Transferring Return Predictability to the Beijing Stock Exchange

Yuxuan Zhang, Yun Shi*

School of Statistics, East China Normal University, Shanghai 200062, China

Abstract

We investigate whether return predictability in the Beijing Stock Exchange (BSE) can be enhanced by transferring information from the Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE). Models that incorporate external market information achieve substantially higher out-of-sample performance than those based exclusively on BSE data, increasing the monthly out-of-sample R^2 from 0.36% to as high as 1.53%. These improvements in predictive accuracy are accompanied by stronger portfolio performance over the available out-of-sample window, with the annualized Sharpe ratio of long-short portfolios rising from 1.31 to 3.80. Our findings provide early and suggestive evidence that cross-market transfer learning can help address data scarcity in young and emerging equity markets.

Keywords: Return predictability, Transfer learning, Machine learning, Beijing Stock Exchange, Chinese stock market

JEL Classification: C53, C58, G1, G17

*Corresponding author. Email: yshi@fem.ecnu.edu.cn.

1 Introduction

Can return predictability be transferred from mature stock markets to a young market with limited historical data? This question is particularly relevant in China, where the Beijing Stock Exchange (BSE), launched in 2021, operates alongside the long-standing Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE). The BSE was established to support innovative small and medium-sized enterprises (SMEs) and has become an important component of China's capital market. Existing studies on the BSE have mainly focused on its institutional framework and policy implications (Lu and Ye, 2022; Ye, 2022). While traditional factor models have provided early insights into the BSE (Zhang, 2024), they represent only the first steps in charting a predictive landscape that remains largely unexplored in this emerging market.

A central challenge in studying return predictability on the BSE is the scarcity of historical observations caused by its short trading history. Although machine learning (ML) models have demonstrated strong predictive power in global equity markets (Gu et al., 2020; Chen et al., 2024; Choi et al., 2025) and Chinese A-share markets (Leippold et al., 2022; Liu et al., 2022; Zhu, 2025), these data-demanding models may perform poorly when applied to the BSE. Estimating high-dimensional predictive models using limited local data can lead to overfitting and unstable out-of-sample forecasts. Consequently, improving return predictability in the BSE requires a paradigm shift that moves away from isolated, single-market modeling toward frameworks capable of leveraging information from more mature and related markets.

One natural approach to achieving this objective is transfer learning (TL), a framework designed to enhance model performance in data-sparse target domains by leveraging knowledge from related, data-rich source domains (Pan and Yang, 2009; Weiss et al., 2016; Zhuang et al., 2020). In the financial domain, transfer learning has been increasingly applied to financial forecasting (Ye and Dai, 2022; Zhang et al., 2023), option pricing (Chen et al., 2026), financial text analysis (Kraus and Feuerriegel, 2017; Ergun and Sefer, 2025), and trading decision-making (Jeong and Kim, 2019). In empirical asset pricing, Hellum et al. (2025) show that return predictability contains a strong transferable component across international equity markets and propose the generalized elastic net (GENet) to combine global and local predictive information. Their framework suggests that when a target market has limited data, a predictive model learned from related markets may

provide a useful reference for estimating local expected returns.

This paper applies this idea to China's multi-tier stock market. The SSE and SZSE provide natural source markets for the BSE because they share the same national institutional environment, but have much longer trading histories and richer cross-sectional information. If return-predictive relationships are partly common across Chinese stock markets, information learned from the SSE and SZSE should help predict BSE stock returns. At the same time, the BSE differs from the two mature exchanges in listing standards, firm characteristics, liquidity conditions, and investor composition. Therefore, the key empirical question is not only whether predictive information transfers across markets, but also how external information should be combined with limited local BSE information.

To answer this question, we examine monthly stock return prediction on the BSE using firm-level characteristics. We compare a local model estimated solely with BSE data with transfer learning models that borrow predictive information from the SSE, the SZSE, or their pooled sample. The transfer specifications include hard transfer, which directly applies source market predictive coefficients to the BSE, as well as soft transfer and two-stage estimation based on the GENet framework of Hellum et al. (2025). The results show that cross-market transfer substantially improves out-of-sample predictive accuracy relative to the local benchmark. These improvements also translate into superior portfolio performance, and the transfer learning advantage is especially pronounced when the available BSE training history is limited.

The contributions of this study are threefold. First, it provides early evidence on return predictability in the BSE, complementing studies that mainly examine its institutional and asset-pricing features. Second, it compares the relative benefits of different source markets by evaluating transfer models based on the SSE, the SZSE, and their pooled sample. Third, it sheds light on informational linkages across China's multi-tier stock markets, showing that predictive signals learned from mature exchanges are associated with improved return prediction in a young and data-scarce market over the available out-of-sample period.

The remainder of this paper is organized as follows: Section 2 introduces the data and cross-market transfer design. Section 3 presents the empirical results. The conclusion and further research directions are stated in Section 4.

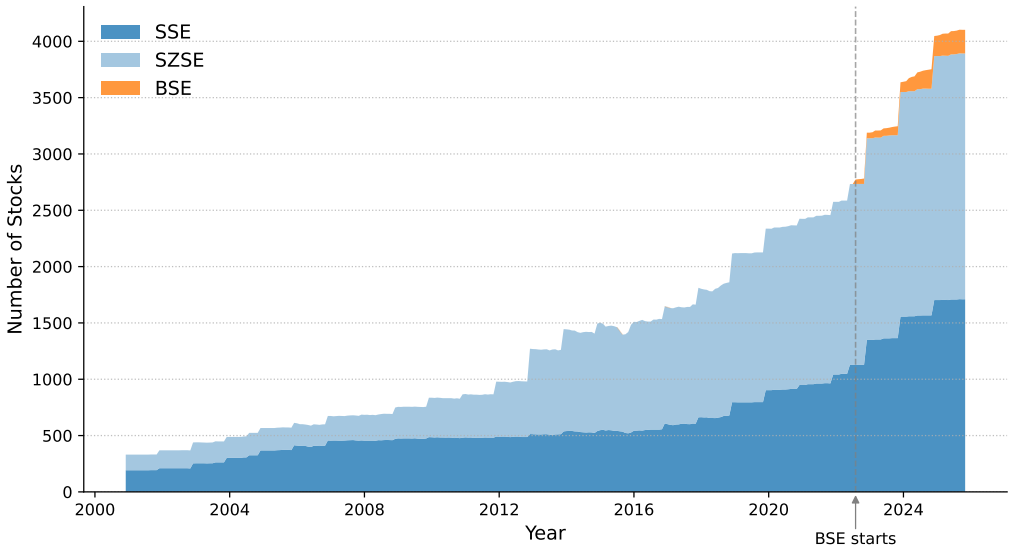


Fig. 1. Monthly Number of Stocks Across Exchanges. This figure plots the monthly number of stocks listed on the Shanghai Stock Exchange (SSE), Shenzhen Stock Exchange (SZSE), and Beijing Stock Exchange (BSE) from November 2000 to October 2025. Firms listed on the BSE enter the sample beginning in July 2022.

2 Data and Cross-Market Transfer Design

2.1 Data

In this study, we use a dataset of more than 4,100 A-share stocks listed on the SSE, SZSE, and BSE from November 2000 to October 2025. Figure 1 plots the monthly number of stocks across the three exchanges. By the end of the sample period, the constructed sample contains 1,710 SSE stocks, 2,182 SZSE stocks, and only 210 BSE stocks. While the SSE and SZSE provide long and relatively rich historical samples, BSE observations enter the panel only from July 2022, highlighting the data scarcity problem that motivates our cross-market transfer learning approach.

Following Hellum et al. (2025), we define stock returns net of the contemporaneous cross-sectional average within each market, making the prediction target more comparable across exchanges and more suitable for cross-market transfer learning. Specifically, for stock i in market m and month $t+1$, the dependent variable is defined as

$$r_{i,m,t+1} = R_{i,m,t+1} - \bar{R}_{m,t+1}, \qquad \bar{R}_{m,t+1} = \frac{1}{N_{m,t+1}} \sum_{i=1}^{N_{m,t+1}} R_{i,m,t+1},$$

where $R_{i,m,t+1}$ denotes the realized monthly stock return and $N_{m,t+1}$ is the number of

stocks in market m in month $t+1$. The predictor set consists of 77 firm-level characteristics constructed following Leippold et al. (2022) for the Chinese stock market, covering dimensions such as size, beta, momentum, liquidity. We exclude characteristics that require more than three years of historical data to construct, given the relatively short sample history of the BSE. Stock return and quarterly financial statement are obtained from the Wind and CSMAR databases, respectively. Missing financial statement variables are forward-filled when appropriate, so that previously disclosed information remains available until subsequent reports are released. Detailed definitions and categories of the firm-level characteristics are provided in the Online Appendix. Before model estimation, all continuous characteristics are ranked cross-sectionally within each market and month and rescaled to the interval $[-1, 1]$.

2.2 Transfer Specifications

We consider the following predictive asset pricing model:

$$r_{i,m,t+1} = \theta_m' x_{i,m,t} + \varepsilon_{i,m,t+1} = (\theta_c + \theta_{l,m})' x_{i,m,t} + \varepsilon_{i,m,t+1}, \tag{1}$$

where $r_{i,m,t+1}$ denotes the market-demeaned return of stock i in market m at time $t+1$, $x_{i,m,t}$ is a p-dimensional vector of firm-level characteristics, and $\varepsilon_{i,m,t+1}$ is the error term. The market-specific predictive coefficient θ_m is decomposed into a China-wide component θ_c and a local component $\theta_{l,m}$. The China-wide component captures common predictive patterns shared across Chinese stock markets, whereas the local component captures market-specific deviations. This decomposition justifies the use of transfer learning to leverage information from mature markets for prediction in emerging venues.

The local benchmark (Local) estimates the BSE predictive coefficient using only BSE observations. Let D_{BSE} denote the set of BSE training observations. Specifically, we use the Elastic Net (ENet) estimator

$$\hat{\theta}_{\text{BSE}}^{\text{Local}} = \arg\min_{\theta \in \mathbb{R}^p} \left\{ \frac{1}{|D_{\text{BSE}}|} \sum_{(i,t) \in D_{\text{BSE}}} \left(r_{i,\text{BSE},t+1} - \theta' x_{i,\text{BSE},t} \right)^2 + P_{\lambda,\alpha}(\theta) \right\}, \tag{2}$$

where $P_{\lambda,\alpha}(\theta)$ is the elastic net penalty defined as

$$P_{\lambda,\alpha}(\theta) = \lambda \left(\alpha \|\theta\|_1 + \frac{1-\alpha}{2} \|\theta\|_2^2 \right). \tag{3}$$

Here, $\lambda \geq 0$ controls the overall strength of regularization, and $\alpha \in [0,1]$ determines the relative weight between the L_1 and L_2 penalties. This model serves as a purely local benchmark because it does not use any source-market information.

To evaluate whether source-market information improves BSE return prediction, we compare the local benchmark with three transfer specifications: hard transfer (Hard), soft transfer (Soft), and two-stage estimation (2Stage). Let $s \in \{\text{SSE}, \text{SZSE}, \text{POOL}\}$ denote the source market, where POOL refers to the combined SSE and SZSE sample. Hard directly transfers source-market coefficients to the BSE, while Soft and 2Stage allow BSE data to adjust the transferred information. To implement these local adjustments, we adopt the Generalized Elastic Net (GENet) proposed by Hellum et al. (2025), which replaces the zero-centered shrinkage target in ENet with a source-market prior.

Given a prior coefficient vector $\hat{\theta}_0$, GENet estimates the BSE predictive coefficient as

$$\hat{\theta}_{\text{BSE}}^{\text{GENet}}(\hat{\theta}_0) = \arg\min_{\theta \in \mathbb{R}^p} \left\{ \frac{1}{|D_{\text{BSE}}|} \sum_{(i,t) \in D_{\text{BSE}}} (r_{i,\text{BSE},t+1} - \theta' x_{i,\text{BSE},t})^2 + P_{\lambda,\alpha}\left(\theta - \nu \hat{\theta}_0\right) \right\}. \tag{4}$$

Here, $\hat{\theta}_0$ is a prior coefficient vector estimated from source-market information, and $\nu \in [0,1]$ controls the strength of transfer. A larger ν shrinks the BSE estimate more strongly toward the source-market prior, whereas a smaller ν allows greater reliance on BSE-specific information. When $\nu = 0$, GENet reduces to the local ENet benchmark.

As shown by Hellum et al. (2025), GENet also has a useful residual-learning interpretation. Writing $\theta = \nu\widehat{\theta}_0 + \widetilde{\theta}$, the final prediction can be expressed as $\widehat{r}_{i,\text{BSE},t+1} = \nu\widehat{\theta}_0' x_{i,\text{BSE},t} + \widetilde{\theta}' x_{i,\text{BSE},t}$. The first term represents the transferred source-market prediction, while the second term captures the BSE-specific correction learned from local residual variation. This interpretation highlights that transfer learning does not simply replace local information with source-market information; rather, it uses source-market estimates as a structured starting point and allows BSE data to adjust the prediction.

Based on this framework, we consider four model specifications, summarized in Table 1. Local serves as the purely BSE-based benchmark. Hard transfer tests whether

Table 1
Transfer specifications. This table summarizes the four model specifications used in the empirical analysis. The columns indicate whether each model uses source-market information, whether it allows BSE-specific adjustment, and how the predictive coefficient is estimated.

Model	Source information	BSE adjustment	Estimation idea
Local	No	Yes	Estimate ENet using BSE data only.
Hard	Yes	No	Apply source-market coefficients directly to BSE.
Soft	Yes	Yes	Estimate BSE coefficients using GENet with source-market coefficients as the prior.
2Stage	Yes	Yes	Estimate a pooled source–BSE prior, then re-estimate BSE coefficients using GENet.

source-market predictive relationships can be directly transferred to BSE without local adaptation. Soft transfer uses source-market coefficients as the GENet prior and allows BSE data to estimate a local correction. The two-stage estimation differs from soft transfer in how the prior is constructed. Its first-stage estimate is obtained from a pooled source–BSE training sample, and the second stage further adapts this prior to BSE-specific return patterns through GENet.

2.3 Empirical Evaluation Design

In the baseline analysis, we adopt a fixed out-of-sample evaluation design. All model parameters, source-market priors, and hyperparameters are estimated using observations available up to December 2024. The fitted models are then evaluated on BSE returns from January 2025 to October 2025. This design ensures that no observations from the baseline test period are used in model estimation, source-market prior construction, or hyperparameter selection.

For each model and source-market choice, the hyperparameters are selected within the corresponding training sample using a 10-fold time-series cross-validation procedure. To avoid look-ahead bias, we deviate from the cross-sectional splitting approach of Hellum et al. (2025) and strictly preserve the temporal ordering of the data. Specifically, the training sample is divided into ten consecutive, non-overlapping blocks. For each iteration $k = 2, \ldots, 10$, the model is trained on the first $k-1$ blocks and validated on the subsequent

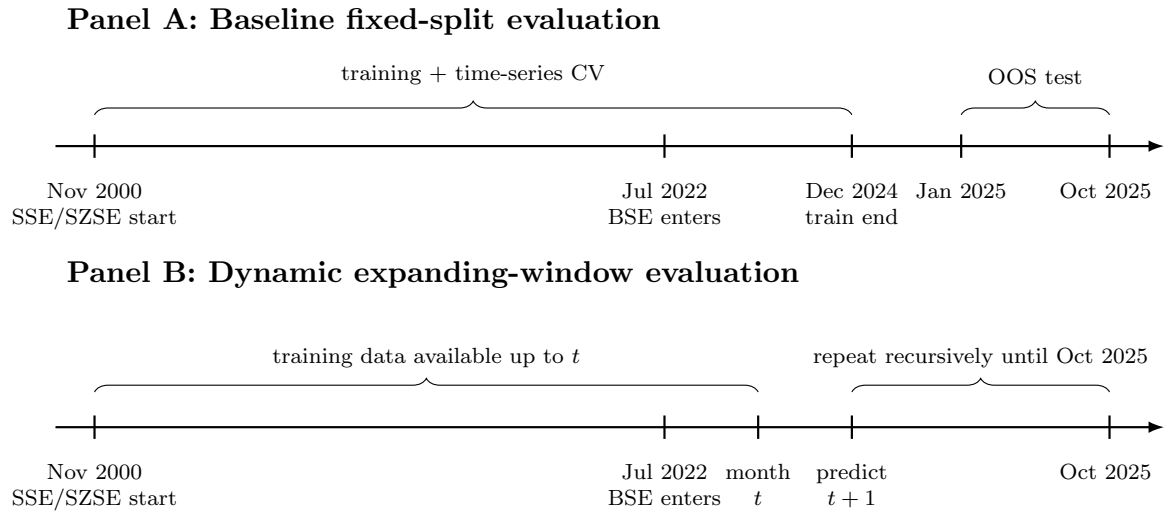


Fig. 2. Empirical evaluation design. Panel A illustrates the baseline fixed-split out-of-sample evaluation. Panel B illustrates the dynamic expanding-window evaluation, where all model components are re-estimated recursively using training data available up to each month t, including source-market information and the expanding BSE sample.

block k, thereby expanding the training window over time. The optimal hyperparameter combination is selected by minimizing the average validation Mean Squared Error (MSE) across folds. We use Bayesian optimization implemented with Optuna and run 100 trials for each model, searching over $\alpha, \nu \in [0, 1]$ and $\lambda \in [10^{-5}, 10^{-2}]$ on a logarithmic scale.

The baseline fixed-split design is complemented by a dynamic expanding-window analysis. In this exercise, the BSE training sample is allowed to expand over time, and model estimation is repeated recursively for each test month. For a prediction of month $t+1$, source-market priors, BSE local models, BSE-specific adjustment terms, and hyperparameters are estimated using only information available up to month t. Hyperparameters are selected within each expanding training window using the same time-series cross-validation procedure described above. The fitted model is then applied to firm characteristics observed at month t to predict BSE returns in month $t+1$. This dynamic design allows us to examine how the value of transfer learning changes as the target-market sample gradually accumulates.

2.4 Performance Measures

Our evaluation of model performance incorporates both statistical and economic dimensions. Following Gu et al. (2020), we quantify predictive accuracy using the monthly

out-of-sample R^2 on the BSE:

$$R^2_{\mathrm{OOS},t,S} = 1 - \frac{\sum_{i\in\mathcal{T}_t}\left(r_{i,t+1} - \hat{r}^{(S)}_{i,t+1}\right)^2}{\sum_{i\in\mathcal{T}_t} r^2_{i,t+1}}, \tag{5}$$

where $\mathcal{T}_t$ denotes the set of BSE stocks in the test sample at month t, and $\hat{r}^{(S)}_{i,t+1}$ is the predicted market-demeaned return generated by model specification S, with S referring to either the local benchmark or a transfer learning model. We report the average monthly out-of-sample R^2 over the test period for each model. To examine whether differences in predictive accuracy are statistically significant, we further conduct the Diebold and Mariano (2002) test based on pairwise forecast errors.

We then evaluate the economic value of the forecasts through portfolio-sorting analysis. At the end of each month t, BSE stocks in the test sample are sorted into five quintiles according to their predicted returns from each model specification. Stocks within each quintile are value-weighted using their market capitalization at the end of month t, and all portfolios are rebalanced monthly. Based on these quintile portfolios, we construct two strategies. The long–short strategy buys stocks in the highest predicted-return quintile and sells stocks in the lowest predicted-return quintile. The long-only strategy holds only the highest predicted-return quintile. The long–short strategy captures the return spread implied by the model forecasts, while the long-only strategy reflects a more implementable setting under short-selling constraints in the Chinese stock market.

3 Empirical Results

3.1 Predictive Gains from Source Markets

Figure 3 summarizes the main out-of-sample results. Panel A evaluates predictive accuracy using the average monthly R^2_{OOS}, while Panel B reports the annualized Sharpe ratio of the long-short portfolios formed based on predicted returns from each model.

The local model, which relies exclusively on BSE data, delivers the weakest predictive performance, with an average monthly R^2_{OOS} of 0.36%. In contrast, all transfer learning models achieve substantially higher predictive accuracy, with R^2_{OOS} ranging from 1.01% to 1.53%. These gains correspond to an improvement of roughly three to four times

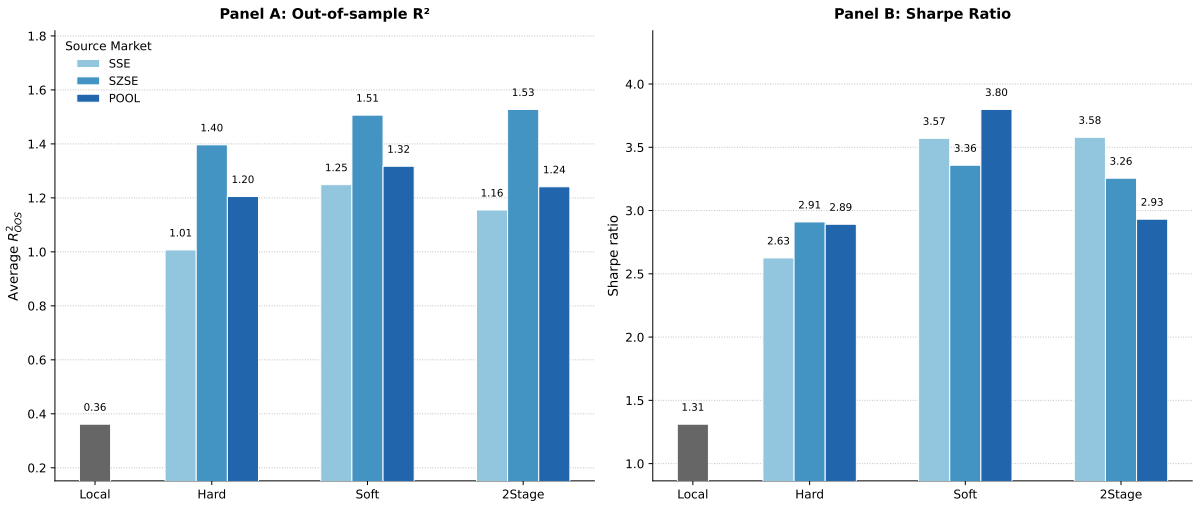


Fig. 3. Out-of-sample performance of local and transfer learning models. Panel A reports the average monthly out-of-sample R^2. Panel B reports the annualized Sharpe ratio of the corresponding long-short portfolios. SSE, SZSE, and POOL indicate the source markets used for transfer learning.

relative to the local benchmark, suggesting that predictive relations learned from mature Chinese stock markets contain useful information for BSE return prediction.

Panel B shows that the gains in predictive accuracy are also reflected in portfolio performance. The local model generates an annualized Sharpe ratio of 1.31 for the long-short portfolio, whereas transfer learning models raise the Sharpe ratio to between 2.63 and 3.80. The highest Sharpe ratio is achieved by soft transfer using the pooled source market, reaching 3.80. This result suggests that while the SZSE provides the strongest source for statistical prediction, combining information from both mature exchanges can further improve the risk-adjusted performance of trading strategies.

Table 2 further evaluates whether the predictive gains are statistically significant using pairwise Diebold–Mariano tests. The first row shows that every transfer learning specification significantly outperforms the local benchmark. The hard transfer models are significant at the 10% level, while all soft transfer models are significant at the 5% level. The two-stage models also deliver positive and significant improvements relative to the local model. These results suggest that the advantage of transfer learning is not limited to a single model specification or source market.

Comparisons among transfer learning models are more mixed. Although SZSE-based specifications tend to perform strongly in terms of predictive accuracy, no single transfer method uniformly dominates across all pairwise comparisons. Overall, the evidence sug-

Table 2
Comparison of monthly out-of-sample predictive accuracy using Diebold-Mariano tests. This table reports pairwise Diebold–Mariano test statistics for forecast errors across competing models. Positive (negative) values indicate that the model in the column outperforms (underperforms) the model in the corresponding row. Stars indicate statistical significance where *, **, and *** correspond to $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively.

	Hard (SSE)	Hard (SZSE)	Hard (POOL)	Soft (SSE)	Soft (SZSE)	Soft (POOL)	2Stage (SSE)	2Stage (SZSE)	2Stage (POOL)
Local	1.68*	1.83*	1.74*	2.55**	2.19**	2.40**	2.17**	1.82*	1.69*
Hard (SSE)		2.04**	1.87*	1.10	2.48**	1.11	0.60	1.89*	1.59
Hard (SZSE)			-2.05**	-0.40	0.94	-0.21	-0.62	1.29	-2.81***
Hard (POOL)				0.31	2.18**	0.46	-0.02	1.81*	0.81
Soft (SSE)					1.13	0.85	-1.35	0.67	-0.13
Soft (SZSE)						-0.84	-1.39	-0.05	-2.15**
Soft (POOL)							-2.47**	0.49	-0.31
2Stage (SSE)								0.87	0.15
2Stage (SZSE)									-2.15**

gests that the main performance gain comes from incorporating cross-market predictive information, rather than from the choice of a particular transfer specification alone.

The selected transfer parameters provide further evidence that source-market information plays an important role in the fitted BSE models. The optimal ν values are close to one across all source markets and transfer methods, ranging from 0.9612 to 0.9997, as reported in the Online Appendix. This indicates that the cross-validation procedure assigns substantial weight to the transferred source-market prior, while still allowing the BSE data to provide local adjustment.

3.2 Portfolio Gains from Transferred Signals

We next examine whether the predictive gains translate into economically meaningful portfolio performance. Following the portfolio construction described in Section 2, we sort BSE stocks into quintiles based on predicted returns from each model and evaluate both long-short and long-only strategies.

For the long-short strategies, the upper panel shows a clear separation between the local benchmark and the transfer learning models. The local portfolio increases only moderately during the out-of-sample period and experiences a noticeable decline after July 2025. In contrast, most transfer-based portfolios follow a more persistent upward trend and remain well above the local benchmark throughout most of the test period.

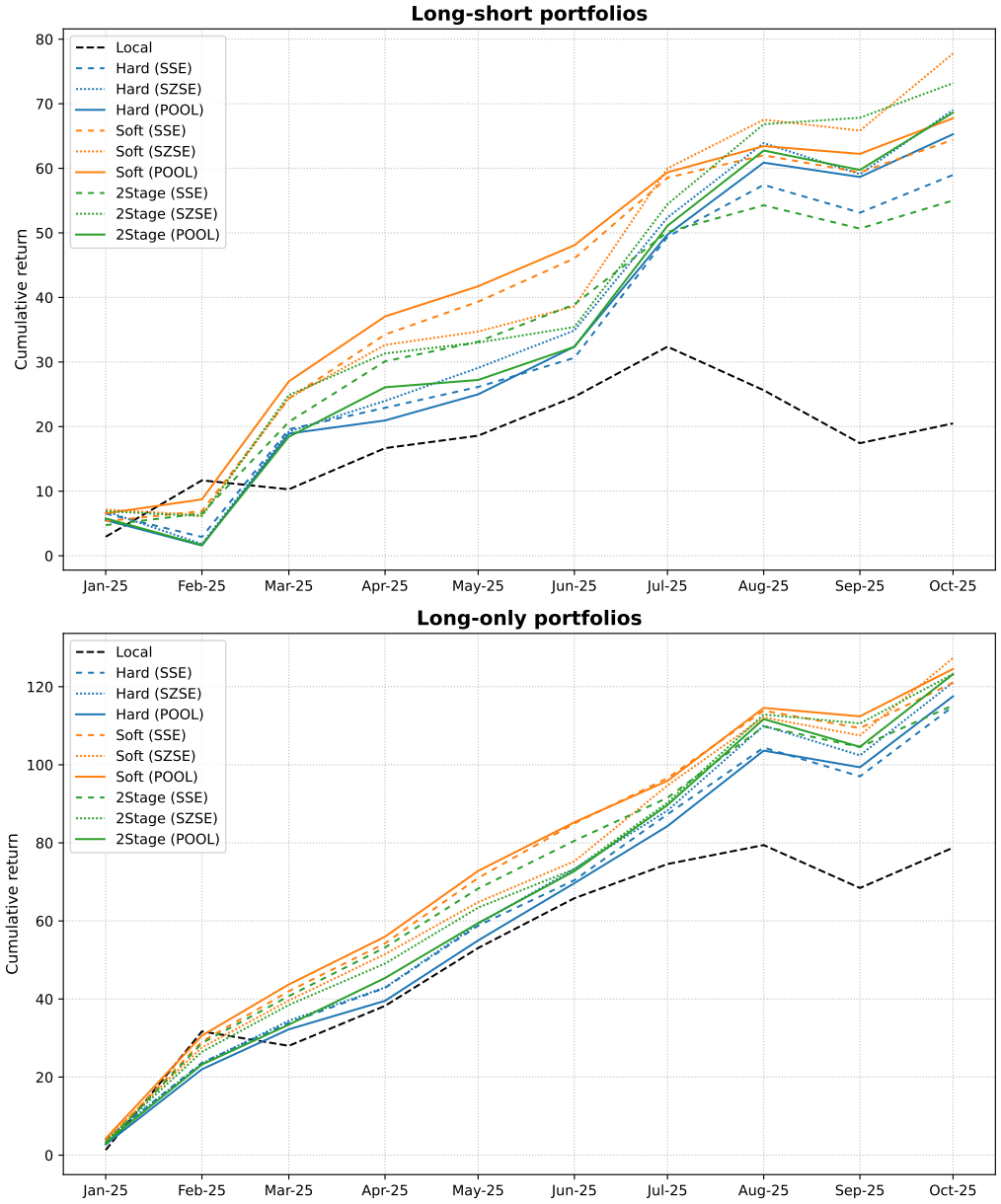


Fig. 4. Cumulative portfolio performance of local and transfer learning models. The upper panel reports long-short portfolios, and the lower panel reports long-only portfolios. Portfolios are formed according to predicted returns, value-weighted within each portfolio, and rebalanced monthly. Cumulative returns are reported in percent.

By the end of the test period, the local long-short portfolio earns a cumulative return of about 20%, whereas most transfer learning portfolios reach cumulative returns between 50% and 80%. This result suggests that cross-market transfer improves not only return forecasts, but also the cross-sectional ranking of BSE stocks.

Given the restrictions and implementation difficulties associated with short selling in the Chinese stock market, Figure 4 also reports long-only portfolio performance in the lower panel. All long-only portfolios generate positive cumulative returns, partly

Table 3
Portfolio performance based on model predictions. This table reports the performance of long-short and long–only portfolios formed each month by sorting stocks into quintiles based on predicted returns. Portfolios are value-weighted and rebalanced monthly over the out-of-sample period from January 2025 to October 2025. "Avg" denotes the average realized monthly return (%); "Std" is the standard deviation of monthly returns (%); "S.R." represents the annualized Sharpe ratio; "Skew" and "Kurt" refer to return skewness and kurtosis, respectively; "Max DD" reports the maximum portfolio drawdown (%); and "Turnover" reports the average monthly one-way portfolio turnover (%).

	Local	Hard			Soft			2Stage		
		SSE	SZSE	POOL	SSE	SZSE	POOL	SSE	SZSE	POOL
Long-Short										
Avg	1.99	4.91	5.57	5.32	5.20	6.08	5.40	4.56	5.80	5.54
Std	4.93	6.32	6.48	6.22	4.92	6.14	4.81	4.29	6.03	6.39
S.R.	1.31	2.63	2.91	2.89	3.57	3.36	3.80	3.58	3.26	2.93
Skew	-0.66	0.68	0.11	0.58	1.18	0.79	1.49	0.65	1.04	0.33
Kurt	2.47	3.09	3.26	3.27	5.28	2.84	6.21	4.11	3.22	2.76
Max DD	11.29	3.40	4.75	3.71	1.64	1.01	0.73	2.34	0.71	3.91
Turnover	20.38	121.10	117.98	124.57	78.56	100.05	84.10	74.86	105.66	114.77
Long-Only										
Avg	6.36	8.09	8.40	8.21	8.44	8.72	8.61	8.17	8.51	8.50
Std	9.79	5.67	5.78	5.48	6.87	6.18	6.79	7.07	6.03	5.78
S.R.	2.21	4.86	4.96	5.11	4.19	4.82	4.33	3.94	4.82	5.02
Skew	1.54	-0.27	-0.44	-0.06	1.31	0.73	1.60	1.37	0.94	-0.36
Kurt	6.82	5.64	5.28	4.62	6.94	6.32	7.49	7.09	5.71	5.45
Max DD	6.13	3.61	3.62	2.09	2.10	2.18	1.01	2.51	1.08	3.36
Turnover	12.48	67.22	64.61	69.48	50.83	56.93	54.04	49.47	58.31	64.21

reflecting favorable BSE market conditions during the test period. However, the transfer learning portfolios consistently outperform the local benchmark, indicating that their advantage is not solely driven by the overall market trend. Instead, transferred predictive signals help identify stocks with stronger subsequent performance within the BSE universe.

Table 3 provides a more detailed comparison of portfolio performance. For long-short portfolios, the local model generates an average monthly return of 1.99%, an annualized Sharpe ratio of 1.31, and a maximum drawdown of 11.29%. In comparison, transfer learning portfolios deliver substantially higher average monthly returns, ranging from 4.56% to 6.08%, and Sharpe ratios between 2.63 and 3.80. The highest average return is achieved by soft transfer from the SZSE. Transfer learning models also reduce downside risk, with maximum drawdowns mostly below 5%.

The long-only results show a similar pattern. The local model achieves an average monthly return of 6.36% and a Sharpe ratio of 2.21. Transfer learning portfolios in-

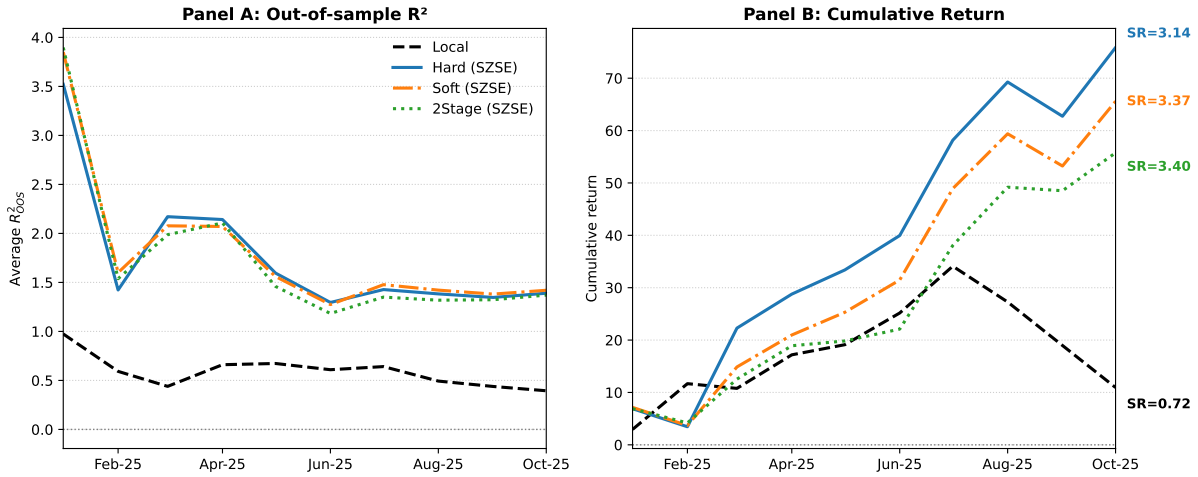


Fig. 5. Dynamic out-of-sample performance starting with 30 months of BSE training history. Panel A reports the average monthly out-of-sample R^2. Panel B reports cumulative long-short portfolio returns, with the final Sharpe ratio labeled at the end of each line. All transfer learning models in this figure use the SZSE as the source market. The test period is aligned with the baseline out-of-sample evaluation period.

crease average monthly returns to 8.09%–8.72% and raise Sharpe ratios to as high as 5.11. Moreover, the volatility and maximum drawdown of most transfer-based long-only portfolios are lower than those of the local benchmark, suggesting that the improvement is not achieved by simply taking on more risk.

Taken together, the portfolio results suggest that the statistical gains documented earlier are accompanied by economically meaningful portfolio improvements over the short out-of-sample evaluation window. Nevertheless, given the limited length of the baseline test period, these results should be interpreted as early evidence rather than as proof of a long-run performance pattern.

3.3 Transfer Gains with Expanding BSE Samples

The results above are based on a fixed training and test split. Since the BSE is a newly established market, the amount of available target-market history changes substantially over time. We therefore further examine dynamic out-of-sample performance as the BSE training window expands. This exercise allows us to evaluate whether the transfer learning advantage is concentrated only in the early stage of the sample or remains present as more local BSE observations become available. Given the strong baseline performance of SZSE-based transfer models, we focus on the SZSE as the source market in this analysis.

Figure 5 reports the dynamic performance when the initial BSE training window con-

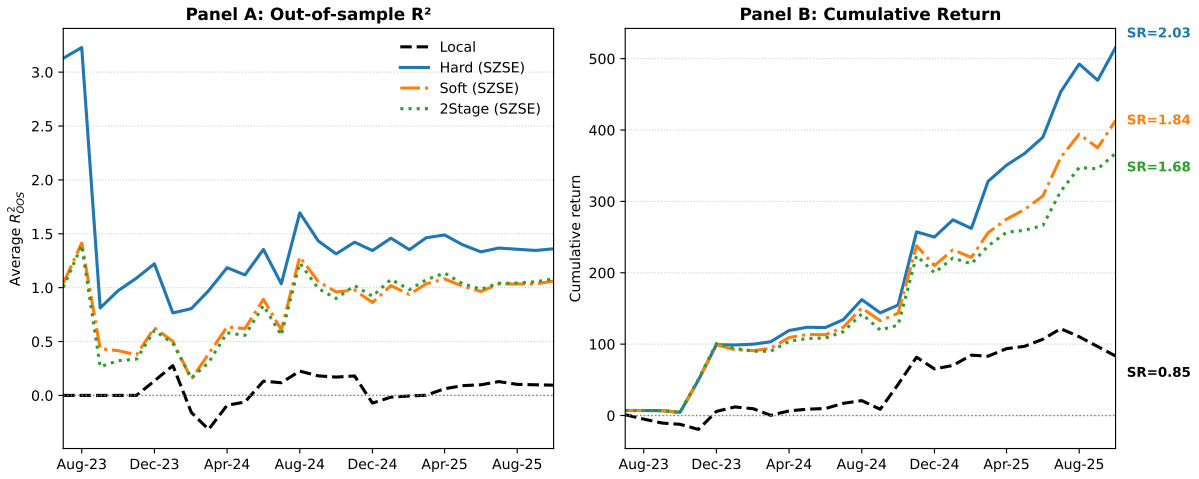


Fig. 6. Dynamic out-of-sample performance starting with 12 months of BSE training history. Panel A reports the average monthly out-of-sample R^2. Panel B reports cumulative long-short portfolio returns, with the final Sharpe ratio labeled at the end of each line. All transfer learning models in this figure use the SZSE as the source market. This figure illustrates model performance when the BSE training history is more limited.

tains 30 months of observations, which aligns the evaluation period with the baseline out-of-sample test. Panel A shows that all transfer learning models consistently outperform the local benchmark in terms of out-of-sample predictive accuracy. Although the R^2_{OOS} declines after the first test month, the transfer models remain above the local model throughout most of the evaluation period. This pattern suggests that the predictive advantage of cross-market transfer is not driven by a single month, but persists during the baseline test period.

Panel B of Figure 5 presents the corresponding cumulative long-short portfolio returns. The local portfolio generates only a modest cumulative gain and its performance weakens toward the end of the test period. By contrast, the transfer-based portfolios follow a much stronger upward trend and end the period with substantially higher cumulative returns. The final Sharpe ratios are 3.14, 3.37, and 3.40 for Hard, Soft, and 2Stage, respectively, compared with 0.72 for the local benchmark. These results confirm that the dynamic predictive gains also translate into economically meaningful portfolio performance.

Figure 6 considers a more data-scarce setting in which the initial BSE training window contains only 12 months of observations. In this case, the local model is substantially weaker and more unstable. Its average out-of-sample R^2 stays close to zero for much of the sample and even turns negative in early 2024, whereas the SZSE-based transfer models remain mostly positive throughout the period. Hard transfer performs particularly well,

with average R^2 values often above 1% and reaching about 1.7% around August 2024.

The economic performance in this data-scarce setting shows an even stronger separation. The cumulative long-short return of the local portfolio remains far below those of the transfer-based portfolios. Hard transfer delivers the strongest cumulative performance, followed by Soft and 2Stage. The final Sharpe ratios are 2.03, 1.84, and 1.68 for Hard, Soft, and 2Stage, respectively, all higher than the local benchmark of 0.85. This result suggests that when the BSE has very limited historical data, directly borrowing predictive information from a mature and related source market provides a particularly effective benchmark.

Overall, the dynamic evidence reinforces the main finding that transfer learning is especially valuable under target-market data scarcity. When the BSE training history is very short, Hard transfer performs particularly well because it relies primarily on well-estimated source-market information. As more BSE observations become available, adaptive transfer methods such as Soft and 2Stage become more competitive because they combine source-market signals with BSE-specific information. These results indicate that cross-market transfer remains useful throughout the process in which the BSE gradually accumulates its own trading history.

3.4 Robustness Checks

We conduct two robustness checks to examine whether the main findings are sensitive to implementation frictions and model specification. First, we incorporate transaction costs into the portfolio analysis. Second, we consider a non-linear extension to assess whether the advantage of transfer learning remains when the predictive function is allowed to be non-linear.

3.4.1 Trading Frictions

Although the portfolio results above suggest sizable economic gains from transfer learning, these gains are based on gross returns before implementation costs. This issue is particularly relevant because the strategies are rebalanced monthly and the transfer-learning portfolios generally exhibit higher turnover than the local benchmark. As shown in Table 3, the average monthly one-way turnover of transfer-based long-short portfolios ranges from 74.86% to 124.57%, compared with 20.38% for the local benchmark.

Table 4
Portfolio Performance after Transaction Costs. This table reports portfolio performance after transaction costs. Portfolios are rebalanced monthly. The long-short portfolio buys stocks in the highest predicted-return quintile and shorts stocks in the lowest predicted-return quintile. The long-only portfolio buys stocks in the highest predicted-return quintile. Monthly returns are reported in percent.

	Monthly return					Sharpe ratio				
Long-Short										
Transaction costs	0 bps	20 bps	40 bps	60 bps	80 bps	0 bps	20 bps	40 bps	60 bps	80 bps
Local	1.99	1.92	1.84	1.77	1.70	1.31	1.26	1.22	1.17	1.12
Hard (SSE)	4.91	4.48	4.04	3.60	3.17	2.63	2.38	2.14	1.90	1.65
Hard (SZSE)	5.57	5.14	4.72	4.30	3.87	2.91	2.69	2.46	2.23	2.00
Hard (POOL)	5.32	4.87	4.42	3.97	3.52	2.89	2.65	2.40	2.15	1.90
Soft (SSE)	5.20	4.92	4.63	4.35	4.07	3.57	3.38	3.18	2.99	2.79
Soft (SZSE)	6.08	5.72	5.36	5.00	4.64	3.36	3.15	2.95	2.74	2.54
Soft (POOL)	5.40	5.10	4.80	4.50	4.19	3.80	3.59	3.37	3.16	2.94
2Stage (SSE)	4.56	4.29	4.02	3.75	3.48	3.58	3.35	3.13	2.90	2.67
2Stage (SZSE)	5.80	5.42	5.04	4.66	4.27	3.26	3.04	2.82	2.61	2.39
2Stage (POOL)	5.54	5.12	4.71	4.30	3.88	2.93	2.72	2.50	2.27	2.05
Long-Only										
Transaction costs	0 bps	20 bps	40 bps	60 bps	80 bps	0 bps	20 bps	40 bps	60 bps	80 bps
Local	6.36	6.32	6.27	6.23	6.18	2.21	2.19	2.18	2.16	2.15
Hard (SSE)	8.09	7.85	7.60	7.36	7.12	4.86	4.72	4.58	4.44	4.30
Hard (SZSE)	8.40	8.17	7.93	7.70	7.47	4.96	4.83	4.70	4.57	4.44
Hard (POOL)	8.21	7.96	7.71	7.46	7.21	5.11	4.96	4.82	4.67	4.52
Soft (SSE)	8.44	8.26	8.08	7.90	7.71	4.19	4.10	4.01	3.92	3.83
Soft (SZSE)	8.72	8.51	8.31	8.10	7.90	4.82	4.71	4.60	4.49	4.38
Soft (POOL)	8.61	8.41	8.22	8.02	7.83	4.33	4.23	4.13	4.03	3.93
2Stage (SSE)	8.17	7.99	7.81	7.63	7.45	3.94	3.86	3.79	3.71	3.63
2Stage (SZSE)	8.51	8.30	8.09	7.88	7.67	4.82	4.70	4.58	4.46	4.34
2Stage (POOL)	8.50	8.27	8.04	7.81	7.58	5.02	4.89	4.76	4.63	4.50

A similar pattern is observed for long-only portfolios, although their turnover levels are lower. These higher turnover levels suggest that transferred predictive signals lead to more active portfolio rebalancing, making transaction costs an important consideration. To provide a more conservative assessment, we recompute portfolio performance under proportional transaction costs ranging from 0 to 80 basis points.

Table 4 shows that transaction costs reduce portfolio returns and Sharpe ratios, as expected, but the relative advantage of transfer learning remains robust. For long-short portfolios, the local model declines from a monthly return of 1.99% and a Sharpe ratio of 1.31 at 0 bps to 1.70% and 1.12 at 80 bps. In contrast, transfer-learning portfolios still deliver much stronger performance under the highest cost assumption. For example, Soft transfer from the SZSE generates a monthly return of 4.64% and a Sharpe ratio of 2.54 at 80 bps, while Soft transfer from the pooled source market maintains a Sharpe ratio of

2.94. The long-only results show a similar pattern. At 80 bps, the local model achieves a monthly return of 6.18% and a Sharpe ratio of 2.15, whereas transfer-learning portfolios still generate monthly returns between 7.12% and 7.90%, with Sharpe ratios as high as 4.52. These results suggest that the economic gains from cross-market transfer are not driven solely by ignoring trading frictions.

3.4.2 Nonlinear Transfer with Random Fourier Features

Section 2.2 shows that GENet can be interpreted as a residual-learning procedure, in which a source-market prediction is first transferred to the BSE and a BSE-specific correction is then learned from local residual variation. This representation naturally extends beyond linear predictive models. Instead of restricting the transferred component to the linear function $\widehat{\theta}_0' x_{i,\mathrm{BSE},t}$, we allow it to be a general prediction function $\widehat{f}_0(x_{i,\mathrm{BSE},t})$ estimated from source-market information. The BSE residual is defined as $\widetilde{r}_{i,\mathrm{BSE},t+1} = r_{i,\mathrm{BSE},t+1} - \nu \widehat{f}_0(x_{i,\mathrm{BSE},t})$, and a local correction function $\widehat{g}_{\mathrm{BSE}}(x_{i,\mathrm{BSE},t})$ is estimated using BSE observations. The final nonlinear transfer prediction is therefore $\widehat{r}_{i,\mathrm{BSE},t+1} = \nu \widehat{f}_0(x_{i,\mathrm{BSE},t}) + \widehat{g}_{\mathrm{BSE}}(x_{i,\mathrm{BSE},t})$. This formulation preserves the same transfer logic as GENet.

We implement this nonlinear extension using random Fourier features (RFF; Rahimi and Recht, 2007, 2008). RFF maps the original firm-characteristic vector into a higher-dimensional randomized feature space, which approximates nonlinear kernel functions while retaining the computational simplicity of regularized linear regression. Specifically, the source-market prediction function $\widehat{f}_0(\cdot)$ is estimated using RFF-transformed predictors and Elastic Net regularization. For hard transfer, the source-market RFF model is directly applied to BSE stocks. For soft transfer and two-stage estimation, we first compute the BSE residual after removing the transferred source-market prediction, and then estimate the local correction function $\widehat{g}_{\mathrm{BSE}}(\cdot)$ using Elastic Net on the BSE residuals. The hyperparameters are selected using the same time-series cross-validation procedure described in Section 2.3. The number of random features is fixed in the empirical implementation to keep the robustness check comparable across models.

Figure 7 reports the results. Panel A shows that the local RFF model performs poorly, with an average monthly out-of-sample R^2 of -0.39%. This negative value suggests that introducing non-linearity alone does not resolve the data scarcity problem in the BSE.

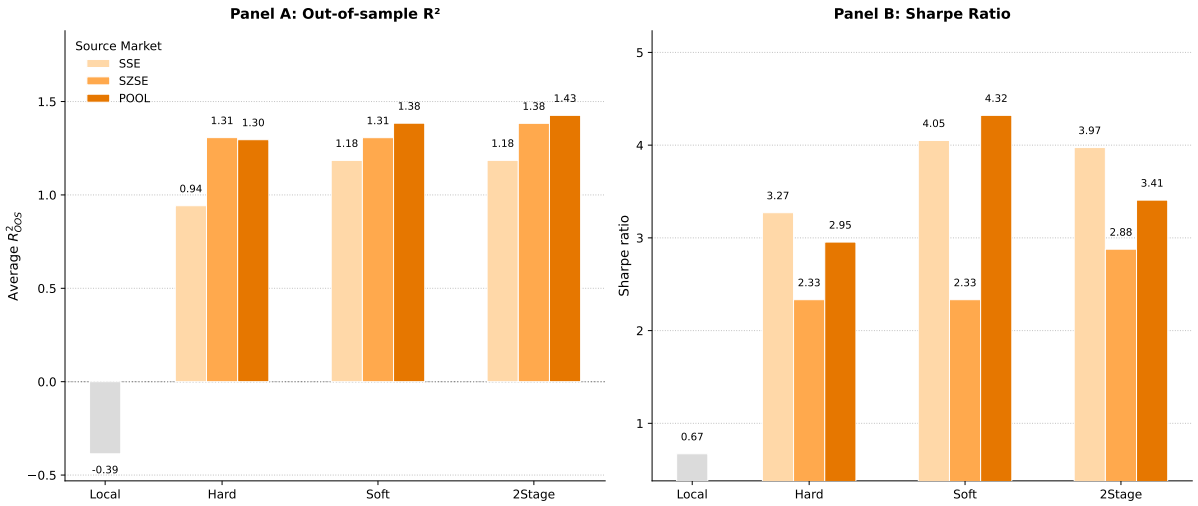


Fig. 7. Non-linear transfer learning performance based on random Fourier features (RFF). Panel A reports the average monthly out-of-sample R^2 of BSE return predictions. Panel B reports the annualized Sharpe ratio of the corresponding long-short portfolios. The RFF specification replaces the linear source-market predictor with a non-linear prediction function constructed from random Fourier features, while retaining the same transfer learning framework.

When the target-market sample is limited, a more flexible model may overfit local noise and generate weaker out-of-sample forecasts. In contrast, all RFF-based transfer learning models achieve positive out-of-sample R^2 values, ranging from 0.94% to 1.43%. The best predictive performance is obtained by 2Stage using the pooled source market, with an average R^2_{OOS} of 1.43%. This pattern likely reflects the greater data requirement of non-linear models. Compared with a single source market, the pooled sample provides a larger and more diversified training set, which helps stabilize the estimation of RFF-based non-linear predictive relationships.

Panel B shows that the RFF-based transfer models also deliver strong portfolio performance. The local RFF model generates a Sharpe ratio of only 0.67, whereas transfer learning models raise the Sharpe ratio to between 2.33 and 4.32. The highest Sharpe ratio is achieved by soft transfer using the pooled source market. This evidence suggests that non-linear transfer models continue to produce economically meaningful stock rankings. Overall, the RFF results show that the transfer learning advantage is not specific to the baseline linear specification. Rather, model flexibility becomes valuable when supported by informative and sufficiently rich source-market signals. Detailed portfolio performance of the RFF-based models is reported in Table OA.2 of the Online Appendix.

4 Conclusion

We examine whether predictive information from mature Chinese equity markets can improve stock return forecasting in the Beijing Stock Exchange, a nascent market characterized by severely limited historical data. Using the Generalized Elastic Net, we transfer predictive knowledge from the Shanghai and Shenzhen stock exchanges to the BSE and evaluate multiple transfer learning specifications under different source domain configurations.

Our results provide early evidence that cross-market transfer learning can improve BSE return prediction. Relative to the local benchmark, which uses only BSE data and achieves an average monthly out-of-sample R^2 of 0.36%, transfer learning models raise the R^2 to between 1.01% and 1.53%. The predictive gains also translate into economically meaningful portfolio performance. For long-short portfolios, transfer learning increases the average monthly return from 1.99% to 4.56%–6.08% and raises the annualized Sharpe ratio from 1.31 to as high as 3.80. The improvement is not achieved by taking on greater downside risk, as the maximum drawdown of the local strategy is 11.29%, while most transfer-based strategies have maximum drawdowns below 5%.

Dynamic analyses and robustness checks further support the main conclusion. The dynamic results show that transfer learning is most valuable when the BSE training history is very short. In the most data-scarce setting, hard transfer performs particularly well because it relies directly on well-estimated source-market signals. As more BSE observations become available, the performance of Soft and 2Stage improves, suggesting that these adaptive transfer methods can better exploit local information to refine the transferred signals. The results also remain robust after accounting for transaction costs. In the RFF extension, the local model performs poorly, but RFF-based transfer models still deliver positive out-of-sample R^2 and strong Sharpe ratios.

Overall, this study provides early evidence on machine learning-based return prediction in the BSE and highlights the potential of transfer learning as a promising approach to the data scarcity challenge in young equity markets. Future research may further examine broader sources of transferable information, time-varying transfer mechanisms, and the economic channels through which predictive signals are shared across China's multi-tier capital markets.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (NSFC Grant Numbers: 71931004). All errors are our own.

References

Chen, H., Cheng, Y., Liu, Y., and Tang, K. (2026). Teaching economics to the machines. NBER Working Paper 34713, National Bureau of Economic Research. Available at SSRN 6093305.

Chen, L., Pelger, M., and Zhu, J. (2024). Deep learning in asset pricing. *Management Science*, 70(2):714–750.

Choi, D., Jiang, W., and Zhang, C. (2025). Alpha go everywhere: Machine learning and international stock returns. *The Review of Asset Pricing Studies*, 15(3-4):288–331.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144.

Ergun, Z. E. and Sefer, E. (2025). Finsentiment: predicting financial sentiment through transfer learning. *Intelligent Systems in Accounting, Finance and Management*, 32(3):e70015.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Hellum, O., Pedersen, L. H., and Rønn-Nielsen, A. (2025). How global is predictability? the power of financial transfer learning. Available at SSRN 4620157.

Jeong, G. and Kim, H. Y. (2019). Improving financial trading decisions using deep q-learning: Predicting the number of shares, action strategies, and transfer learning. *Expert Systems with Applications*, 117:125–138.

Kraus, M. and Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems*, 104:38–48.

Leippold, M., Wang, Q., and Zhou, W. (2022). Machine learning in the chinese stock market. *Journal of Financial Economics*, 145(2):64–82.

Liu, Q., Tao, Z., Tse, Y., and Wang, C. (2022). Stock market prediction with deep learning: The case of China. *Finance Research Letters*, 46:102209.

Lu, L. and Ye, J. (2022). Beijing stock exchange and multi-tier capital markets: How china alters share listing and trading rules to promote sme finance? *The Company Lawyer*, 43(7):233–237.

Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

Rahimi, A. and Recht, B. (2007). Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20.

Rahimi, A. and Recht, B. (2008). Uniform approximation of functions with random bases. In *2008 46th annual allerton conference on communication, control, and computing*, pages 555–561. IEEE.

Weiss, K., Khoshgoftaar, T. M., and Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, 3(1):9.

Ye, N. (2022). Catching chinese unicorns: How the launch of beijing stock exchange reinforces china's securities law reform to promote sme finance and market inclusion. *International Company and Commercial Law Review*, 33(5).

Ye, R. and Dai, Q. (2022). A relationship-aligned transfer learning algorithm for time series forecasting. *Information Sciences*, 593:17–34.

Zhang, D., Lin, R., Wei, T., Ling, L., and Huang, J. (2023). A novel deep transfer learning framework with adversarial domain adaptation: application to financial time-series forecasting. *Neural Computing and Applications*, 35(34):24037–24054.

Zhang, H. (2024). The applicability of fama-french three-factor model to beijing stock market. *International Journal of Global Economics and Management*, 3:382–389.

Zhu, T. (2025). Latent factor model in asset pricing: A deep learning approach in the Chinese stock market. *Finance Research Letters*, 86:108519.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76.