Journal of International Money and Finance 131 (2023) 102794

Contents lists available at ScienceDirect

Journal of International Money and Finance

journal homepage: www.elsevier.com/locate/jimf

Price comovement and market segmentation of Chinese A- and H-shares: Evidence from a panel latent-factor model

Yingjie Dong[a], Wenxin Huang[b], Yiu-Kuen Tse[c,*]

[a] *China Institute of Finance and Capital Markets, Beijing, China*
[b] *Antai College of Economics & Management, Shanghai Jiao Tong University, Shanghai, China*
[c] *School of Economics, Singapore Management University, Singapore*

ARTICLE INFO

Article history:
Available online 4 January 2023

JEL Codes:
C410
G120

Keywords:
Common latent factors
Cross-listing
Cup-Lasso method
Market segmentation
Panel model

ABSTRACT

This paper examines the price comovement of cross-listed Chinese A- and H-shares using a panel model with latent factors and a heterogeneous long-run structure. Our model is more flexible than the cointegration system and is estimated using the data-driven Cup–Lasso method. The long-run H-share price discounts are heterogeneous across different groups of stocks. We have identified both stationary and nonstationary latent factors in the price differentials, which are driven by different economic variables. By analyzing the factor loadings of the nonstationary latent factor, we identify some trading-friction and information-friction variables that have effects on the price convergence between the A- and H-shares.

1. Introduction

It has been increasingly popular for firms to cross-list their shares on multiple exchanges. If cross-listed shares are driven by the same fundamentals and shareholders of the firms have the same voting rights and claims on future cash flows, prices of the cross-listed shares should be cointegrated in equilibrium. When trading is frictionless, deviations from the equilibrium relationship should be transitory and will give rise to arbitrage opportunities. Arbitrage activities, however, may be impeded by trading frictions. These impediments may cause the markets to be segmented and the cointegration relationship to be violated. For shares cross-listed in segmented multiple markets, the existence of persistent shocks delinking the prices of the cross-listed shares may be a more fundamental explanation for the volatility of the price differentials.

If trading frictions are persistent, the cross-listed shares may exhibit price disparity even in the long run. When frictions are constant, the price disparity may be stable so that the prices of cross-listed shares are cointegrated with a cointegration coefficient different from unity. The prices of shares cross-listed in multiple markets may be influenced by shocks specific to the market in which they are traded, which may be transitory or persistent. In sum, the prices of cross-listed shares may be cointegrated with a coefficient of unity or different from unity, and their price differential may be persistent (nonstationary) or transitory (stationary). Empirical models for the price comovement of cross-listed shares must be flexible enough to entertain these possibilities.

[*] Corresponding author at: School of Economics, Singapore Management University, Singapore 178903, Singapore.
E-mail address: yktse@smu.edu.sg (Y.-K. Tse).

https://doi.org/10.1016/j.jimonfin.2022.102794
0261-5606/© 2022 Elsevier Ltd. All rights reserved.

In this paper we study the shares of the companies cross-listed in the mainland Chinese market (A-shares) and the Hong Kong market (H-shares), collectively called AH shares. Scherrer (2021) finds that cross-listing can enhance the information environment. Shan et al. (2022) find that the Chinese stocks in A-share and H-share markets may provide different diversification opportunities for international investors. As these markets are in the same time zone and are in close geographical proximity to each other, they have attracted much research interest. Existing empirical studies on the AH shares mainly focus on estimating cointegration relationships between their prices (Wang and Jiang (2004),Su et al. (2007), Cai et al. (2011)). While these authors find some support for cointegration, they also report that the relationships change over time. Su et al. (2007) report different cointegration equations over different subsamples. Cai et al. (2011) apply a nonlinear Markov error correction model to the AH share prices and conclude that there is an "upturn in the overall cointegration relation." A recent study by Wang and Chong (2018) finds that the AH share markets "began to cointegrate" after the launches of the Connect Programs. Recently, Carpenter et al. (2021) show there was a rise in stock price informativeness of the AH shares due to reforms in the 2000s. These studies show that the AH share prices may not have a stable long-run cointegration relationship. Empirical analysis assuming a cointegration system may be misspecified, leading to spurious results.

Therefore, it is desirable to adopt a model set-up that encompasses possible long-run structures of the price comovement of the AH shares without enforcing the cointegration assumption. To this effect, we use the panel model with common latent factors and heterogeneous long-run structure recently proposed by Huang et al. (2021). The common latent factors may consist of stationary and/or nonstationary components. The long-run discount or premium may be heterogeneous across different groups of stocks. We follow the methodology developed by Huang et al. (2021), which uses a Lasso-based approach to iteratively estimate the unobserved group-specific long-run coefficients and common latent factors. This framework allows the AH share prices to violate cointegration. Using this comprehensive data-driven model, we first investigate whether common latent long-run (short-run) factors exist that will persistently (transitorily) change the relationship between the AH share prices over time. We then examine the underlying economic variables that determine the persistent and transitory components of the AH share price differentials. Our approach has several advantages over existing studies. First, we use a panel model to take account of the underlying structures across different stocks. Second, our model encompasses cointegration, with flexibility for the price differentials to be nonstationary. Third, we use the robust Cup–Lasso method for estimation, which permits inference using standard tests (Huang et al. (2021)). This is the first paper to adopt a stationary and/or nonstationary framework to study the long-run relationship between the AH share prices. Our methodology is robust and data-driven, which circumvents the arbitrary assumption of a stable linear cointegration system.

The main findings of this paper are as follows. Using a panel data set from 2013 to 2018 with 71 cross-listed AH firms, we identify two endogenous levels of long-run discounts of H-shares versus their corresponding A-shares, with one nonstationary common latent factor representing a permanent common trend and one stationary common latent factor representing transitory common shocks. The existence of the nonstationary factor indicates that the AH share prices are not cointegrated. We use the nonstationary factor to identify periods of structural breaks of the AH share markets, which we find to be more informative and reliable than methods using events based on regulatory reforms.

We investigate economic variables affecting the common latent factors at the macro level. Our analysis shows that the permanent common trend is more reliant on the Hong Kong stock market than the A-share market. On the other hand, the A-share market has a much higher impact on the transitory common shocks than the Hong Kong market. We further analyze the factor loadings of the nonstationary latent factor at the micro level based on firm characteristics. We find that trading-friction variables, such as leverage and illiquidity of the H-shares, reduce H-share price convergence, while volatility of H-shares enhances price convergence. In addition, an information-friction variable, measured by the number of analysts covering the firm, improves price convergence. Unlike previous studies, we document a latent persistent trend component in the AH share price differentials and identify the determinants of this component.

The rest of this paper is organized as follows. We give a brief description of the background of the cross-listed A- and H-shares in Section 2. Section 3 outlines the econometric model we are using: a panel model with cross-sectional dependence and endogenous grouping. In Section 4, we present our empirical results. We conclude in Section 5. In Appendix A, we summarize some recent developments of the Chinese stock market. Additional results are reported in the Online Appendix.

2. Background of A- and H-shares

Stocks listed on the Shanghai Stock Exchange (SSE) and the Shenzhen Stock Exchange (SZSE) are traded in renminbi (RMB). H-shares of firms incorporated in mainland China are traded in Hong Kong dollars (HKD) on the Stock Exchange of Hong Kong (SEHK). In the last decade, several programs (such as the QFII program, QDII program, the RQFII program, the Shanghai-Hong Kong Stock Connect Program and the Shenzhen–Hong Kong Stock Connect Program) were introduced to facilitate cross-boundary investment. Appendix A provides a historical account of these reforms.

Although capital flows between the mainland China and Hong Kong markets have been liberalized over time, the SSE/SZSE and SEHK markets maintain some substantial differences. First, there are language and culture differences between investors in mainland China and Hong Kong. A-share investors are predominantly from mainland China, where Chinese is the native language. Listed firms in the SSE/SZSE are only required to disclose information in Chinese. On the other hand, many of the investors in the Hong Kong market are foreigners. Second, the SSE/SZSE and SEHK markets have different investor profiles. While most trading accounts in China belong to retail investors, the majority of trading in the SEHK comes from institutional investors. Third, the listing requirements in the SEHK are more stringent than those in the SSE/SZSE. The SEHK has additional listing requirements for firms incorporated in mainland China, and the accounting standards for the SSE/SZSE and the SEHK are different. Finally, trading costs, such as transaction fees and taxes, are different in the two markets.

Differences between the SSE/SZSE and SEHK markets raise impediments to arbitrage, causing significant price differences.[1] It has been well recorded that H-shares are traded at large discounts to their corresponding A-shares. In our sample period of 2013 to 2018, the average discount of the H-shares (on a currency-adjusted basis) based on the ratio of the logarithmic prices was 19.5%.

3. Panel model with latent factors

We use a panel model with latent cross-sectional dependence to analyze the comovement of the cross-listed AH share prices. Our model incorporates a nonstationary component to represent possible common latent permanent trends and a stationary component to represent possible common latent transitory shocks. This model provides a more general framework than the traditional cointegration set-up for studying market segmentation. To capture the heterogeneity of the H-share discounts, we adopt a latent group structure where firms of the same group are assumed to have the same discount. These discounts are, however, heterogeneous across different groups. We follow the model set-up proposed by Huang et al. (2021), who describe a robust data-driven procedure for estimation and model identification.

3.1. Model set-up and analysis

For each AH firm $i, i = 1, \cdots, N$, on day t, for $t = 1, \cdots, T$, we denote the logarithmic price of the H-share by y_{it} and the logarithmic price of the corresponding A-share by x_{it}. The basic cointegration system of (x_{it}, y_{it}) traditionally used in the literature is given by

$$y_{it} = \beta_i x_{it} + e_{it}, \qquad i = 1, \cdots, N; t = 1, \cdots T, \tag{1}$$

and

$$x_{it} = x_{i,t-1} + \varepsilon_{it}, \tag{2}$$

where e_{it} and ε_{it} are uncorrelated disturbances with a mean of zero, and they are identically distributed over time. If markets are perfectly integrated, β_i in Eq. (1) is equal to unity for all firms (Hasbrouck (1995),Eun and Sabherwal (2003)). When markets are delinked with constant friction, Eq. (1) may still hold, albeit with values of β_i different from unity. Thus, if e_{it} are stationary over time, (x_{it}, y_{it}) are cointegrated, and they comove with each other in the long run. In the SSE/SZSE and SEHK markets, which are not perfectly integrated as described in Section 2, there is no guarantee that the A- and H-share prices of the same firm are cointegrated. Thus, a more general model loosening the cointegration relationship between the A- and H-share prices is more appropriate for capturing their dynamics.

In this paper, we generalize the postulated cointegration relationship between the AH share prices by assuming that the error term e_{it} in Eq. (1) is cross-sectionally dependent on a multi-factor structure. Following Huang et al. (2021), we assume that there are r_1 common latent nonstationary factors in the AH price differential e_{it}, denoted by f_{ut}, for $t = 1, \cdots, T$, where f_{ut} are $r_1 \times 1$ vectors. Similarly, we assume that there are r_2 common latent stationary factors, denoted by f_{st}, for $t = 1, \cdots, T$, where f_{st} are $r_2 \times 1$ vectors. The long-run price differential e_{it} is then given by[2]

$$e_{it} = \lambda'_{1i} f_{ut} + \lambda'_{2i} f_{st} + u_{it} = \lambda'_i f_t + u_{it}, \tag{3}$$

where λ_{1i} and λ_{2i} are, respectively, $r_1 \times 1$ and $r_2 \times 1$ vectors of factor loadings, $\lambda_i = (\lambda'_{1i}, \lambda'_{2i})'$ and $f_t = (f'_{ut}, f'_{st})'$. We also denote $f_u = (f'_{u1}, \cdots, f'_{uT})'$ and $f_s = (f'_{s1}, \cdots, f'_{sT})'$. The nonstationary common latent factors $f_u$capture the long-run (persistent) common factors affecting the changes to the price differentials of all AH shares, while the stationary common latent factors f_s capture the short-run (transitory) comovement common to the price differentials. The factor loadings λ_i vary with individual firms. In particular, λ_{1i} determines the magnitude of the adjustment of the H-share price of firm i to the long-run level, while λ_{2i} determines the transitory changes. The idiosyncratic components u_{it} are assumed to be cross-sectionally independent.

Huang et al. (2021) propose a latent group structure for the long-run cointegration parameters β_i. Let the sample of firms be partitioned into K groups, where K may be from 1 to N, and its value is to be determined endogenously by the data. We denote the groupings by G_k, for $k = 1, \cdots, K$, so that $G_m \bigcap G_n$ is null for $m \neq n$, and $\bigcup_{k=1}^K G_k = \{1, \cdots, N\}$. The long-run discount/premium parameters are then given by $\beta_i = \alpha_k$ for $i \in G_k$, where α_k are distinct for $k = 1, \cdots, K$. Thus, β_i are heterogeneous across groups but homogeneous within the same group, and each firm's group membership is endogenous. This endogenous grouping structure facilitates a balance between parameter parsimony and possible model misspecification. It allows the

[1] See Jacobs and Weber (2015) and Farago and Hjalmarsson (2019) for further analysis of the structure of pairs-trading strategies.

[2] We will use the term "long-run" price differential to refer to the deviation of y_{it} from its long-run level $\beta_i x_{it}$, that is, e_{it}. This is different from the "actual" price differential $y_{it} - x_{it}$.

long-run discount/premium parameters to be heterogeneous across groups so that we can benefit from the use of the panel data. It also overcomes some problems associated with the estimation and inference of nonstationary time series.[3]

The common latent factors f_u and f_s, as well as the latent grouping structure of the H-share discount/premium β_i, can be estimated jointly using the penalized principal component (PPC) method via the continuous-updated-Lasso (Cup–Lasso) approach (Huang et al. (2021)). Compared to other methods in the existing literature, the Cup–Lasso method has several advantages. First, it renders more reliable estimation and inference. When the dependent variable y_{it} and the unobserved factors f_t are serially correlated, the traditional least squares–based estimates, such as the least squares estimates of Bai (2009) and the more advanced penalized least squares estimates of Huang et al. (2020) suffer from the problem of biased inference. Specifically, the computed common latent nonstationary factors f_u may be inconsistent due to spurious regression. On the other hand, as is shown in Huang et al. (2021), the Cup–Lasso estimate renders consistent and unbiased estimates of the parameters β_i as well as the latent common factors f_u and f_s.

Second, traditional approaches, such as Froot and Dabora (1999) and Wang and Jiang (2004), investigate economic variables related to price comovement by directly regressing the cross-listed price differentials on some selected economic variables. As the price differentials may contain both common latent factors, and idiosyncratic errors, if these economic variables are only related to the common latent factors, the regression results may be spurious due to the idiosyncratic errors (Bai and Ng (2006)). Our approach enables us to model the estimated common latent factors directly (without idiosyncratic errors) to avoid spurious regression.

Third, this is the first paper that imposes a latent group-specific structure on the H-share discount/premium parameters β_i. In the absence of any prior information on β_i, classification based on economic variables or on the assumption of a complete heterogeneous discount may be misleading. In contrast, we identify the latent group-specific structure using the data-driven Cup–Lasso method, which has the desirable property that the unobserved group-specific coefficients can be calculated as if the individuals' membership was known.[4]

We combine Eqs. (1) and (3) to obtain

$$y_{it} = \beta_i x_{it} + \lambda_i' f_t + u_{it} = \beta_i x_{it} + \lambda_{1i}' f_{ut} + \lambda_{2i}' f_{st} + u_{it}, \tag{4}$$

which, together with Eq. (2), form our system for modeling (x_{it}, y_{it}). We shall call this set-up the HJPS model. Huang et al. (2021) propose the penalized principal component (PPC) method to estimate and identify the parameters of the HJPS model. The next subsection provides a summary of the PPC procedure, the details can be found in Huang et al. (2021).

3.2. The PPC estimation method

For each AH firm, our focus is to estimate the group-specific cointegration parameters between the AH share price, β_i, with possible common latent stationary and nonstationary factors f_{st} and f_{ut}. We employ the penalized principal component (PPC) method, proposed by Huang et al. (2021), to simultaneously identify the unobserved group structures in the cointegration relationship and estimate the latent factors. Since the unobserved group structures and latent factors are obtained by minimizing a penalized principal component objective function, our identification and estimation results are purely data-driven, which is more robust to regulatory reforms in the AH markets. We summarize the main procedure of the PPC estimation method below.

1. *Determine the number of stationary and nonstationary factors.* We first note that the latent nonstationary factors are essential to the PPC estimation, but the stationary factors do not affect the consistency of the cointegration parameter estimates despite having a second-order bias (see Huang et al. (2021)). Therefore, we first employ the level data to determine the number of nonstationary factors r_1 without any assumption on the unobserved stationary factors. The resulting residuals are then used to obtain the number of stationary factors r_2. Specifically, r_1 and r_2 are obtained by the information criterion (3.11) and (3.13) in HPJS, respectively.
2. *Determine the number of groups.* The number of groups may be based on economic models or determined by data-driven information criteria. In this paper, we minimize the BIC-type information criterion IC_3 as in (3.14) of HPJS to obtain a good estimate for the number of groups.
3. *Initiate cointegration parameters and nonstationary factors using LS estimates.* Given the number of nonstationary factors, the LS estimator $\left(\tilde{\beta}_i, \widetilde{F}_u\right)$ is the solution to the following set of nonlinear equations:

$$\tilde{\beta}_i = \left(x_i' M_{\widetilde{F}_u} x_i\right)^{-1} x_i' M_{\widetilde{F}_u} y_i, \tag{5}$$

$$\widetilde{F}_u \widetilde{V}_{1,NT} = \left[\frac{1}{NT^2} \sum_{i=1}^{N} \left(y_i - x_i\tilde{\beta}_i\right)\left(y_i - x_i\tilde{\beta}_i\right)\right] \widetilde{F}_u, \tag{6}$$

[3] See Huang et al. (2020) for detailed discussions of these issues.

[4] For additional information on the advantages of nonparametric and endogenous grouping, see Su et al. (2016),Huang et al. (2020), and Huang et al. (2021).

where $M_{\widetilde{F}_u} = I_T - \frac{1}{T^2}\widetilde{F}_u\widetilde{F}'_u, \frac{1}{T^2}\widetilde{F}'_u\widetilde{F}_u = I_{r_1}, \widetilde{F}_u = \left(\tilde{f}_{u1}, \ldots, \tilde{f}_{uT}\right)'$ and $\widetilde{V}_{1,NT}$ is a diagonal matrix consisting of the r_1 largest eigenvalues of the matrix inside the square brackets in (6), arranged in descending order.

4. *Identify unknown group structure and estimate group-specific cointegration parameters.* Using the initial estimates of $\tilde{\beta}_i$ and $\widetilde{F}_u$ as starting values, we minimize the following PPC criterion function to obtain estimates of $(\boldsymbol{\beta}, \boldsymbol{\alpha}, F_u)$:

$$Q_{NT}^{\lambda,K}(\boldsymbol{\beta}, \boldsymbol{\alpha}, F_u) = Q_{NT}(\boldsymbol{\beta}, F_u) + \frac{\lambda}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\|\beta_i - \alpha_k\|, \tag{7}$$

where $Q_{NT}(\boldsymbol{\beta}, F_u) = \frac{1}{NT^2}\sum_{i=1}^{N}(y_i - x_i\beta_i)'M_{F_u}(y_i - x_i\beta_i)$, and $\lambda = \lambda(N, T)$ is a tuning parameter. Minimizing the PPC criterion function in (7) produces the Classifier-Lasso estimators (see Su et al. (2016)) $\left(\hat{\beta}_i, \hat{\alpha}_k, \widehat{F}_u\right)$ of (β_i, α_k, F_u), with $\widehat{F}_u = \left(\hat{f}_{u1}, \ldots, \hat{f}_{uT}\right)'$. Note that

$$\widehat{F}_u V_{1,NT} = \left[\frac{1}{NT^2}\sum_{i=1}^{N}\left(y_i - x_i\hat{\beta}_i\right)\left(y_i - x_i\hat{\beta}_i\right)'\right]\widehat{F}_u, \tag{8}$$

where $\frac{1}{T^2}\widehat{F}'_u\widehat{F}_u = I_{r_1}$ and $V_{1,NT}$ is a diagonal matrix consisting of the r_1 largest eigenvalues of the matrix inside the square brackets in (8), arranged in descending order. The resulting estimated groups are defined as

$$\widehat{G}_k = \left\{i \in \{1, 2, \ldots, N\} : \hat{\beta}_i = \hat{\alpha}_k\right\} \text{for} k = 1, \ldots, K. \tag{9}$$

5. *Obtain the stationary common factors.*Given the estimates $\hat{\beta}_i, \hat{\alpha}_k$, and $\widehat{F}_u$, we compute the cointegration residuals $\hat{r}_{it} = y_{it} - \hat{\beta}'_i x_{it} - \hat{\lambda}'_{1i}\hat{f}_{ut}$.We then employ the standard procedure in stationary panel models with interactive fixed effects (see Bai (2009)). The LS estimator $\hat{F}_s$ is the solution to the following eigenvalue problem:

$$\widehat{F}_s\widetilde{V}_{2,NT} = \left[\frac{1}{NT}\sum_{i=1}^{N}\hat{r}_i\hat{r}_i\right]\widehat{F}_s, \tag{10}$$

where $\frac{1}{T}\widehat{F}'_s\widehat{F}_s = I_{r_2}, \widehat{F}_s = \left(\hat{f}_{s1}, \ldots, \hat{f}_{sT}\right)'$ and $V_{2,NT}$ is the diagonal matrix consisting of the r_2 largest eigenvalues of the matrix inside the square brackets in (10), arranged in descending order.

6. *Compute the Cup–Lasso estimator.* We iterate the above steps to update the estimates of the long-run cointegration parameters $\left(\hat{\beta}_i, \hat{\alpha}_k\right)$, nonstationary common factors $\hat{f}_{ut}$, and stationary common factors $\hat{f}_{st}$ until numerical convergence is achieved. The final estimates are called the Cup–Lasso estimates.

3.3. Model implications

The coefficients β_i (or α_k) capture the discount/premium of the logarithmic H-share prices relative to their corresponding A-share prices. If cross-listed shares are traded in integrated markets without frictions, β_i are expected to be 1 for all firms. When markets are not integrated, β_i may differ from 1 with a latent group-specific pattern. When $\beta_i < 1$ ($\beta_i > 1$), an H-share of firm i is sold at a discount (premium) to the associated A-share. The determination of the number of groups K, as well as the empirical estimates of the coefficients β_i, are both data-driven.

If $f_u = 0$, prices of cross-listed A- and H-shares are cointegrated in the classical sense.[5] When shares are cross-listed in segmented markets, f_u may not be zero and the nonstationary common latent factors f_u will persistently change the long-run relationship between the AH shares. For individual firm i the effect of the shocks on its H-share price is $\lambda_{1i}f_{ut}$, which depends on factor loadings λ_{1i}. Unlike the common latent factors f_u, which depend on macroeconomic shocks, factor loadings λ_{1i}depend on microeconomic factors specific to each firm. The size of $\lambda_{1i}f_{ut}$ has implications for price correction (convergence) to the long-run level $\beta_i x_{it}$, which may be used to calibrate the extent of market segmentation and the success of arbitrage. Finally, the stationary common latent factors f_s capture the short-run transitory movement of the AH long-run price differential. It will be of interest to understand the macroeconomic determinants driving the stationary common latent factors f_s.

4. Empirical results

We apply the HJPS methodology to model the price comovement of the AH shares and examine the determinants of the latent factors at the macro level. The factor loadings of the nonstationary component are then studied at the firm level, which sheds light on factors affecting arbitrage activities.

[5] Note that we will call β_i the "cointegration parameters" even if $f_u \neq 0$.

4.1. Data

As of 2019, there were 112 firms with cross-listed shares in mainland China and Hong Kong. To analyze their share prices, we use a balanced panel data from January 1, 2013, to November 31, 2018. We delete stocks listed after January 2013 and delete stocks with long trading pauses within the sample period (more than 120 trading days). After this filtering, we have 71 AH firms with 1416 common trading days in our sample.[6] Of all firms in the sample, 58 firms are in the SSE and 13 firms are in the SZSE. There are 29 firms in the manufacturing industry, which is the industry with the largest number of cross-listed firms.

We compile trading data, dividends, firm fundamentals, market type, industry classification, and stock splits from the China Stock Market & Accounting Research Database. Trading data of the SSE A-share Index, SZSE A-share Index, Hang Seng (HS) Index and foreign exchange (FX) rate data are compiled from Bloomberg. Adjusted stock prices are computed to remove the influence of stock splits and dividend payments.

4.2. Estimation results

We fit the HJPS model as described in Section 3 using the cross-listed AH shares' logarithmic prices. To remove the influence of the firms' price levels, we first compute the average of the estimated volatilities of the firm's A- and H-share logarithmic prices for each firm over the sample period. We then standardize both the A- and H-share logarithmic prices by this average (see Bai (2009)). The H-share prices are converted to RMB prior to the logarithmic transformation.

Using the information criteria in Huang et al. (2021), we find that there is one nonstationary common latent factor and one stationary common latent factor, with two distinct groups of long-run discount/premium (i.e., $r_1 = r_2 = 1$ and $K = 2$). The presence of the nonstationary component provides direct evidence that the cross-listed A- and H-share prices are not cointegrated.[7] It also shows that assuming a cointegration system in our sample period will result in model misspecification.

In the second stage of the PPC procedure, we compute the model parameters and nonstationary/stationary factors using the Cup–Lasso method. We identify the membership of the two groups of firms with different long-run H-share discounts. There are 44 firms (Group 1) with an estimated αvalue of 0.75, representing an H-share discount of 25% off their corresponding A-shares. We call this group of stocks the low-discount group. Another 27 firms (Group 2) have an estimated αvalue of 0.49, so this group of H-shares is traded at a discount of 51% off their corresponding A-shares. We call this group of stocks the high-discount group. Specifically, 10 out of 13 (77%) firms in the SZSE and 34 out of 58 (59%) firms in the SSE belong to Group 1. In Group 2, six firms are from the mining industry and ten firms are from the manufacturing industry. All 12 firms from the financial industry are in Group 1.[8] These results show the flexibility of the PPC method in identifying the number of groups as well as the determination of the groupings, without any *a priori assumption.*

Table 1 summaries some statistics of the estimated nonstationary and stationary common latent components. These statistics are provided for all stocks as well as separately for low- and high-discount groups. We can see that f_{ut} are positive at all time points, whereas λ_{1i} may be positive or negative for different firms. On the other hand, both f_{st} and λ_{2i} may take positive or negative values. The magnitudes of the nonstationary component $\lambda_{1i}f_{ut}$ are larger than those of the stationary component $\lambda_{2i}f_{st}$. The range of the mean of $\lambda_{1i}f_{ut}$ over all firms in the low-discount group is (0.61, 1.06), while the range for the high-discount group is (1.82, 3.14). Thus, the high-discount group has larger upward price correction than the low-discount group. This adjustment reduces the level of discount for the high-discount group. We shall revisit this point later in the paper.

4.3. Nonstationary latent factor and market segmentation

We now examine whether the estimated nonstationary common latent factor f_u contains useful information related to market integration between the mainland China and Hong Kong stock markets. Several questions are of interest. First, does the nonstationary component evolve smoothly or does it display abrupt changes exemplified by structural breaks? Second, if there are structural breaks, do these breaks coincide with reforms in the SSE/SZSE? Third, do the market reforms provide improved market integration?.

To examine whether the nonstationary common latent factors f_u are smooth or whether there are structural breaks in the series, we use the nonparametric multipower variation method of Barndorff–Nielsen et al. (2006) for jump detection.[9] To check whether the extent of market integration between the SSE/SZSE and SEHK changes after the detected jumps, we use the market-integration measure proposed by Kapadia and Pu (2012). Let R^A and R^H denote the return of A- and H-shares, respectively, over the same period. The mainland China and Hong Kong markets are integrated if the A- and H-share prices move in the same direction (i.e., $R^A \times R^H > 0$). Following Kapadia and Pu (2012), we define $\hat{\kappa}_i$ for firm i as

[6] See the Online Appendix for the details of the firm names, trading codes, and industry codes.

[7] We test for each AH firm whether the A- and H-share prices are cointegrated using Engle–Granger cointegration tests. For 71 firms, 31 firms are cointegrated, while 40 firms are not.

[8] Among the 31 firms that are cointegrated, 22 firms are from Group 1, and nine firms are from Group 2.

[9] See Appendix A1 of Dong and Tse (2017) for detailed descriptions of the jump detection procedures.

Table 1
Summary statistics of the estimated panel latent-factor model.

Parameters/factors	Mean	Median	Min	Max	No of Obs
Panel A: All stocks					
Nonstationary factors f_{ut}	2.49	2.48	1.77	3.06	1416
Nonstationary factor loadings λ_{1i}	0.61	0.66	−0.84	3.86	71
Mean of $\lambda_{1i}f_{ut}$ over all firms	1.51	1.50	1.07	1.85	1416
Stationary factors f_{st}	0.00	−0.09	−0.60	0.53	1416
Stationary factor loadings λ_{2i}	0.36	0.39	−1.62	2.34	71
Mean of $\lambda_{2i}f_{st}$ over all firms	0.00	−0.03	−0.22	0.19	1416
Mean of NT ratio	0.19	0.18	0.15	0.22	1416
Mean of LNA ratio	0.81	0.79	0.76	0.88	1416
Mean of LNH ratio	1.02	1.01	0.91	1.21	1416
Panel B: Low-discount stocks					
Nonstationary factor loadings λ_{1i}	0.35	0.42	−0.84	1.42	44
Mean of $\lambda_{1i}f_{ut}$ over all firms	0.87	0.86	0.61	1.06	1416
Stationary factor loadings λ_{2i}	0.50	0.47	−0.97	2.14	44
Mean of $\lambda_{2i}f_{st}$ over all firms	0.00	−0.04	−0.30	0.27	1416
Mean of NT ratio	0.09	0.09	0.07	0.11	1416
Mean of LNA ratio	0.82	0.82	0.80	0.85	1416
Mean of LNH ratio	1.01	1.00	0.90	1.19	1416
Panel C: High-discount stocks					
Nonstationary factor loadings λ_{1i}	1.03	0.89	−0.18	3.86	27
Mean of $\lambda_{1i}f_{ut}$ over all firms	2.56	2.55	1.82	3.14	1416
Stationary factor loadings λ_{2i}	0.13	0.15	−1.62	2.34	27
Mean of $\lambda_{2i}f_{st}$ over all firms	0.00	−0.01	−0.08	0.07	1416
Mean of NT ratio	0.33	0.32	0.26	0.41	1416
Mean of LNA ratio	0.79	0.76	0.69	0.94	1416
Mean of LNH ratio	1.03	1.01	0.92	1.44	1416

Notes: NT ratio is defined in Eq. (14), LNA ratio is defined in Eq. (15), and LNH ratio is defined in Eq. (16).

$$\hat{\kappa}_i = \sum_{\tau=1}^{M-1}\sum_{t=1}^{M-\tau} I\left(R_{i,t}^{A,\tau} \times R_{i,t}^{H,\tau} > 0\right), \tag{11}$$

where $I(\cdot)$ is the indicator function, M is the number of return observations in the sample period, $R_{i,t}^{A,\tau} = x_{i,t+\tau} - x_{i,t}$, and $R_{i,t}^{H,\tau} = y_{i,t+\tau} - y_{i,t}$. A measure of price concordance of the two markets may be computed using Kendall's tau correlation defined as

$$\kappa_i = \frac{4\hat{\kappa}_i}{M(M-1)} - 1, \tag{12}$$

for firm *i*. The tau statistic takes values within a range of $[-1,1]$; we will use it as a measure for market integration and price concordance, with a value of 1 implying perfect concordance.

As a jump may not be associated with a structural break, we further test whether the extent of market integration before and after a jump is equal. Rejection of equality would support that there is a structural break. We apply the tau statistic to test the differences in market integration over two adjacent periods. The jump is disregarded if the difference is not significant. With this refinement, we finally conclude that there are seven regimes in our sample period.[10] Fig. 1 plots the nonstationary common latent factor f_{ut} over time, with the dates of the regime changes marked and the regimes labelled R1 to R7. For comparison, we also plot the SSE A-share Index, the SZSE A-share Index, and the Hang Seng Index (HSI). We can see that while the SSE and SZSE market indices move very closely together, they are not always in line with the HSI. This is especially true for regimes R3 and R6. It is clear that the mainland China and Hong Kong markets have their own shocks and underlying fundamentals, which may cause the overall markets to move differently.

To examine the extent of market integration and price concordance of SSE/SZSE and SEHK throughout our sample period, we compute the tau statistics in each regime for all AH firms. The results are reported in Table 2, which summarizes the mean of the tau statistics over all firms in each regime, the differences of the means of tau between consecutive regimes, and the p-values of the tests of significant difference in the mean values of tau over consecutive regimes. We can see that the differences are all statistically significant at the 5% level, which supports a changing degree of market integration in the regimes. As these regimes are identified based on the nonstationary common latent factor, this provides evidence that

[10] In the sample period of 2013 to 2018, there are 13 detected jumps in the nonstationary common latent factor f_u, identified using the nonparametric multipower variation method. This result is robust whether we set the significance level of the jump detection statistic to be 1%, 0.5% or 0.1%. If two jumps are very close to each other (less than 20 trading days), the second jump is deleted. We then test whether the market integration between the mainland China and Hong Kong stock markets changes before and after the detected jumps using the nonparametric tau statistics. If the tau statistics over two consecutive regimes do not change significantly, we combine these two regimes into one. With this refinement we finally come up with seven regimes in the sample period.

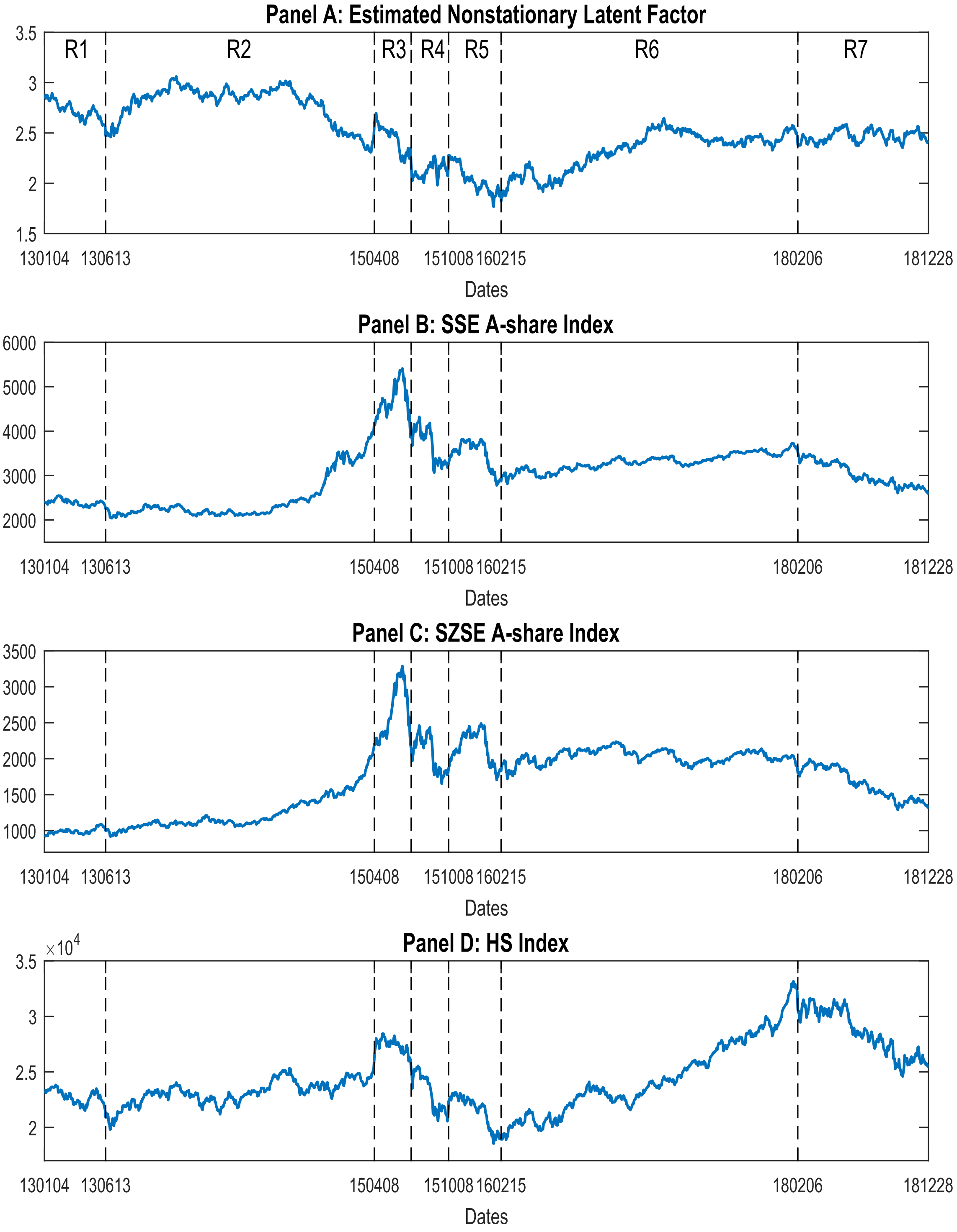

Fig. 1. Nonstationary latent factor and market indices (dates in yymmdd).

Table 2
Market integration and regime changes.

			Kendall's tau		
Regime	Start	End	Mean	Diff	p-value
R1	130104	130612	0.329		
R2	130613	150407	0.277	−0.052	0.000
R3	150408	150705	0.367	0.090	0.000
R4	150706	151007	0.420	0.055	0.011
R5	151008	160214	0.316	−0.108	0.000
R6	160215	180205	0.256	−0.059	0.000
R7	180206	181228	0.382	0.126	0.000

Notes: Regimes R1 to R7 are determined using the nonparametric multipower method on the nonstationary factor f_{ut}. Regime "Start" and "End" are dates in yymmdd format. Kendall's tau statistic is defined in Eqs. (11) and (12). We report the mean of tau (Mean) in each regime, the difference in tau between consecutive regimes (Diff), and the p-value of the test for the difference being zero.

the common latent factor has important information for market integration.[11] These findings also suggest that the AH share market integration varies over time in a nonmonotonic manner.[12]

As there were several reform initiatives in the SSE/SZSE and SEHK in the last decade, it will be interesting to investigate whether these reforms brought about changes in market integration. Wang and Chong (2018) argue that the Shanghai-Hong Kong Stock Connection Program and the Shenzhen–Hong Kong Stock Connection Program are effective in promoting financial integration. Ning et al. (2017) suggest that the foreign exchange reform in August 2015 may have impact on the A- and H-shares. Based on these prior studies, we select five events for further examination as follows: (1) E1, the approval of the Shanghai-Hong Kong Stock Connect Program on April 10, 2014, (2) E2, the launch of the Shanghai-Hong Kong Stock Connect Program on November 17, 2014, (3) E3, the approval of the Shenzhen–Hong Kong Stock Connect Program on August 16, 2016, (4) E4, the launch of the Shenzhen–Hong Kong Stock Connect Program on December 05, 2016, and (5) E5, the foreign exchange reform on August 11, 2015. We test whether the mainland China and Hong Kong stock market integration changes significantly before and after these five events, with 50 days before and after each event date as our sample periods for computing the tau statistics. Table 3 reports the results. We can see that there is only one event that caused significant change in market integration, namely, E4 (the launch of Shenzhen–Hong Kong Stock Connection Program). This finding has two important implications. First, reforms may take time to facilitate changes, and market integration may only be improving slowly. Second, model-based and data-driven approaches may be more efficient in identifying changes in market integration.

We further examine how the extent of market integration evolved over time from 2013 to 2018. To this effect we compute the tau statistics using add-drop samples of 100 daily observations each to compute a time series of tau statistics. In addition to the whole sample of firms, we also separate the data into two groups: the high-discount group with $\hat{\beta} = 0.49$ and the low-discount group with $\hat{\beta} = 0.75$. Fig. 2 plots the results.[13] Panel B presents the time series plots separately for the two groups of stocks. It shows that the relative degree of price concordance of the two groups of stocks varies over time, and that firms with larger long-run price discounts do not necessarily have lower price concordance.

To investigate the differences in the actual discounts of the two groups of shares, we plot the mean of the ratios of the logarithmic H-share price to the associated logarithmic A-share price for all stocks as well as for the two groups of low- and high-discount stocks separately.[14] Fig. 3 shows the results. We can see that in Regime R1 and earlier part of Regime R2, the long-run high-discount group of stocks were *actually* trading at a lower discount than the long-run low-discount group of stocks. It is only from the later part of Regime R2 onward that the high-discount group of stocks started to trade at a higher discount than the low-discount group. Thus, while the estimated parameter $\hat{\beta}$ separates the data into two groups, the group with the higher long-run discount does not necessarily have a higher actual discount. We have to take into account the H-share price adjustment due to the nonstationary and stationary components. Erroneous conclusions may be drawn if these two components are ignored. Indeed, the actual average discount of the long-run high-discount group of H-shares was 22.2%, which is not much higher than that of the long-run low-discount group of 17.9%.

[11] Note that we do not claim that *each* detected jump in f_u indicates a structural change of the integration between the mainland China and Hong Kong stock markets. We emphasize, however, that the estimated nonstationary common latent factor f_u contains important information for market integration, and we can identify structural changes in market integration using them.

[12] We perform robustness checks by sampling transactions every two days and then repeating the calculations following all previous procedures. Results are reported in Table A2 of the Online Appendix. All results are similar to the current findings except that now the market integration does not change significantly between regimes R3 and R4.

[13] Note that the results in Table 2 are computed using the observations in each regime, whereas the results in Fig. 2 are based on moving windows of 100 daily observations.

[14] Note that in the empirical estimation of our model, the prices are standardized by the volatilities. Thus, the estimated long-run discount level $\hat{\beta}$ differs in meaning from the *actual* discount by a factor dependent on the volatility ratios of the A- and H-share prices.

Table 3
Market integration and selected events.

Event	Start	End	Kendall's tau		
			Mean	Diff	*p*-value
E1	140121	140409	0.232		
	140411	140627	0.248	0.017	0.446
E2	140827	141114	0.310		
	141118	150130	0.321	0.011	0.627
E3	160531	160815	0.257		
	160817	161107	0.256	−0.001	0.957
E4	160912	161202	0.290		
	161206	170223	0.218	−0.072	0.002
E5	150528	150810	0.373		
	150812	151102	0.384	0.019	0.503

Notes: This table reports Kendall's tau statistics in periods surrounding selected reform events in the SSE/SZSE and the SEHK. E1 is the approval of the Shanghai-Hong Kong Stock Connect Program, E2 is the launch of the Shanghai-Hong Kong Stock Connect Program, E3 is the approval of the Shenzhen–Hong Kong Stock Connect Program, E4 is the launch of the Shenzhen–Hong Kong Stock Connect Program and E5 is the foreign exchange reform. Event "Start" and "End" are dates in yymmdd format. We report the mean of tau (Mean) in each period, the difference in tau between consecutive periods (Diff), and the *p*-value of the test for the difference being zero.

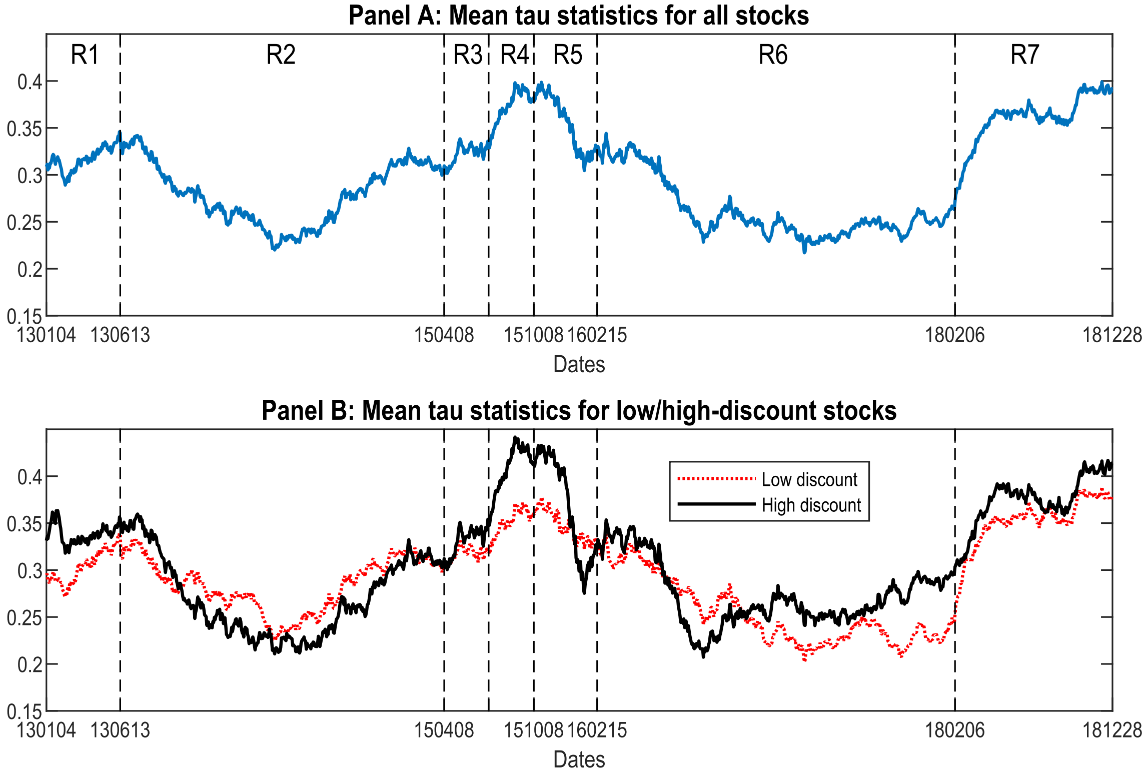

Fig. 2. Mean tau statistics (dates in yymmdd).

4.4. Macro-level analysis of the nonstationary latent factor

We now investigate how economic variables affect the persistent common trend of the AH price differential. Specifically, we regress changes in f_{ut} on the mainland China and Hong Kong stock market index returns as well as the percentage change in the exchange rate between RMB and HKD.[15] The empirical model is as follows

$$\triangle f_{ut} = \theta_0 + \theta_1 Asei_t + \theta_2 Hsi_t + \theta_3 Rmbhk_t + \epsilon_t, \tag{13}$$

[15] Fung et al. (2022) find that the magnitude of the H-share discount is related to the expected RMB appreciation. Thus, we include the exchange rate variable in our regression analysis of the computed nonstationary latent factor (Section 4.4) and transitory common shock (Section 4.7).

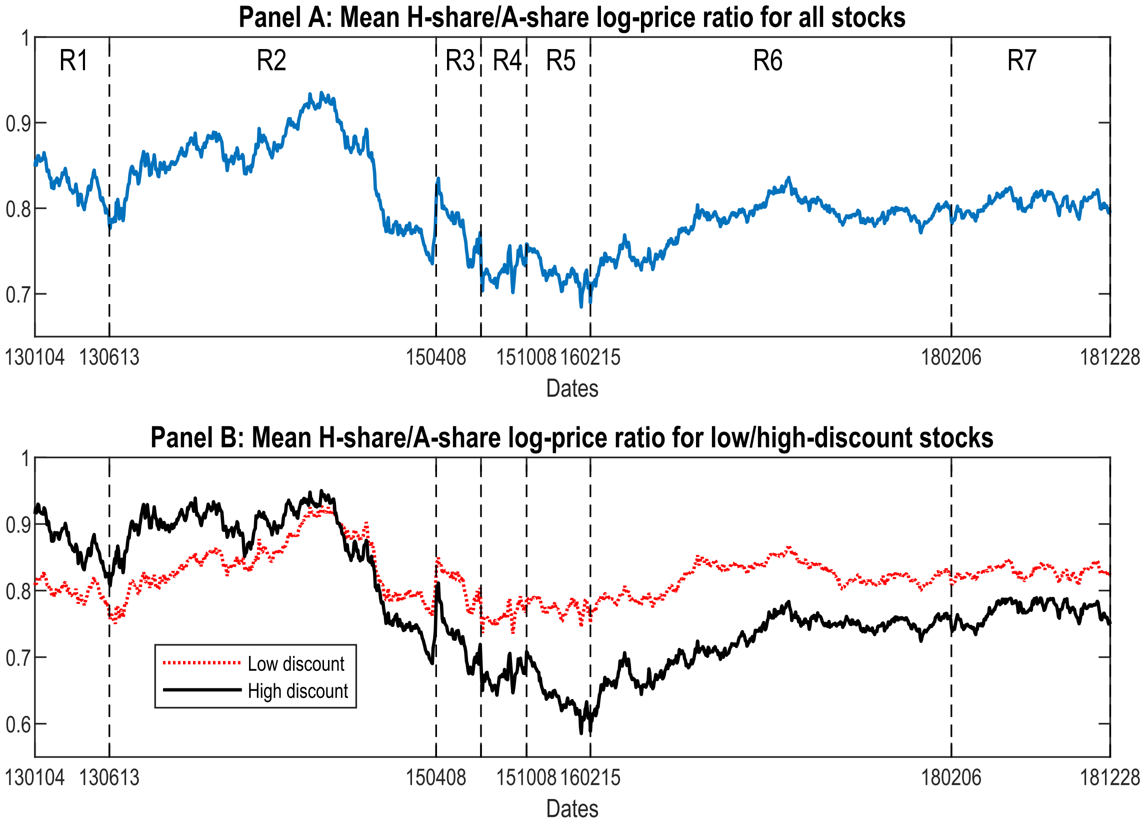

Fig. 3. Mean H-share/A-share log-price ratio (dates in yymmdd).

where $\triangle f_{ut} = f_{ut} - f_{u,t-1}$, $Asei_t$ is the mainland China stock market index return on day t computed as the mean of the t^{th} -day SSE and SZSE index returns, Hsi_t is the HSI return on day t, and $Rmbhk_t$ is the percentage change in the exchange rate between RMB and HKD on day t. To control for the influence of structural changes, we remove the trading days when there are detected jumps for f_{ut}.

We run regressions for both the full model with all explanatory variables and the reduced models with only one explanatory variable, with the results presented in Table 4. We can see that Model U1 has a low adjusted R^2 of 0.0026 and the regressor *Asei* is statistically insignificant. In contrast, in Model U2, *Hsi* is statistically significant and the regression has an adjusted R^2 of 0.4164. When all explanatory variables are in, as in Model U4, the adjusted R^2 increases to 0.5249. In this model *Hsi* and *Asei* are both statistically significant, albeit with different signs. The coefficient of *Hsi* has a larger magnitude and also a larger *t*-value. On the other hand, *Rmbhk* is not statistically significant. It is noted that the coefficients of the variables *Asei* and *Rmbhk* have different signs in Model U4 than in Models U1 and U2. This may be due to missing variables in the latter two models.[16] Overall, there is evidence that the nonstationary common latent factor is affected by both the A- and H-share indices, with the influence of the latter being higher.

4.5. Micro-level analysis of the nonstationary factor loadings

While the nonstationary common latent factors f_u affect all AH shares at the macro level, their effects on individual firms are different as there are different factor loadings λ_{i1} for different firms. In our model, the effect of the nonstationary factor on the H-share price of firm $i$at time t is $\lambda_{i1} f_{ut}$. This effect is higher for higher factor loading λ_{i1}, as f_{ut} are all positive (see Table 1). Given f_{ut}, a larger factor loading λ_{i1} implies higher upward adjustment of the H-share price. Thus, a firm with a larger factor loading will on average have a bigger price discount reduction and better AH share price convergence. We now investigate the economic channels that may affect the factor loadings of the AH firms. We postulate that factors on an individual firm level that encourage arbitrage activities will give rise to higher factor loadings and thus better price convergence. We use

[16] We also include the mainland China and Hong Kong stock market risks and the expected exchange rates in the regression model, and they all have negligible effects on the nonstationary factor. The calculated variance inflation factor (VIF) values of independent variables in Model U4 are all below 1.40, suggesting that multicollinearity may not be a concern for our current model set-up.

Table 4
Regression results for the nonstationary common latent factor.

Variables	Model U1		Model U2		Model U3		Model U4	
	Coefficient	*t*-value	Coefficient	*t*-value	Coefficient	*t*-value	Coefficient	*t*-value
Asei	0.916	1.59					−0.612***	−14.82
Hsi			1.497***	23.09			1.974***	34.91
Rmbhk					−0.752**	−2.02	0.263	0.92
Intercept	−0.000	−0.27	−0.000	−0.71	−0.000	−0.16	−0.000	−0.59
R^2	0.0033		0.4168		0.0035		0.5260	
Adj-R^2	0.0026		0.4164		0.0028		0.5249	
F-value	2.52		533.17		4.07		410.24	

Notes: These are the estimation results for the ordinary least squares regressions of the differenced nonstationary common latent factor f_{ut} on selected economic variables. *Asei* is the mainland China stock market index return, which is the mean of the SSE and SZSE index returns, *Hsi* is the Hang Seng Index return, and *Rmbhk* is the rate of change of the exchange rate between RMB and HKD. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. The *t*-values are calculated based on robust standard errors.

firm-characteristic variables, variables related to trading frictions, and variables related to information frictions as variables that may explain variations in the factor loadings.

For basic firm-characteristic variables, we use the market type of A-shares, market size, and the price-to-book ratios as explanatory variables. The market type variable *Mkttype* equals 1 if the A-share is listed in the SSE and 0 if in the SZSE. The variable *Mktsize* is the logarithmic market value of the firm, and *Pbratio* is the price-to-book ratio of the firm. We expect *Mktsize* to have positive effect on the factor loading, as large firms may attract more active arbitrage. On the other hand, the effects of *Mkttype* and *Pbratio* may be indeterminate.

Comovement of prices of cross-listed shares is expected to improve with a reduction in trading frictions. Following Augustin et al. (2019), we consider the following trading-friction variables: firm leverage, *Levg*, and equity volatility, $Volat.a$ and $Volat.h$ (for A- and H-share markets, respectively), Amihud (2002) illiquidity measures $Illiq.a$ and $Illiq.h$ (for A- and H-share markets, respectively) and trading volume $Trvolume.a$ and $Trvolume.h$ (for A- and H-share markets, respectively), as proxies for trading frictions. *Levg* is firm leverage, which is the ratio of total debt to total assets. $Volat.a$ and $Volat.h$ are the mean daily realized range volatility estimates (standardized annualized volatility estimates in percentage) of A-share and H-share prices, respectively. $Illiq.a$ and $Illiq.h$ are the scaled Amihud illiquidity measures of A- and H-share prices, respectively. Specifically, $Illiq_{i,t} = |R_{i,t}/Trvolume_{it}| \times 10^4$, where $R_{i,t}$ is the daily stock return of firm i on day t, and $Trvolume_{it}$ is the logarithmic monetary value of traded shares of firm i on day t. We use the mean value over time as our illiquidity measure. We expect leverage and illiquidity to have negative effects on factor loadings. On the other hand, volatility may be indeterminate.

Improvement in cross-border information flows is expected to enhance the value of the information of a firm's stock price and cross-listed price comovement and will also increase the attention of cross-border investors (Augustin et al. (2019)). Thus, we use three measures of information frictions: the number of analysts covering the firm (*Analystnum*), the number of reports produced (*Reportnum*), and a variable of audit services (*Big*4). *Big*4 is equal to 1 if the audit service is provided by a big-four accounting firm, and 0 otherwise. We expect high analyst coverage to increase factor loading. Table 5 summarizes some descriptive statistics of the selected explanatory variables.

We run ordinary least squares regressions of the nonstationary common latent factor loadings λ_{1i} on the economic variables. Table 6 reports the results. From Model L1 we see that market size *Mktsize* is significantly positive, suggesting that a large firm attracts more active price correction. The results from Model L2 show that leverage, *Levg*, and illiquidity of the H-shares, $Illiq.h$, are statistically significant. As their coefficients are negative, these variables reduce price convergence. On the other hand, all variables for the A-share market have no effect on price convergence. This finding suggests that price correction is not dependent on trades in the A-share market. Also, trading volumes do not have significant effects on price adjustment. This equation has good explanatory power, with an adjusted R^2 of 0.3401.

The results from Model L3 show that the number of analysts, *Analystnum*, has significant positive effects on the factor loadings, thus enhancing price convergence. This model has good explanatory power with an adjusted R^2 of 0.2225. In Model L4 we include all three sets of economic variables in the regression. While the adjusted R^2 increases to 0.4166, quite a number of variables are statistically insignificant, probably due to multicollinearity. Finally, in Model L5 we summarize the results after deleting some insignificant variables. Leverage, *Levg*, volatility and illiquidity in the H-share market, $Volat.h$ and $Illiq.h$, respectively, as well as the analysts variable, *Analystnum*, are significant. The variable *Mktsize* has dropped out of the regression, as it may be correlated with *Analysnum* and is subsumed by it. On the other hand, $Volat.h$ has a significant positive coefficient, suggesting that arbitrageurs may actually prefer stocks with larger price movement. This model has an adjusted R^2 of 0.4041. Overall, the results confirm that only variables in the H-share market are relevant for determining arbitrage activities, while characteristics for the A-share market do not play a role.

Table 5
Descriptive statistics of explanatory variables for nonstationary factor loading.

Variables	Mean	Std Dev	Min	Max
Panel A: Basic firm-characteristic variables				
Mkttype	0.817	0.390	0.000	1.000
Mktsize	17.605	1.377	14.743	21.022
Pbratio	1.164	0.599	0.373	3.027
Panel B: Trading-friction variables				
Levg	0.510	0.232	0.032	0.936
Volat.a	29.930	9.063	3.629	47.152
volat.h	31.207	5.503	18.824	44.216
Illiq.a	8.470	1.790	4.480	12.600
Illiq.h	9.265	2.188	4.950	14.900
Trvolume.a	19.031	0.925	16.855	21.320
Trvolume.h	17.749	1.639	13.676	21.055
Panel C: Information-friction variables				
Analystnum	15.888	9.234	0.331	36.948
Reportnum	34.710	24.587	0.494	133.250
Big4	0.711	0.450	0.000	1.000

Notes: For the basic firm characteristics, the market type variable, *Mkttpye*, equals 1 if the A-share is listed on the SSE and 0 if it is listed on the SZSE. The market size variable, *Mktsize*, is the logarithmic market value of the firm, and *Pbratio* is the price-to-book ratio. For the trading frictions variables, *Levg* is the firm leverage, which is equal to the ratio of total debt to total assets. *Volat.a* and *Volat.h* are the mean daily realized range volatility estimates (standardized annualized volatility estimates in percent) of A- and H-shares, respectively. *Illiq.a* and *Illiq.h* are the calculated scaled Amihud illiquidity measure of A- and H-shares, respectively. Specifically, $Illiq = |R_{it}/Trvolume_{it}| \times 10^4$, where R_{it} is the daily stock return of firm i on day t, and $Trvolume_{it}$ is the logarithmic monetary value of traded shares of firm i on day t. For the information frictions variables, *Analystnum* is the number of analysts covering the firm, *Reportnum* is the number of reports covered, and *Big4* is a dummy variable of auditing services, which equals 1 if the audit service is provided by a big-four accounting firm, and 0 otherwise.

Table 6
Regression results for the nonstationary factor loadings.

Variables	Model L1		Model L2		Model L3		Model L4		Model L5	
	Coeff.	*t*-value	Coeff.	*t*-value	Coeff.	*t*-value	Coeff.	*t*-value	Coeff.	*t*-value
Panel A: Basic firm-characteristic variables										
Mkttype	−0.054	−0.20					−0.624**	−2.34		
Mktsize	0.260***	4.30					−0.112	−0.59		
Pbratio	0.374	1.55					0.128	0.51		
Panel B: Trading-friction variables										
Levg			−1.064***	−3.17			−1.229***	−2.73	−1.204***	−2.99
Volat.a			0.006	0.70			0.005	0.80		
Volat.h			0.080	1.66			0.071*	1.80	0.062**	2.54
Illiq.a			−0.033	−0.40			0.010	0.12		
Illiq.h			−0.304**	−2.03			−0.391***	−2.67	−0.297***	−4.28
Trvolume.a			−0.075	−0.72			−0.125	−0.85		
Trvolume.h			0.160	1.56			0.183	1.45		
Panel C: Information-friction variables										
Analystnum					0.059**	2.31	0.026	0.83	0.029***	3.07
Reportnum					−0.009	−0.97	−0.005	−0.48		
Big4					0.233	1.55	0.102	0.59		
Intercept	−4.366***	−4.70	0.180	0.07	−0.198	−1.34	3.518	1.11	1.591***	3.61
R^2	0.2660		0.4061		0.2558		0.5249		0.4382	
Adj-R^2	0.2331		0.3401		0.2225		0.4166		0.4041	
F-value	10.97		10.34		10.36		6.10		14.46	

Notes: These are the results of the ordinary least squares estimation for the regressions of the nonstationary common factor loadings λ_{1i} on selected economic variables. Definitions of the explanatory variables can be found in Table 5. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. The *t*-values are calculated based on robust standard errors.

4.6. Components of H-share price discount

The price of an H-share depends on the long-run discounted A-share price $\beta_i x_{it}$, its response to the nonstationary component $\lambda_{1i} f_{ut}$, and its response to the stationary component $\lambda_{2i} f_{st}$. It will be interesting to see the relative contribution of each component to the H-share price. To this effect, we compute several ratios to help assess the relative impact of each component.

We first define the measure NT ratio (nonstationary component to total response) for firm i at time t as the response of the nonstationary component to the sum of the long-run and nonstationary component as follows

$$\mathrm{NT}_{it} = \frac{\lambda_{1i} f_{ut}}{\beta_i x_{it} + \lambda_{1i} f_{ut}}. \tag{14}$$

We compute a time series of the average NT ratio for all stocks, as well as for each group of high- and low-discount stocks. Fig. 4 presents the results. From Table 1 we can see that the overall NT ratio varies between 0.15 and 0.22. On the other hand, the NT ratio of the high-discount group, with a range of (0.26, 0.41), is higher than that of the low-discount group, with a range of (0.07, 0.11). Thus, the long-run high-discount group has proportionally larger adjustment (H-share price increase) due to the nonstationary component. Indeed, the response due to the nonstationary component increases the price of the H-share in the high-discount group much higher than the long-run level, resulting in a reduction in the size of the discount. From Panel C of Fig. 4 we can see that the NT ratio for the high-discount group in Regime R1 and the early part of Regime R2 is quite high. This large positive correction results in the high-discount group having a lower *actual* discount than the low-discount group in this period, as shown in Panel B of Fig. 3.

We further examine the aggregate of the long-run price level and the nonstationary response in relation to the A-share price. We call this ratio the LNA (long-run and nonstationary component to A-share price) ratio, which is defined as

$$\mathrm{LNA}_{it} = \frac{\beta_i x_{it} + \lambda_{1i} f_{ut}}{x_{it}}. \tag{15}$$

This ratio may be taken as a measure of the systematic part of the H-share price as a fraction of its associated A-share price. Fig. 5 presents the time series plot of the average LNA ratio for all stocks, as well as for each group of high- and low-discount stocks. Panel B shows that the LNA ratio is quite stable for the low-discount group. Table 1 shows that the range of the LNA ratio for the low-discount group is (0.80, 0.85). In contrast, there is more variation in the LNA ratio for the high-discount group, with a range of (0.69, 0.94). These results are in line with the pattern in Fig. 3. In particular, the high LNA ratio in Regime R1 and the earlier part of R2 for the high-discount stocks brings the actual discount of this group of stocks lower in the earlier period of the sample.

We further compute the ratio of the sum of the long-run price level and the nonstationary response in relation to the *actual* H-share price. We call this ratio the LNH (long-run and nonstationary component to H-share price) ratio, which is defined as

$$\mathrm{LNH}_{it} = \frac{\beta_i x_{it} + \lambda_{1i} f_{ut}}{y_{it}}. \tag{16}$$

Fig. 6 presents the time series plot of the average LNH ratio for all stocks as well as for each group of high- and low-discount stocks. Note that if there is no stationary component ($f_s = 0$), the LNH ratio should be close to unity. From Table 1 we can see that the means of the LNH ratios are very close to 1, for all stocks as well as for each of the two discount groups. Furthermore, it can be seen that the graph of the LNH ratio for the low-discount stocks is quite stable and is close to unity, with values within the range of (0.90, 1.19). In contrast, the graph of the high-discount shares is quite volatile, with the ratio in the range of (0.92, 1.44).

4.7. Analysis of the transitory common shock

We now investigate the economic variables that are related to the (transitory) stationary common latent factor f_s underlying the comovement of the AH price differential. First, we choose mainland China and Hong Kong stock market risks as our explanatory variables, which are denoted by $Risk_A$ and $Risk_H$, respectively. We use the realized range method of Alizadeh et al. (2002) to compute the daily realized variance and treat the calculated values as proxies for the market risks. To compute the mainland China stock market risk $Risk_A$, we calculate the realized range estimate of both the SSE A-share Index and the SZSE A-share Index, and then compute $Risk_A$ as the cross-sectional mean value. Similarly, $Risk_H$ is computed as the realized range volatility estimate of the HSI. Following Wang and Jiang (2004), we also include the relative level of risk aversion $ReRisk_t$ at time t as an explanatory variable, which is computed as

$$ReRisk_t = \frac{|Risk_{At} - Risk_{Ht}|}{Risk_{At} + Risk_{Ht}}, \tag{17}$$

where $Risk_{At}$ and $Risk_{Ht}$ are the realized range volatility estimates of the mainland China and Hong Kong stock market indices on day t, respectively.

Second, we use mainland China and Hong Kong stock market sentiments as explanatory variables. We compute the logarithmic trading volume of SEHK as a proxy for market sentiments in Hong Kong and denote this variable by $Sentiment_H$. Similarly, $Sentiment_A$ is computed as the mean of the trading volumes of SSE and SZSE A-shares. As the volume variables are trended, we take the first difference of the sentiment variables as regressors and denote them by $\Delta Sentiment_H$ and $\Delta Sentiment_A$. Cai et al. (2011) suggest a volume-based market-sentiment measure using the relative trading volume between markets. We follow Cai et al. (2011) and include the relative AH trading volume $ReSentiment_t$ at time t as an explanatory variable, which is defined as

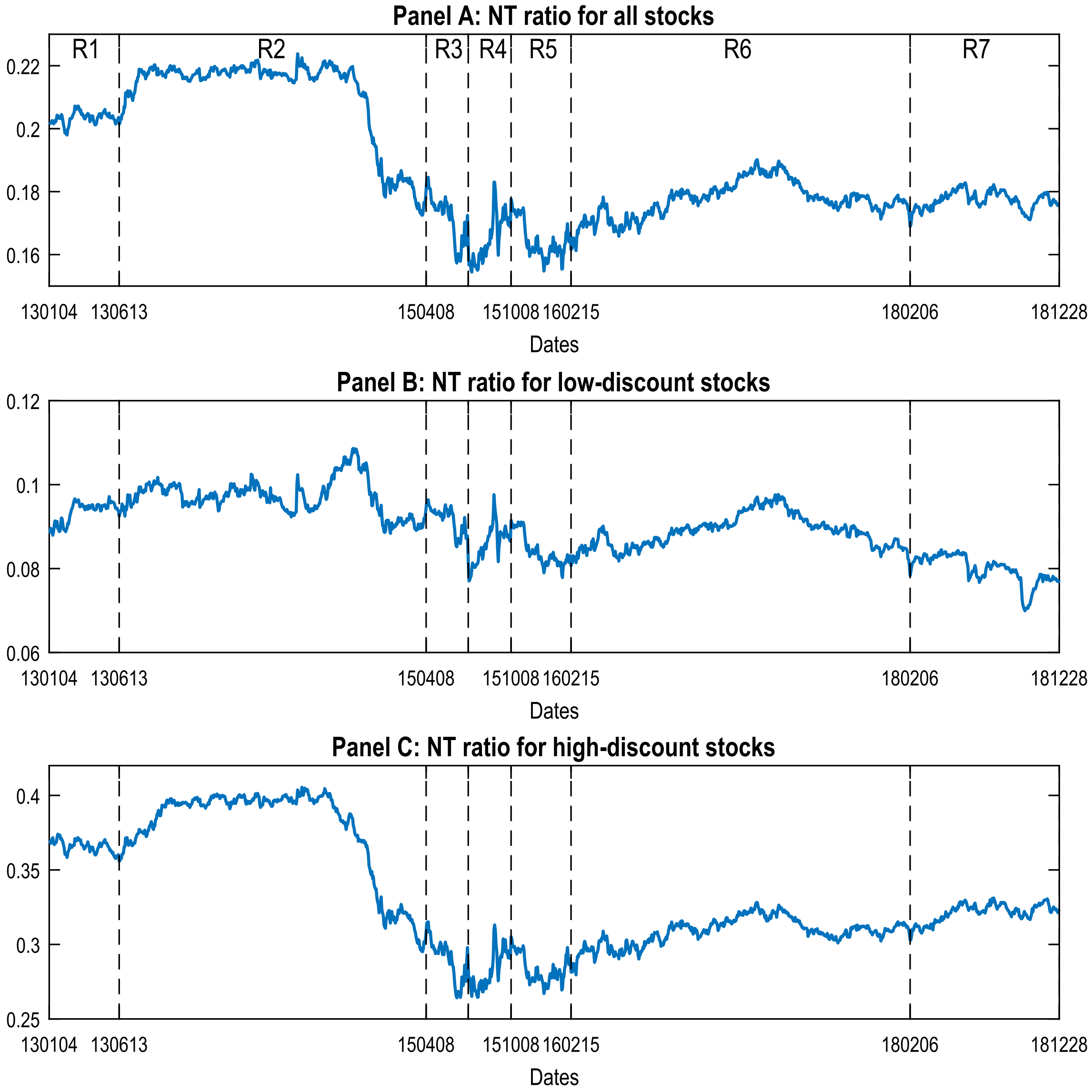

Fig. 4. NT ratio (dates in yymmdd).

$$ReSentiment_t = \frac{|Sentiment_{At} - Sentiment_{Ht}|}{Sentiment_{At} + Sentiment_{Ht}}, \qquad (18)$$

where $Sentiment_{At}$ and $Sentiment_{Ht}$ are A- and H-share trading volumes for firm i on day t, respectively.

Finally, we also include the expected devaluation in the RMB, denoted as $\Delta Z_t, t = 1, \cdots, T$, in our evaluation. ΔZ_t is calculated as a simple average of the daily exchange rate fluctuations between six neighboring currencies and the US dollar. We follow Wang and Jiang (2004) and select the six neighboring currencies as the Indonesian rupiah, Malaysian ringgit, Singaporean dollar, South Korean won, Taiwanese dollar, and Thai baht. [17]

In summary, our model can be written as

[17] Although the exchange rate of the RMB against the HKD (*Rmbhk*) can be used to reflect the realized RMB devaluation, the expected devaluation in the RMB may not show up directly. If a devaluation of the RMB is expected, we would rationally expect H-share prices to drop. Thus, we include the expected exchange rate data in our regression analysis. Since the expected exchange rate data is unobservable, we follow Wang and Jiang (2004) and use the mean daily exchange rate fluctuations (variance) between six neighboring countries' currencies against the US dollar as a proxy measure. See Wang and Jiang (2004) for more discussions. We also regress the common latent stationary factor f_{st} on *Rmbhk*. However, the computed R^2 is 0.0007, and the coefficient of *Rmbhk* is statistically insignificant based on robust standard errors.

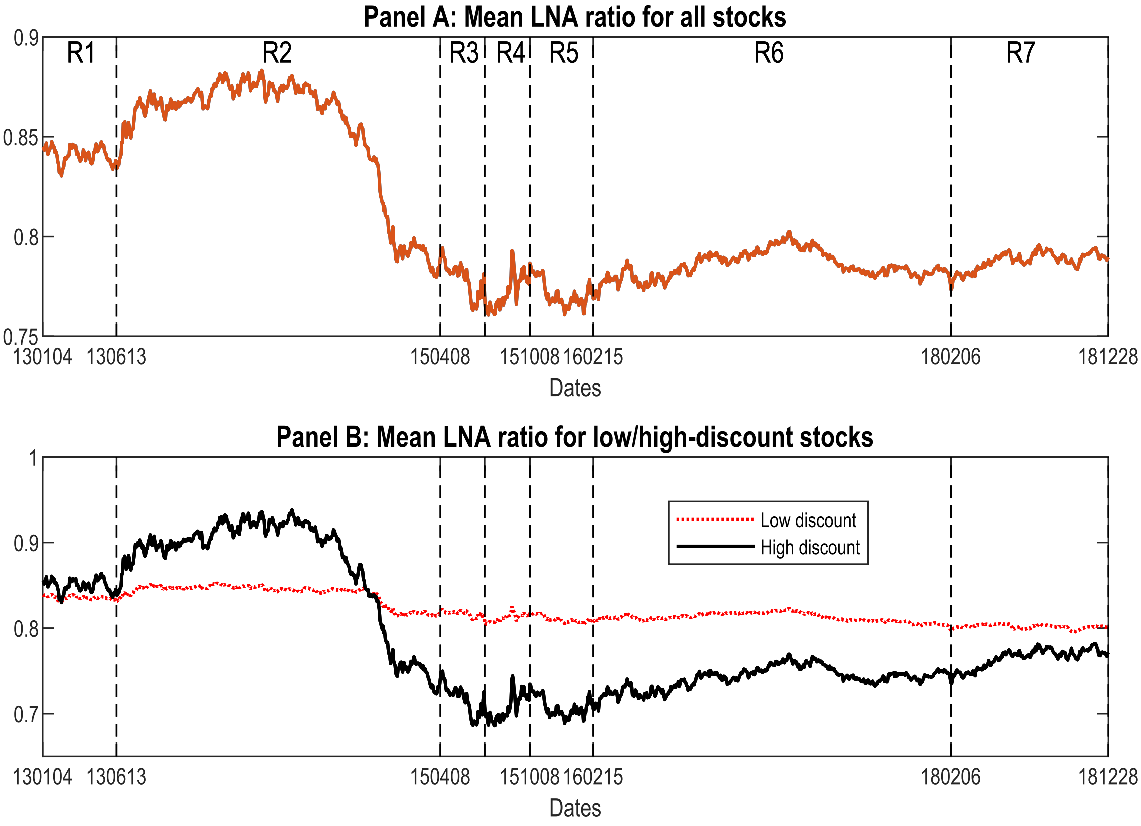

Fig. 5. Mean LNA ratio (dates in yymmdd).

$$f_{st} = \theta_0 + \theta_1 Risk_{Ht} + \theta_2 \Delta Sentiment_{Ht} + \theta_3 Risk_{At} + \theta_4 \Delta Sentiment_{At} + \theta_5 ReRisk_t + \theta_6 ReSentiment_t + \theta_7 \Delta Z_t + \varepsilon_t. \tag{19}$$

We estimate the regressions for the full model with all explanatory variables and the reduced models with subsets of explanatory variables. Table 7 summarizes the results. It can be seen that Model S2, which involves variables for the A-share market, has a higher adjusted R^2 (0.1884) than Model S1 (0.0144), which involves variables from the H-share market. Model S3, which uses relative measures of risk and sentiments, has a lower adjusted R^2 (0.1405) than Model S2. Model S5, which includes all variables selected, has the highest adjusted R^2 of 0.2329. All variables, except for $\Delta Sentiment_H$, are statistically significant. Model S4 shows that ΔZ provides less explanatory power than the H-share market variables. Overall, it can be seen that variations in the common latent transitory component depend predominantly on the A-share market.

5. Conclusion

We have investigated the price comovement of cross-listed AH shares in the mainland China and Hong Kong stock markets. We apply the panel model with latent cross-sectional dependence proposed by Huang et al. (2021) (the HJPS model), which incorporates possible common latent stationary and/or nonstationary components, as well as an endogenous heterogeneous group-specific pattern of the long-run H-share discount/premium. Utilizing this flexible set-up, we are able to analyze the complex comovement of the cross-listed share prices without presuming an equilibrium relationship under the classical cointegration framework.

We use the nonstationary factor to identify periods of structural breaks in the AH share markets. These breaks do not coincide with the reforms of the AH-share markets and show that the nonstationary common latent factor provides better results for the identification of structural breaks in the joint market. We find that market integration is not monotonic chronologically, and that the AH-share markets have higher price concordance during periods of market turmoil.

We perform macro-level analysis of the nonstationary trend and stationary shocks. The nonstationary component is more dependent on the Hong Kong market than the mainland China market. On the other hand, the stationary component is more dependent on the mainland China market variables, including risk variables and sentiment variables. We find that firm lever-

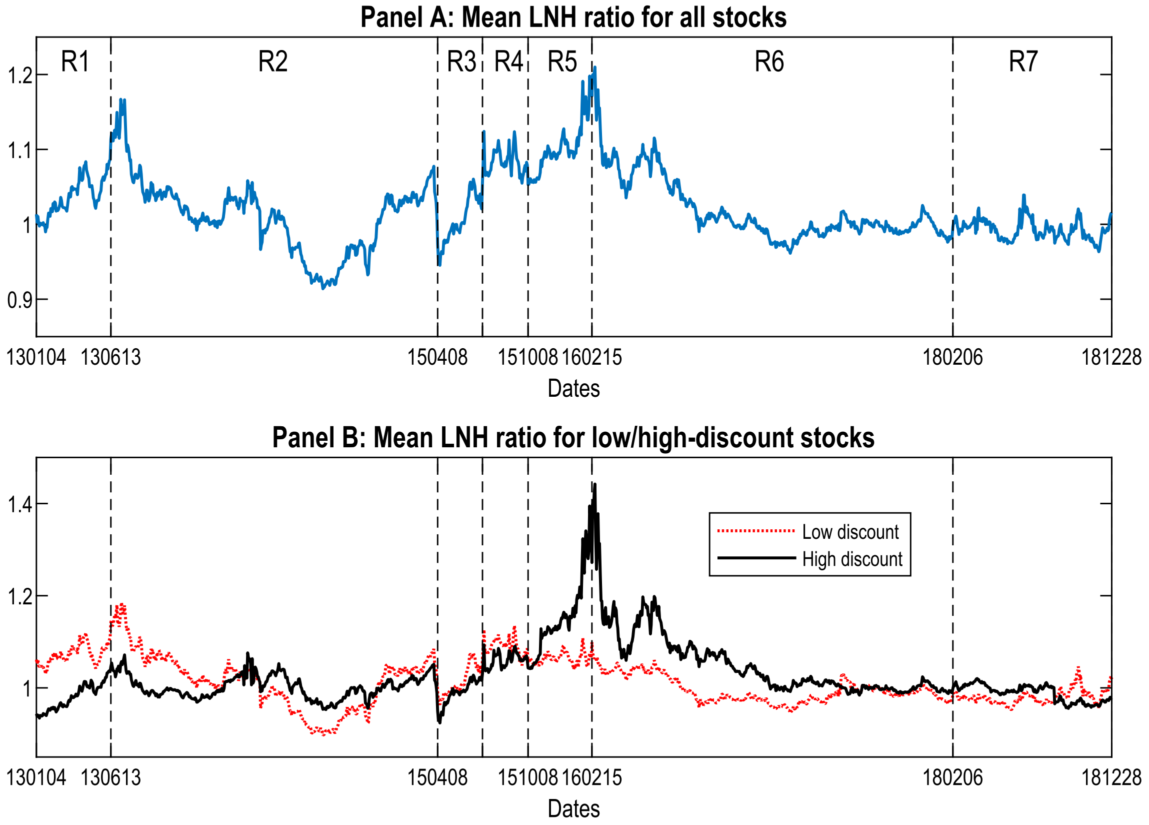

Fig. 6. Mean LNH ratio (dates in yymmdd).

Table 7
Regression results for the stationary common latent factor.

Variables	Model S1		Model S2		Model S3		Model S4		Model S5	
	Coefficient	*t*-value	Coefficient	*t*-value	Coefficient	*t*-value	Coefficient	*t*-value	Coefficient	*t*-value
$Risk_H$	-0.653^{***}	-4.66							0.399^{**}	2.14
$\Delta Sentiment_H$	0.036	1.26							-0.042	-1.60
$Risk_A$			-1.091^{***}	-17.80					-0.956^{***}	-10.05
$\Delta Sentiment_A$			0.132^{***}	3.27					0.163^{***}	3.97
ReRisk					-0.586^{***}	-13.22			-0.219^{***}	-3.93
ReSentiment					-4.261^{***}	-5.75			-3.451^{***}	-4.54
ΔZ							-862.415^{**}	-2.49	-692.103^{***}	-3.06
Intercept	0.067^{***}	4.37	0.168^{***}	15.11	0.346^{***}	9.39	0.014	1.42	0.342^{***}	8.13
R^2	0.0158		0.1895		0.1418		0.0147		0.2367	
Adj-R^2	0.0144		0.1884		0.1405		0.0140		0.2329	
F-value	10.87		159.41		141.58		6.20		70.35	

Notes: These are the estimation results for the ordinary least squares regressions of the stationary common latent factor f_{st} on selected variables. $Risk_A$ and $Risk_H$ are trading volatilities, $\Delta Sentiment_A$ and $\Delta Sentiment_H$ are sentiment variables measured as the differenced logarithmic trading volumes, for the mainland China and Hong Kong stock markets, respectively. Relative trading volatility and volume of A- and H-shares are denoted by *ReRisk* and *ReSentiment*, respectively. Expected devaluation in the RMB is ΔZ. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. The *t*-values are calculated based on robust standard errors.

age and illiquidity of the H-share reduce the level of H-share price correction, whereas volatility in the H-share price and the number of analysts for the firm promote price convergence.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Yingjie Dong acknowledges partial support provided by the projects 71803021 and 71790604 of the National Natural Science Foundation of China, and the project 19ZDA101 of the National Social Science Fund of China. Wenxin Huang acknowledges fundings from the National Natural Science Foundation of China (No. 72003127 and No. 72271159) and the Shanghai Sailing Program (No. 20YF1420100). The content is solely the responsibility of the authors and does not necessarily represent the official views of the China Institute of Finance and Capital Markets.

Appendix A. Reforms in the Chinese stock markets

The China Securities Regulatory Commission (CSRC) implemented several reform programs in the last two decades. In November 2002, the CSRC started to allow licensed foreign institutional investors to invest in RMB-denominated A-shares under the Qualified Foreign Institutional Investor (QFII) program, with each QFII institutional investor given a quota. In 2005, the CSRC initiated a reform of nontradable shares. Before the reform, A-shares were divided into nontradable (the majority) and tradable shares. As a consequence, firms suffered from a lack of liquidity. In the 2005 split-share reform, nontradable shareholders paid tradable shareholders (via giving shares tradable shareholders as well as repurchasing shares) to gain liquidity. In 2006, China launched the Qualified Domestic Institutional Investor (QDII) scheme to allow mainland institutional investors to invest in offshore stocks and bonds. This arrangement provided mainland investors opportunities to access foreign markets. In 2011, China continued to loosen her capital controls by establishing the RMB Qualified Foreign Institutional Investor (RQFII) program, under which foreign institutional investors are allowed to invest in Chinese RMB-denominated bonds. The QFII and RQFII quotas have expanded steadily since their inception.

In November 2014, the Hong Kong Securities and Futures Commission (SFC) and the China Securities Regulatory Commission (CSRC) successfully launched the Shanghai-Hong Kong Stock Connect Program. Eligible mainland Chinese investors in the SSE as well as Hong Kong and international investors in the SEHK are able to trade eligible shares in each other's markets through the trading and clearing facilities of their home exchange. In December 2016, another cross-boundary investment channel, the Shenzhen–Hong Kong Stock Connect program, was launched to connect the SZSE and the SEHK. The Connect Programs operated steadily and smoothly since their launches. As of January 2021, the China-bound daily quota is RMB 52 billion, and the HK-bound daily quota is RMB 42 billion.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jimonfin.2022.102794.

References

Alizadeh, S., Brandt, M.W., Diebold, F.X., 2002. Range-based estimation of stochastic volatility models. J. Finance 57 (3), 1047–1091.
Augustin, P., Jiao, F., Sarkissian, S., Schill, M.J., 2019. Cross-listings and the dynamics between credit and equity returns. Rev. Finan. Stud. 33, 112–154.
Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77 (4), 1229–1279.
Bai, J., Ng, S., 2006. Evaluating latent and observed factors in macroeconomics and finance. J. Econ. 131 (1–2), 507–537.
Barndorff-Nielsen, O.E., Shephard, N., Winkel, M., 2006. Limit theorems for multipower variation in the presence of jumps. Stoch. Process. Their Appl. 116 (5), 796–806.
Cai, C.X., McGuinness, P.B., Zhang, Q., 2011. The pricing dynamics of cross-listed securities: The case of Chinese A- and H-shares. J. Bank. Finance 35 (8), 2123–2136.
Carpenter, J.N., Lu, F., Whitelaw, R.F., 2021. The real value of China's stock market. J. Financ. Econ. 139, 679–696.
Dong, Y., Tse, Y.-K., 2017. Business time sampling scheme with applications to testing semi-martingale hypothesis and estimating integrated volatility. Econometrics 5 (4), 51.
Eun, C.S., Sabherwal, S., 2003. Cross-border listings and price discovery: Evidence from US-listed Canadian stocks. J. Finance 58 (2), 549–575.
Farago, A., Hjalmarsson, E., 2019. Stock price co-movement and the foundations of pairs trading. J. Finan. Quant. Anal. 54 (2), 629–665.
Froot, K.A., Dabora, E.M., 1999. How are stock prices affected by the location of trade? J. Financ. Econ. 53 (2), 189–216.
Fung, J.K., Girardin, E., Hua, J., 2022. How does the exchange-rate regime affect dual-listed share price parity? Evidence from China's A- and H-share markets. J. Int. Money Finance 129, 102738.
Hasbrouck, J., 1995. One security, many markets: Determining the contributions to price discovery. J. Finance 50 (4), 1175–1199.
Huang, W., Jin, S., Phillips, P.C.B., Su, L., 2021. Nonstationary panel models with latent group structures and cross-section dependence. J. Econ. 221 (1), 198–222.
Huang, W., Jin, S., Su, L., 2020. Identifying latent grouped patterns in cointegrated panels. Econ. Theory 36 (3), 410–456.
Jacobs, H., Weber, M., 2015. On the determinants of pairs trading profitability. J. Finan. Mark. 23, 75–97.
Kapadia, N., Pu, X., 2012. Limited arbitrage between equity and credit markets. J. Financ. Econ. 105 (3), 542–564.
Ning, Y., Wang, Y., Su, C.-W., 2017. How did China's foreign exchange reform affect the efficiency of foreign exchange market? Physica A: Stat. Mech. Its Appl. 483, 219–226.
Scherrer, C.M., 2021. Information processing on equity prices and exchange rate for cross-listed stocks. J. Finan. Mark. 54, 100634.
Shan, C., Tang, D.Y., Wang, S.Q., Zhang, C., 2022. The diversification benefits and policy risks of accessing China's stock market. J. Empir. Finance 66, 155–175.
Su, L., Shi, Z., Phillips, P.C.B., 2016. Identifying latent structures in panel data. Econometrica 84 (6), 2215–2264.
Su, Q., Chong, T.T.-L., Yan, I.K.-M., 2007. On the convergence of the Chinese and Hong Kong stock markets: A cointegration analysis of the A and H shares. Appl. Finan. Econ. 17, 1349–1357.

Wang, Q., Chong, T.T.-L., 2018. Co-integrated or not? After the Shanghai-Hong Kong and Shenzhen-Hong Kong stock connection schemes. Econ. Lett. 163, 167–171.

Wang, S.S., Jiang, L., 2004. Location of trade, ownership restrictions, and market illiquidity: Examining Chinese A- and H-shares. J. Bank. Finance 28 (6), 1273–1297.