
		15	Hotels, Restaurants, & Leisure
		16	Retailing
5	Consumer Staples	17	Consumer Staples
6	Health Care	18	Health Care
7	Financials	19	Real Estate
		20	Other Financials
8	Information Technology	21	Software & Services
		22	IT (Hardware & Equipment)
9	Telecommunication Services	23	Telecommunication Services
10	Utilities	24	Utilities

Ind. No	CHE1 Industry
1	Miscelaneous
2	Automobile
3	Chemicals
4	Consumer Products
5	Diversified
6	Electronics
7	Energy
8	Finance
9	Food & Beverage
10	Healthcare
11	Information Technology
12	Infrastructure
13	Manufacturing
14	Metals & Mining
15	Real Estate
16	Retail
17	Telecommunications
18	Tourism
19	Transportation

Table 5: Significance of the ex-market industry factors, January 1998 – June 2005.

Factor	F-stat	Prop |t|>2	Beta	T-stat (Beta)
Market	745.04	97.78	1.01	78.36
Information Technology	10.72	44.44	0.20	4.88
Steel	6.13	36.67	-0.09	-2.13
Metals & Mining	5.32	21.11	0.05	1.04
Utilities	4.96	27.78	-0.08	-2.14
Transportation Infrastructure	4.87	23.33	-0.20	-5.04
Financials	4.84	25.56	0.03	0.36
Automobiles & Components	4.25	26.67	-0.05	-1.52
Real Estate	4.13	23.33	0.05	1.54
Software & Services	3.59	24.44	0.12	2.27
Health Care	3.26	23.33	-0.03	-1.06
Chemicals	3.00	24.44	0.00	-0.20
Media	2.88	23.33	0.05	0.67
Household Durables	2.79	25.56	-0.04	-1.09
Energy	2.78	23.33	-0.03	-0.54
Retailing	2.73	23.33	0.02	0.72
Transportation	2.31	17.78	-0.06	-1.38
Apparel, Luxury & Leisure Products	2.26	18.89	-0.04	-1.63
Materials	1.83	16.67	-0.10	-3.67
Industrial Conglomerates	1.65	16.67	0.13	4.66
Hotels Restaurants & Leisure	1.63	14.44	0.02	0.67
Consumer Staples	1.62	6.67	-0.05	-2.80
Capital Goods	1.44	10.00	0.00	0.41
Commercial Services & Supplies	1.30	8.89	0.00	-0.03
Telecommunication Services*	1.25	6.45	-0.13	-0.78

* Note: Telecommunication Services first appears in January 2003.

Table 6: CHE2 and CHE1 risk factor compositions.

CHE2 Risk Indices and descriptors	CHE1 Risk Indices and descriptors
Size	**Size**
ASSI - Log of total assets	Log of total market capitalization
LCAP - Log of total market capitalization	
LFLO - Log of float market capitalization	
Volatility	**Volatility**
CMRA - Cumulative return range in the last 24 months	Cumulative return range in the last 12 months
DHILO - Daily high price minus low price*	Historical sigma
DVRAT - Daily return variance ratio*	Squared return of the previous month
HBETA - Historical beta*	
HSIGMA - Historical sigma*	
	Market Sensitivity
	Historical beta
	Historical beta times historical sigma
Downside Volatility	
DDNBT - Downside beta*	
DDNCR - Downside correlation*	
DDNSR - Downside volatility ratio*	
Momentum	**Momentum**
RSTR12 - Relative strength, past 12 months	Historical alpha
RSTR24 - Relative strength, past 24 months	Relative strength
Trading Activity	
MNDTO3 - Mean daily turnover, 3 months*	
MNDTO6 - Mean daily turnover, 6 months*	
MNDTO12 - Mean daily turnover, 12 months*	
TOBT - Liquidity, turnover beta*	

* Descriptors that use daily data.

Table 6 (continued).

CHE2 Risk Indices and descriptors	CHE1 Risk Indices and descriptors
Value	**Value**
BTOP - Book to price	Current earnings to price
CTOP - Cash flow (dividend) to price	Book to price
CTP5 - Five year average cash flow to price	
ETOP - Earnings to price	
ETP5 - Five year average earnings to price	
	Yield
	Current dividend yiels
	Indicator of extremely low yield
Growth	
EGRO - Five year earnings grouth	
ROA - Return on assets	
ROA5 - Five year average return on assets	
ROE - Return on equity	
ROE5 - Five year average return on equity	
SUE - Standardized unexpected earnings	
Leverage	
BLEV - Book leverage	
DTOA - Debt to total assets	
MLEV - Market leverage	
Stock Exchange Indicator	**Stock Exchange Indicator**
SZSE - Shenzheng stock exchange indicator	Hong Kong stock exchange indicator
B-share Residual Factors	
BUNIV - B-share universe indicator	
BSZSE - B-shares traded on SZSE indicator	

Table 7: Performance of the CHE2 risk factors.

Factor	F-stat	Prop |t|>2 (%)	Sigma (Ann %)	Alpha (Ann %)	T-stat (Alpha)	Beta	T-stat (Beta)
			<u>Sample period: 1997/07 - 2001/06</u>				
SIZE	17.64	72.92	5.21	***-12.61***	***-4.81***	0.04	1.64
VOLATIL	8.77	50.00	4.59	***-5.84***	***-2.53***	0.01	0.37
GROWTH	7.59	56.25	4.23	***-5.35***	***-2.52***	***0.06***	***2.70***
SZSE	7.44	41.67	7.33	-3.96	-1.07	***0.09***	***2.34***
MOMENTUM	5.37	31.25	4.24	***4.69***	***2.20***	0.02	0.96
TRADACTV	4.16	39.58	3.04	***-4.96***	***-3.24***	-0.02	-1.40
VOLDOWN	3.75	27.08	2.60	0.72	0.55	***0.05***	***3.56***
VALUE	3.05	25.00	2.94	1.27	0.86	***-0.07***	***-4.48***
LEVERAGE	1.55	12.50	1.65	0.32	0.39	0.00	0.30
			<u>Sample period: 2001/07 - 2005/06</u>				
MOMENTUM	20.10	64.58	4.51	1.76	0.74	***-0.07***	***-2.22***
SIZE	13.02	58.33	3.40	2.11	1.18	0.03	1.08
VOLDOWN	11.14	47.92	2.62	***4.31***	***3.13***	***0.11***	***5.85***
VOLATIL	9.48	47.92	2.69	1.50	1.06	***0.07***	***3.78***
TRADACTV	7.09	45.83	1.94	***-2.63***	***-2.58***	***0.04***	***2.61***
GROWTH	5.52	29.17	2.17	1.42	1.25	-0.02	-1.18
VALUE	4.01	29.17	1.92	***3.48***	***3.44***	0.00	0.25
SZSE	3.24	31.25	2.57	1.02	0.75	0.03	1.61
LEVERAGE	1.43	8.33	0.79	***1.32***	***3.15***	0.01	1.67

Table 8: Performance of B-share residual factors.

Factor	F-stat	Prop |t|>2 (%)	Sigma (Ann %)	Alpha (Ann %)	T-stat (Alpha)	Beta	T-stat (Beta)
			<u>Sample period: 1997/07 - 2001/06</u>				
BUNIV	375.11	72.92	52.48	6.54	0.25	0.24	0.88
BSZSE	11.85	52.08	22.91	-10.24	-0.89	0.01	0.12
			<u>Sample period: 2001/07 - 2005/06</u>				
BUNIV	47.91	58.33	17.80	-2.43	-0.26	-0.08	-0.62
BSZSE	15.79	52.08	14.99	10.46	1.33	0.08	0.76

Table 9: Correlations between risk index returns. The period 1998/01–2005/06 is below the diagonal, and 2001/07–2005/06 is above the diagonal.

	SIZE	VOL	VOLD	TRAD	MOM	LEV	VAL	GRO	SZSE
SIZE		1.1%	19.1%	41.3%	-8.5%	23.8%	-5.6%	2.6%	6.9%
VOLATIL	-6.9%		-2.0%	-9.9%	-19.7%	-11.2%	-29.1%	-20.8%	18.6%
VOLDOWN	7.2%	-13.7%		28.1%	-38.1%	24.5%	5.7%	-6.9%	4.4%
TRADACTV	2.5%	-2.8%	0.9%		1.5%	-8.9%	12.1%	15.3%	8.5%
MOMENTUM	8.9%	-16.5%	-26.3%	-5.3%		-3.7%	17.4%	46.4%	-15.2%
LEVERAGE	-4.2%	-11.8%	17.7%	-8.3%	-24.6%		21.0%	-21.4%	7.5%
VALUE	-23.9%	13.1%	-18.1%	23.6%	-8.0%	3.1%		-16.8%	-3.9%
GROWTH	35.0%	-15.5%	0.6%	7.6%	33.9%	-7.1%	-37.6%		-21.0%
SZSE	20.7%	-9.3%	2.5%	3.0%	11.5%	-4.4%	-19.2%	31.5%	

Table 10: Bias tests for factor risk and factor residual risk.

Factor	Total Risk 1998/01-2001/06	Total Risk 2001/07-2005/06	Residual Risk 1998/01-2001/06	Residual Risk 2001/07-2005/06
		Risk Factors		
GROWTH	0.873	1.042	0.878	1.047
LEVERAGE	1.005	0.874	1.025	0.891
MOMENTUM	0.960	1.104	1.106	1.107
SIZE	1.138	1.114	1.143	1.123
SZSE	***0.710***	0.976	***0.711***	0.968
TRADACTV	0.944	0.942	1.027	0.906
VALUE	***1.356***	0.974	***1.314***	1.006
VOLATIL	1.017	0.887	1.070	0.830
VOLDOWN	0.814	0.969	***0.738***	0.935
BUNIV*	1.111	0.809		
BSZSE*	0.786	0.869		
		Industry Factors		
APPARELS	1.041	1.013	1.075	1.058
AUTO	1.028	1.067	1.026	***1.289***
CAPITAL	1.107	0.982	***0.744***	0.828
CHEMICAL	1.051	1.035	0.904	1.129
COMMERCL	0.942	0.965	0.980	1.128
CONGLOM	***1.260***	0.973	***1.383***	0.810
CONSTPL	1.028	1.018	***0.732***	0.996
DURABLES	1.020	0.997	0.948	1.037
ENERGY	1.078	1.040	1.069	1.016
FINANCE	1.058	1.041	1.037	1.185
HEALTH	1.059	0.980	0.906	1.167
HOTELS	1.131	0.995	0.921	1.188
INFRASTR	0.878	0.883	0.900	1.002
IT	***1.229***	1.028	***1.271***	1.203
MATERIAL	0.959	1.033	0.925	0.882
MEDIA	1.152	0.893	1.200	1.139
METALS	1.202	1.144	1.067	***1.529***
REALEST	0.994	0.937	0.988	1.098
RETAIL	1.137	0.960	0.909	1.031
SOFTWARE	1.156	0.965	1.028	1.027
STEEL	1.068	1.025	1.006	1.169
TELECOM		0.993		***1.990***
TRANSPT	1.013	0.981	0.874	1.003
UTILITY	0.927	1.031	1.203	***1.235***

* Note: For the sample period 1998/01-2001/06, the month of 2001/03 is excluded.

Table 11: Portfolio bias tests.

Panel A. Total Risk Bias Statistics

Portfolios	Top/Bottom	Active	Nuber of Portfolios	Mean Bias	Correct	Underpred.	Overpred.
			Sample period: 1998/01 - 2001/06				
Estimation		NO	1	1.078	1	0	0
Top 50 capt		NO	1	1.170	1	0	0
Min. risk		NO	1	1.528	0	1	0
Random		NO	30	1.045	30	0	0
Industry	Top	NO	13	1.091	11	2	0
Industry	Bottom	NO	13	0.999	13	0	0
Riskindex	Top	NO	8	1.024	8	0	0
Riskindex	Bottom	NO	8	1.036	8	0	0
Top 50 capt		YES	1	0.982	1	0	0
Min. risk		YES	1	0.546	0	0	1
Random		YES	30	0.987	28	0	2
Industry	Top	YES	13	1.086	11	1	1
Industry	Bottom	YES	13	1.068	11	1	1
Riskindex	Top	YES	8	1.018	7	0	1
Riskindex	Bottom	YES	8	1.012	7	0	1
			Sample period: 2001/07 - 2005/06				
Estimation		NO	1	0.975	1	0	0
Top 50 capt		NO	1	0.970	1	0	0
Min. risk		NO	1	1.172	1	0	0
Random		NO	30	0.982	30	0	0
Industry	Top	NO	15	1.024	14	1	0
Industry	Bottom	NO	15	0.953	15	0	0
Riskindex	Top	NO	8	0.991	8	0	0
Riskindex	Bottom	NO	8	0.962	8	0	0
Top 50 capt		YES	1	1.052	1	0	0
Min. risk		YES	1	0.915	1	0	0
Random		YES	30	0.997	30	0	0
Industry	Top	YES	15	1.116	11	4	0
Industry	Bottom	YES	15	0.996	13	1	1
Riskindex	Top	YES	8	1.049	7	1	0
Riskindex	Bottom	YES	8	1.077	8	0	0

Table 11 (continued).

Panel B. Common Risk Bias Statistics

Portfolios	Top/Bottom	Active	Nuber of Portfolios	Mean Bias	Correct	Underpred.	Overpred.
			Sample period: 1998/01 - 2001/06				
Estimation		NO	1	1.069	1	0	0
Top 50 capt		NO	1	1.148	1	0	0
Min. risk		NO	1	1.677	0	1	0
Random		NO	30	1.042	30	0	0
Industry	Top	NO	13	1.106	11	2	0
Industry	Bottom	NO	13	0.975	13	0	0
Riskindex	Top	NO	8	1.023	8	0	0
Riskindex	Bottom	NO	8	1.026	8	0	0
Top 50 capt		YES	1	1.041	1	0	0
Min. risk		YES	1	0.689	0	0	1
Random		YES	30	1.017	28	2	0
Industry	Top	YES	13	1.093	11	2	0
Industry	Bottom	YES	13	1.123	10	3	0
Riskindex	Top	YES	8	1.047	6	1	1
Riskindex	Bottom	YES	8	0.971	7	0	1
			Sample period: 2001/07 - 2005/06				
Estimation		NO	1	0.985	1	0	0
Top 50 capt		NO	1	1.024	1	0	0
Min. risk		NO	1	1.162	1	0	0
Random		NO	30	0.979	30	0	0
Industry	Top	NO	15	1.020	14	1	0
Industry	Bottom	NO	15	0.977	15	0	0
Riskindex	Top	NO	8	0.995	8	0	0
Riskindex	Bottom	NO	8	0.963	8	0	0
Top 50 capt		YES	1	1.089	1	0	0
Min. risk		YES	1	0.975	1	0	0
Random		YES	30	1.022	30	0	0
Industry	Top	YES	15	1.112	10	5	0
Industry	Bottom	YES	15	1.027	15	0	0
Riskindex	Top	YES	8	1.043	8	0	0
Riskindex	Bottom	YES	8	1.083	7	1	0

Table 11 (continued).

Panel C. Specific Risk Bias Statistics

Portfolios	Top/Bottom	Active	Nuber of Portfolios	Mean Bias	Correct	Underpred.	Overpred.
			Sample period: 1998/01 - 2001/06				
Estimation		NO	1	0.532	0	0	1
Top 50 capt		NO	1	1.090	1	0	0
Min. risk		NO	1	0.604	0	0	1
Random		NO	30	1.015	27	3	0
Industry	Top	NO	13	0.916	11	1	1
Industry	Bottom	NO	13	1.051	12	1	0
Riskindex	Top	NO	8	0.730	0	1	7
Riskindex	Bottom	NO	8	0.946	4	1	3
Top 50 capt		YES	1	1.203	1	0	0
Min. risk		YES	1	0.572	0	0	1
Random		YES	30	1.043	27	3	0
Industry	Top	YES	13	0.931	10	1	2
Industry	Bottom	YES	13	1.046	12	1	0
Riskindex	Top	YES	8	0.804	3	1	4
Riskindex	Bottom	YES	8	0.929	6	0	2
			Sample period: 2001/07 - 2005/06				
Estimation		NO	1	0.752	0	0	1
Top 50 capt		NO	1	1.378	0	1	0
Min. risk		NO	1	0.784	0	0	1
Random		NO	30	1.004	28	1	1
Industry	Top	NO	15	0.948	14	0	1
Industry	Bottom	NO	15	1.102	12	3	0
Riskindex	Top	NO	8	0.946	4	1	3
Riskindex	Bottom	NO	8	1.153	5	3	0
Top 50 capt		YES	1	1.494	0	1	0
Min. risk		YES	1	0.840	1	0	0
Random		YES	30	1.056	28	2	0
Industry	Top	YES	15	0.972	14	0	1
Industry	Bottom	YES	15	1.077	12	3	0
Riskindex	Top	YES	8	1.045	5	2	1
Riskindex	Bottom	YES	8	1.063	6	1	1

Table 12: CHE2 and CHE1 risk decompositions for 1 August 2005.

Source	CHE2 Risk Decomposition		CHE1 Risk Decomposition	
	Risk (% Ann Std Dev)	Contribution (% Active Risk)	Risk (% Ann Std Dev)	Contribution (% Active Risk)
	*100 largest ESTU stocks, cash as benchmark**			
Risk Indices	6.46	8.40	2.05	0.50
Industries	23.93	115.13	29.37	102.82
Covariance * 2	N/A	-24.59	N/A	-4.12
Asset Selection	2.30	1.06	2.58	0.79
Active	22.30		28.97	
Benchmark	0.00		0.00	
Total	22.30		28.97	
	*100 largest ESTU stocks, 200 largest ESTU stocks as benchmark**			
Risk Indices	3.62	79.53	3.57	76.00
Industries	0.90	4.94	0.72	3.06
Covariance * 2	N/A	0.48	N/A	1.01
Asset Selection	1.57	15.05	1.83	19.92
Active	4.06		4.09	
Benchmark	23.12		30.22	
Total	22.30		28.97	

* Note: Hong Kong-listed stocks are not included.

