MACHINE LEARNING AND SHRINKAGE IN DYNAMIC PANEL FORECASTING

Magdalena Cornejo (Universidad Torcuato Di Tella - CONICET)
Walter Sosa-Escudero (Universidad de San Andrés - CONICET)

Primera versión: Mayo 2026
Esta versión: Mayo 2026

Documento de Trabajo Nº 183
Departamento de Economía
Universidad de San Andrés

Vito Dumas 284, B1644BID, Victoria,
San Fernando, Buenos Aires, Argentina
Teléfono +54 11 7078 0400
Email: economia@udesa.edu.ar

Machine Learning and Shrinkage in Dynamic Panel Forecasting

Magdalena Cornejo*
Universidad Torcuato Di Tella - CONICET

Walter Sosa-Escudero†
Universidad de San Andrés - CONICET

May 2026

Abstract

This paper studies forecasting in dynamic panel data models with fixed effects. We compare the forecasting accuracy of conventional estimators— pooled OLS, fixed effects, Anderson–Hsiao, and Arellano–Bond—against shrinkage and regularization methods such as Ridge, LASSO, Elastic Net, empirical Bayes maximum likelihood and the recent unbiased risk estimation of Kwon (2026). Monte Carlo evidence shows that shrinkage methods substantially improve out-of-sample accuracy. An empirical application to firm-level leverage dynamics using Compustat data confirms the relevance of these findings for forecasting in corporate finance. Machine learning regularization can improve forecasting performance in dynamic panel settings while preserving the structural framework.

Keywords: Forecasting, Dynamic panel data, Machine learning, Regularization, Corporate finance.

JEL Codes: C53, C58.

*Corresponding author: School of Business, Universidad Torcuato Di Tella and CONICET, email: mcornejo@utdt.edu. 7350 Figueroa Alcorta Av., Buenos Aires, Argentina. +54 11 5169 7156.
Replication files and code are available at https://github.com/magdalenacornejo/forecasting-dynamic-panels

†Department of Economics, Universidad de San Andrés and CONICET. Email: wsosa@udesa.edu.ar

1 Introduction

Fixed effects occupy a central place in the analysis of panel data. In *static* specifications, controlling for fixed effects eliminates biases from unobserved, time-invariant heterogeneity, which explains their ubiquity in empirical research. However, in *dynamic* settings the presence of lagged dependent variables renders conventional strategies, such as within or first-difference transformations, inconsistent, a result formally established by Nickell (1981). This creates a fundamental tension: while fixed effects are essential to capture persistent heterogeneity, their estimation becomes problematic in dynamic environments, especially when the time dimension is limited. The econometric literature has proposed several solutions to this problem. Anderson and Hsiao (1981) introduced the use of valid instruments that arise naturally under first-differencing, an approach subsequently exploited in the GMM framework by Arellano and Bond (1991) and extended by Ahn and Schmidt (1995), Blundell and Bond (1998), and Kiviet (1995); Baltagi (2021) provide a recent, comprehensive overview of these ideas.

While these estimators address consistency, their small-sample performance has been the subject of considerable scrutiny, for two reasons. First, the so-called Nickell bias decreases as the time dimension T grows, but remains empirically relevant for realistic sample sizes. Second, the large-sample properties of instrumental variable estimators often fail to materialize in small samples, and in some cases may even reverse. To the point, the well-known simulation study of Judson and Owen (1999) reports substantial biases persisting even for $T = 30$, and that, in small samples, strategies that use few instruments (like the Anderson-Hsiao estimator) perform better than those that exploit multiple instruments, like the Arellano-Bond GMM estimator.

These limitations become particularly relevant when the objective is *forecasting* rather than consistent estimation. In such cases, the variance inflation from IV-based estimators can outweigh the benefits of bias reduction, making the optimal bias–variance trade-off for prediction fundamentally different from that for estimation.

Hence, a natural concern is whether a purposely biased estimator can deliver gains in predictive accuracy. A natural strategy consists of regularizing fixed effects in dynamic panel models. By bypassing instrumental variables, such approaches avoid the variance inflation associated with IV-based methods. At the same time, by shrinking unit-specific parameters toward a common structure, they introduce bias in a controlled way that may lead to substantial reductions in variance and improved predictive performance. In this context, shrinkage methods provide a natural way to address the bias–variance trade-off that arises in dynamic panel environments with many unit-specific parameters. Unlike traditional econometric procedures that prioritize asymptotic consistency, machine learning regularization methods explicitly allow for controlled bias in exchange for improved predictive accuracy and stability.

This paper examines how machine learning and shrinkage-based regularization methods perform in forecasting dynamic panel models with fixed effects. We consider both penalized machine learning estimators—such as LASSO, Ridge regression, and Elastic Net—and model-based shrinkage procedures, including empirical Bayes and unbiased risk estimation methods. These approaches differ in how the degree of regularization is selected, but all aim to identify an intermediate degree of heterogeneity between pooled and fully heterogeneous specifications. We evaluate their predictive accuracy against both naïve (OLS, FE) and consistent (Anderson–Hsiao, Arellano–Bond) econometric benchmarks using out-of-sample mean squared forecast error as the main performance criterion.

Our results indicate that shrinkage-based estimators can substantially improve predictive performance in dynamic panel settings. In particular, when the dynamic persistence is high (autoregressive parameter $\gamma = 0.8$), shrinkage methods consistently achieve the lowest mean squared forecasting errors across different combinations of individual (N) and temporal (T) dimensions. When persistence is low ($\gamma = 0.2$), the fixed-effects estimator often remains the best-performing alternative, although shrinkage estimators remain highly competitive, and frequently emerge as a clear second-best alternative under a wide range of panel dimensions. Moreover, even in settings where fixed effects are not sparse, the cost of 'unnecessary shrinkage' is negligible, making regularization a broadly applicable forecasting strategy. These findings highlight the potential of modern shrinkage methods to complement—and in several predictive contexts, outperform—traditional methods for dynamic panel analysis.

To illustrate the empirical relevance of these ideas, we apply the proposed framework to corporate capital structure dynamics using firm-level data from Compustat. This setting provides a natural environment characterized by persistent dynamics and substantial firm-specific heterogeneity, making it well suited to assess the predictive gains from shrinkage in practice.

More broadly, the paper contributes to the growing literature that integrates machine learning tools into econometric modeling while maintaining the interpretability and structural discipline of conventional dynamic panel frameworks.

The paper is organized as follows. Section 2 discusses the problem of biases and inconsistencies in dynamic panels, and the potential role of shrinkage procedures for prediction. Section 3 describes the empirical strategy in detail. The results of an extensive Monte Carlo exploration are presented and discussed in section 4. Section 5 illustrates the forecasting performance of each estimator in an empirical application. Section 6 offers concluding remarks.

2 Forecasting with dynamic panel models

Recent advances in machine learning have reshaped forecasting methodologies for dynamic panel data models. Pick and Timmermann (2024) review how predictive gains in panel settings depend on the time dimension T, the cross-sectional dimension N, and the degree of parameter heterogeneity. Babii et al. (2024) provide a comprehensive overview of how machine learning methods can address high-dimensionality, nonlinearities, and structural instability in macroeconomic forecasting, including panel data contexts. As part of an ongoing trend, recent studies show a clear shift from traditional estimation-focused approaches —primarily concerned with bias— toward new methods that prioritize out-of-sample predictive performance.

The panel forecasting literature broadly classifies estimators into homogeneous (pooled), heterogeneous, and shrinkage-based approaches. Homogeneous estimators, such as pooled OLS, assume common parameters across units and perform well when N is large and T is small, as they exploit cross-sectional variation to reduce estimation error. However, when slope or intercept heterogeneity is present, pooled models suffer from specification bias (Pesaran and Smith, 1995), which can deteriorate forecasting performance. On the other hand, fully heterogeneous models, such as unit-by-unit regressions or mean-group estimators (Pesaran et al., 1999), account for heterogeneity but often lead to high variance and overfitting when T is limited.

Shrinkage estimators offer a middle ground by introducing controlled bias to reduce variance. These include Bayesian hierarchical models and penalized regression methods such as LASSO, Ridge regression, and Elastic Net. By borrowing information across units, shrinkage methods partially pool the data and can deliver superior predictive accuracy, especially when the individual time dimension is short. For instance, Liu et al. (2020) develop an Empirical Bayes predictor based on Tweedie's formula, showing significant gains in forecasting individual outcomes in short panels. Their method leverages cross-sectional information to shrink fixed effects toward a common distribution, improving efficiency while retaining unit-level heterogeneity. More recently, Kwon (2026) proposes an alternative shrinkage approach based on an unbiased risk estimation (URE) criterion, which selects the degree of shrinkage by minimizing an estimate of the mean squared error without relying on strong distributional assumptions as in Empirical Bayes methods.

This study incorporates both model-based and machine learning shrinkage methods within a dynamic panel framework, explicitly allowing for sparse heterogeneity, and comparing their forecasting performance across a range of estimators in both short and long panels.

The problem of bias in dynamic panels with small T is well-known since Nickell (1981), and has traditionally been addressed through instrumental variables (IV) and generalized method-of-moments (GMM) techniques, such as Anderson and Hsiao (1982) and Arellano and Bond (1991). While these estimators are consistent as $N \to \infty$, they often suffer from high variance in finite samples and are sensitive to instrument proliferation and weak instruments. Simulation evidence in Judson and Owen (1999) and theoretical developments by Bun and Carree (2005) and Hsiao et al. (2002) show that bias-corrected least squares estimators can outperform GMM in terms of root mean squared error (RMSE), particularly in forecasting contexts. These findings suggest that consistency alone is not sufficient for good forecasting performance—variance and overfitting are also critical concerns.

More recent approaches combine the IV logic with shrinkage tools. Chernozhukov et al. (2024) propose an Arellano–Bond LASSO estimator that selects valid instruments via L_1-penalization, reducing overidentification bias and improving out-of-sample accuracy. This reflects a broader trend: the use of regularization to improve traditional estimators' finite-sample performance under complex or high-dimensional settings.

A growing body of empirical evidence supports the superior forecasting performance of shrinkage-based methods. Brücker and Siliverstovs (2006) find that fixed-effects and hierarchical Bayes estimators outperform pooled OLS, IV, and standard GMM in forecasting international migration flows. Similarly, Baltagi (2008) and Baltagi (2013) advocate the use of best linear unbiased predictors (BLUP) for dynamic panels, which are equivalent to Bayesian shrinkage with Gaussian priors on the individual effects. In the context of spatial panel models, Baltagi et al. (2014) show that accounting for cross-sectional dependence further improves forecast accuracy, emphasizing the importance of modeling both heterogeneity and dependence.

Recent machine learning applications further extend this literature. Camehl (2023) introduces a LASSO-based panel VAR estimator with structured penalties across equations, lags, and units, demonstrating substantial forecasting gains over standard VARs. Yang et al. (2024) provide a general framework for machine learning in panel data models, showing that models accounting for fixed effects and heterogeneity consistently outperform naïve pooled alternatives. To address structural change, Zhang et al. (2022) propose shrinkage-based quantile regressions that detect common breakpoints while allowing heterogeneous effects across units and quantiles. While these contributions focus on high-dimensional systems, general-purpose frameworks, or structural change detection, our approach instead concentrates on dynamic panel models with potentially sparse fixed effects and a systematic comparison of shrinkage and traditional estimators from a forecasting perspective.

Together, these contributions point toward a key insight: in forecasting dynamic panel data, estimators that carefully balance bias and variance—often through regularization or shrinkage—can outperform traditional unbiased methods. This is especially true when the cross-sectional dimension is large, the time dimension is short, and the underlying structure exhibits sparsity or heterogeneity. Our work builds on this literature by evaluating the forecasting performance of shrinkage methods that directly regularize the individual fixed effects, comparing them to classical IV-based estimators and demonstrating their advantages in terms of out-of-sample mean squared error.

3 The empirical setup

This section describes the data-generating process (DGP), simulation design, and the estimators under evaluation. Our Monte Carlo simulations are designed to assess the forecasting performance of both conventional and regularized estimators in dynamic panel data models with sparse fixed effects.

The DGP follows closely those in Judson and Owen (1999); Kiviet (1995); Arellano and Bond (1991), with modifications to incorporate sparsity in the fixed effects. Specifically, we consider a linear dynamic panel model of the form:

$$y_{i,t} = \gamma\, y_{i,t-1} + x_{i,t}'\beta + \eta_i + \varepsilon_{i,t}, \qquad |\gamma| < 1, \quad i = 1, \ldots, N;\ t = 1, \ldots, T + 10, \tag{1}$$

where $x_{i,t}$ is a $(K-1) \times 1$ vector of strictly exogenous regressors, $\varepsilon_{i,t} \sim$ i.i.d. $N(0, \sigma_\varepsilon^2)$ is the idiosyncratic error, and η_i denotes an individual-specific fixed effect.

To introduce structural sparsity in the fixed effects, we define a *sparsity parameter* $\delta \in [0, 1]$, which denotes the fraction of individuals whose fixed effects are set to zero *ex ante*. Formally, we generate

$$\eta_i = \begin{cases} 0, & \text{with probability } \delta, \\ \tilde{\eta}_i, & \text{with probability } 1 - \delta, \end{cases} \qquad \tilde{\eta}_i \overset{i.i.d.}{\sim} N(0, \sigma_\eta^2), \tag{2}$$

where $\sigma_\eta^2 > 0$ is fixed in the simulation design. A value $\delta = 0$ corresponds to the case of no sparsity (all individuals have non-zero fixed effects), whereas $\delta = 1$ corresponds to full sparsity (all $\eta_i = 0$). Intermediate values capture environments with limited or structured heterogeneity across units, allowing us to assess whether shrinkage estimators can exploit this structure to improve forecasting performance.

The covariates $x_{i,t}$ follow an AR(1) process, a common assumption in the dynamic panel data literature:

$$x_{i,t} = \rho\, x_{i,t-1} + \xi_{i,t}, \qquad |\rho| < 1, \quad \xi_{i,t} \sim \text{i.i.d. } N(0, \sigma_\xi^2)\ , \tag{3}$$

with $\xi_{i,t}$ independent of both η_i and $\varepsilon_{i,t}$. All other standard assumptions hold:

$$\begin{aligned} &\sigma_\varepsilon^2 > 0, \quad \sigma_\eta^2 > 0, \\ &E[\varepsilon_{i,t}\,\varepsilon_{j,s}] = 0 \quad \text{for } i \neq j \text{ or } t \neq s, \\ &E[x_{i,t}\,\varepsilon_{j,s}] = 0 \quad \forall\, i, j, t, s. \end{aligned} \tag{4}$$

These assumptions rule out any cross-sectional or serial correlation in the errors and enforce the strict exogeneity of the regressors. We generate data for $T + 10$ periods and discard the first 10 observations as burn-in (to mitigate initialization effects), so that the effective sample length for analysis is T.

In the simulation design, we fix $\sigma_\varepsilon^2 = \sigma_\eta^2 = 1$, $\sigma_\xi^2 = 4$, and $\rho = 0.5$. We consider two values of the autoregressive parameter, $\gamma \in \{0.2, 0.8\}$, to represent low- and high-persistence scenarios. We examine panels with $N \in \{20, 100\}$ cross-sectional units and time lengths $T \in \{5, 10, 20, 30\}$, covering a range from very short to moderately long panels. For each combination of N, T, γ, and δ, we perform 1,000 independent Monte Carlo replications. In each replication, a new dataset is drawn from the DGP specified above (using fixed random seeds for reproducibility) and all candidate estimators are applied. All results and the corresponding `R` code are available from the authors upon request.

Our evaluation focuses on one-step-ahead forecasts over a fixed forecasting horizon H. In each simulation replication, we split the T periods into an in-sample estimation window and an out-of-sample forecast window. Specifically, the first 80% of the time periods (T^*) are used for model estimation, while the remaining 20% (H) constitute the forecast horizon reserved for out-of-sample evaluation. For example, in a panel with $T = 30$, we would estimate each model using $T^* = 24$ periods and then generate forecasts for the last $H = 6$ periods. We compute forecast accuracy measures for each estimator, including the out-of-sample mean error (bias), variance, and mean squared error (MSE). Our primary metric of interest is the out-of-sample MSE, averaged across the H forecast periods and over all replications for a given design.

We compare a broad set of estimation methods. On the classical side, we include pooled OLS (which ignores the fixed effects), the fixed-effects (within) estimator (FE), the Anderson–Hsiao instrumental-variables estimator (AH; Anderson and Hsiao 1981), and the one-step (GMM1) and two-step (GMM2) Arellano–Bond GMM estimators (Arellano and Bond, 1991).

On the regularization side, we consider two classes of shrinkage estimators that differ in how the degree of shrinkage is determined. First, we include model-based shrinkage estimators applied to the fixed effects, namely the canonical Empirical Bayes maximum likelihood estimator (EBMLE) and the unbiased risk estimator (URE) recently proposed by Kwon (2026). These approaches shrink the estimated fixed effects toward zero using data-driven rules: EBMLE selects the shrinkage intensity by maximizing a marginal likelihood under a Gaussian prior on the fixed effects, while URE chooses the shrinkage parameter by minimizing an unbiased estimate of the mean squared error (risk).

Second, we consider machine learning shrinkage methods in which the regularization penalty is imposed exclusively on the fixed effects, leaving the lagged dependent variable and the observed regressors unpenalized: the Least Absolute Shrinkage and Selection Operator (LASSO), Ridge regression, and the Elastic Net (EN). For each regularized method, the penalty parameter is selected via a rolling-origin blocked cross-validation scheme using five sequential time blocks. The optimal value is determined according to the "one-standard error rule" applied to the mean squared validation loss.[1]

The motivation behind all shrinkage estimators is to manage the bias–variance trade-off by effectively imposing an intermediate degree of heterogeneity. Whereas pooled OLS assumes all $\eta_i = 0$ and the FE estimator allows each unit to have a distinct η_i, shrinkage approaches partially pool information across units by shrinking the fixed-effect estimates toward zero. Model-based methods (EBMLE and URE) achieve this through explicit assumptions on the distribution of the fixed effects and risk-based criteria, while machine learning methods rely on penalization and cross-validation. This introduces some bias but can substantially lower estimation variance. Given the sparse fixed-effect structure of our DGP, we expect that such shrinkage will yield lower out-of-sample MSE than both the naïve estimators (OLS and FE) and the conventional IV-based estimators (AH and GMM) that prioritize bias reduction over variance control.

4 The predictive performance of shrinkage methods in dynamic panels

The simulation results align with the theoretical arguments, both in qualitative patterns and in quantitative magnitudes. Table 3 begins by examining the performance of standard and regularized strategies when the goal is to *estimate* the key parameter of the dynamic model: the autoregressive coefficient, γ. Table 3 presents results organized as follows: we evaluate the performance of alternative estimators for two cross-sectional sample sizes ($N = 20$ and $N = 100$) and four time dimensions ($T^* = 4, 8, 16, 24$), which correspond to forecast horizons $H = 1, 2, 4, 6$, respectively. We consider two values for γ: 0.2 and 0.8.

For each configuration, we report the bias, variance and the mean squared error (MSE) of five 'standard' estimation strategies: ordinary least squares (OLS), OLS with individual fixed effects (FE), the Anderson/Hsiao estimator (AH), and two variants of the Arellano-Bond estimator (GMM1 and GMM2). The last five columns report results for *shrinkage* strategies: two model-based approaches, the empirical Bayes maximum likelihood estimator (EBMLE) and the unbiased risk estimator (URE), and three penalized regression methods, LASSO, Ridge, and Elastic Net. Methods are implemented as described in Section 3.

Table 3 focuses on the case of '0% sparsity' —that is, when no fixed effects are set to zero ex ante. A complete set of results covering a broader range of parameters is available upon request; however, the configurations presented here are representative of the qualitative and quantitative patterns observed throughout.

The first rows and columns of Table 3 reflect well-known empirical findings in this literature, such as those in Judson and Owen (1999). For instance, when $N = 20$, $T = 4$, and $\gamma = 0.2$, both OLS and FE estimators exhibit severe bias. The Anderson–Hsiao estimator reduces this bias considerably. The bias–variance trade-off is clearly present: the highly biased OLS estimator shows very low variance, whereas the bias correction achieved by the AH estimator comes at the cost of a much larger variance relative to OLS. For instance, the bias of the AH estimator is -0.00410, compared to -0.23014 of OLS. Variances are 0.01332 and 0.00302, respectively. In this particular case, the AH dominates the OLS estimator in MSE (0.01334 and 0.05599). It is interesting to remark that even when AH dominates both OLS and FE estimators in terms of bias, and that OLS has smaller variance than AH and FE, though severely biased, the FE is still a relevant competitor in terms of MSE (0.01703).

Interestingly, even though the regularized estimators are dominated by the AH strategy in terms of bias, their performance is strong, especially when compared to their non-regularized counterparts, such as OLS or FE. In particular, EBMLE and URE display higher bias than AH, but remain less biased than LASSO, Ridge, and Elastic Net. A similar pattern holds for variance: while OLS exhibits the lowest variance, the regularized alternatives -especially EBMLE and URE- achieve a substantial reduction in variance relative to IV-based competitors (AH, GMM1 and GMM2), which translates into a clear MSE advantage.

Even at this stage of the analysis an early pattern emerges: as studied in the context of parameter estimation (e.g. Judson and Owen, 1999), the Anderson–Hsiao estimator effectively reduces bias, but at a substantial cost in terms of variance. At the other extreme, naïve OLS (which ignores fixed effects) is severely biased but highly precise, while Arellano–Bond-type estimators perform poorly across the board. A first result of this paper is that regularized strategies dominate in terms of MSE —an outcome that, as expected, arises from their ability to produce relatively small biases alongside substantial reductions in variance. The rest of Table 3 reflects this discussion: AH dominates in terms of bias, OLS in terms of variance, while regularized estimators perform relatively well in terms of MSE in several configurations.

Tables 4 and 5 are more closely aligned with the central goal of this paper: assessing the *predictive* performance of dynamic panel models. Specifically, Table 4 presents in-sample forecasting results and Table 5 focuses on the main empirical question of the paper: the *out-of-sample* forecasting performance. Consequently, we discuss Table 5 in more detail.

The most important result of this paper is that the regularized alternatives tend to dominate in terms of MSE, and remain highly competitive —often emerging as the second-best alternative after FE across many configurations. Quantitatively, the differences are substantial. Consider, for example, the case with $N = 20$ and $T = 4$, and $\gamma = 0.2$. In this setting, although the Nickell bias remains substantial given the short time dimension of the panel, it is effectively handled by the IV estimators, but it does not seem to affect the performance of regularized methods.

The key point is that the combined bias–variance profile of the regularized estimators leads to a drastic reduction in MSE. For instance, standard estimators such as GMM2—the least biased estimator in this specification- yield an MSE of approximately 2.27, whereas the URE estimator achieves an MSE of 1.38 —only 61% of that of the GMM estimator, and LASSO, Ridge and Elastic Net achieve an MSE of about 1.77. This result offers an early insight: when the objective is out-of-sample forecasting in dynamic panel settings, the 'price' paid to correct for biases using standard IV-based methods is high, while regularized strategies deliver a dramatically better performance due to their explicit ability to negotiate the bias–variance trade-off.

A relevant dimension to explore is the role of the degree of sparsity. Additional tables report results for alternative levels of sparsity. To preserve space, we present and discuss the results related to sparsity graphically; detailed numerical outputs are available upon request.

Figures 1 and 2 present four diagrams summarizing the out-of-sample MSE for the eight forecasting strategies. In each graph the horizontal axis represents the degree of sparsity in the structure of individual fixed effects, while the vertical axis reports the corresponding MSE values. As detailed in the previous section, zero sparsity means that all individual specific fixed effects are non-zero and different across individual units, and 1 means that all fixed effects are set equal to zero. Intermediate cases correspond to a situation where a proportion (from 0 to 1) of the fixed effects are set to 0.

Figure 1 corresponds to the moderately short-panel case with $T = 10$ (i.e., $T^* = 8$, $H = 2$), whereas Figure 2 shows results for a longer time dimension, $T = 30$ (i.e., $T^* = 24$, $H = 6$). Each diagram presents different combinations of the cross-sectional dimension N and the autoregressive parameter γ.

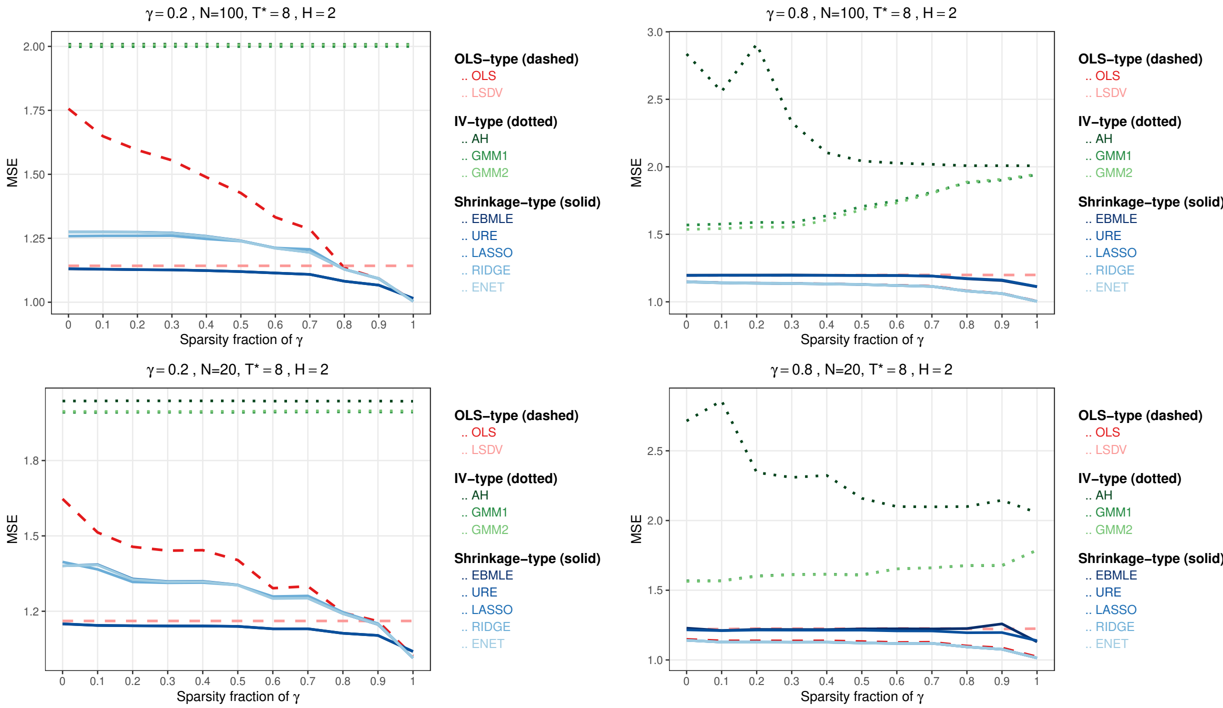

Figure 1: Out-of-sample mean squared error for $T = 10$

Figure 1 presents the out-of-sample predictive performance of the different estimators across varying degrees of fixed-effect sparsity, for both low persistence ($\gamma = 0.2$) and high persistence ($\gamma = 0.8$). Across all configurations, the estimators can be naturally grouped into three categories according to their predictive behavior: OLS-type estimators (OLS and FE), IV-type estimators (AH, GMM1, and GMM2), and regularized estimators (EBMLE, URE, LASSO, Ridge, and Elastic Net).

We begin with the OLS-type estimators, namely pooled OLS and the fixed-effects (LSDV) estimator. A notable result is the relatively stable performance of LSDV across the entire range of sparsity levels. Even when the true fixed effects become increasingly sparse, FE remains competitive in terms of out-of-sample MSE, particularly when sparsity is moderate or low. By contrast, the behavior of pooled OLS depends strongly on the degree of sparsity. When sparsity is low, OLS performs relatively poorly due to the presence of omitted-variable bias arising from unobserved heterogeneity. When sparsity becomes very high—above roughly 80%—the MSE of pooled OLS approaches that of the best-performing estimators.

The second group consists of the IV-type estimators: Anderson–Hsiao (AH), GMM1, and GMM2. Across all configurations in the figure, these estimators systematically display higher out-of-sample MSE than the OLS-type and regularized alternatives. While these methods are designed to mitigate the dynamic panel bias through instrumentation, their forecasting performance appears to be limited by their relatively high variance. As a result, the potential gains from bias reduction are more than offset by the associated increase in variance, leading to weaker predictive accuracy.

Finally, consider the regularized estimators— EBMLE, URE, LASSO, Ridge, and Elastic Net. Their performance differs somewhat across persistence regimes but remains consistently strong throughout the figure. In the low-persistence case ($\gamma = 0.2$), the fixed-effects estimator typically delivers the lowest MSE when sparsity is limited. Nevertheless, the regularized estimators emerge as a very close second-best alternative in this region, with EBMLE and URE often leading the group due to their substantial variance reduction. As sparsity increases, their performance improves steadily, eventually overtaking FE and becoming the best-performing estimators when sparsity becomes sufficiently high.

The advantage of regularized methods becomes even more pronounced in the high-persistence case ($\gamma = 0.8$). In this setting, LASSO, Ridge, and Elastic Net consistently deliver the lowest MSE across the entire range of sparsity levels, clearly dominating both OLS-type and IV-type estimators. EBMLE and URE also perform strongly—typically outperforming standard estimators—but do not match the MSE gains achieved by machine learning–based regularization. This pattern suggests that shrinkage is particularly effective in environments where dynamic persistence is high.

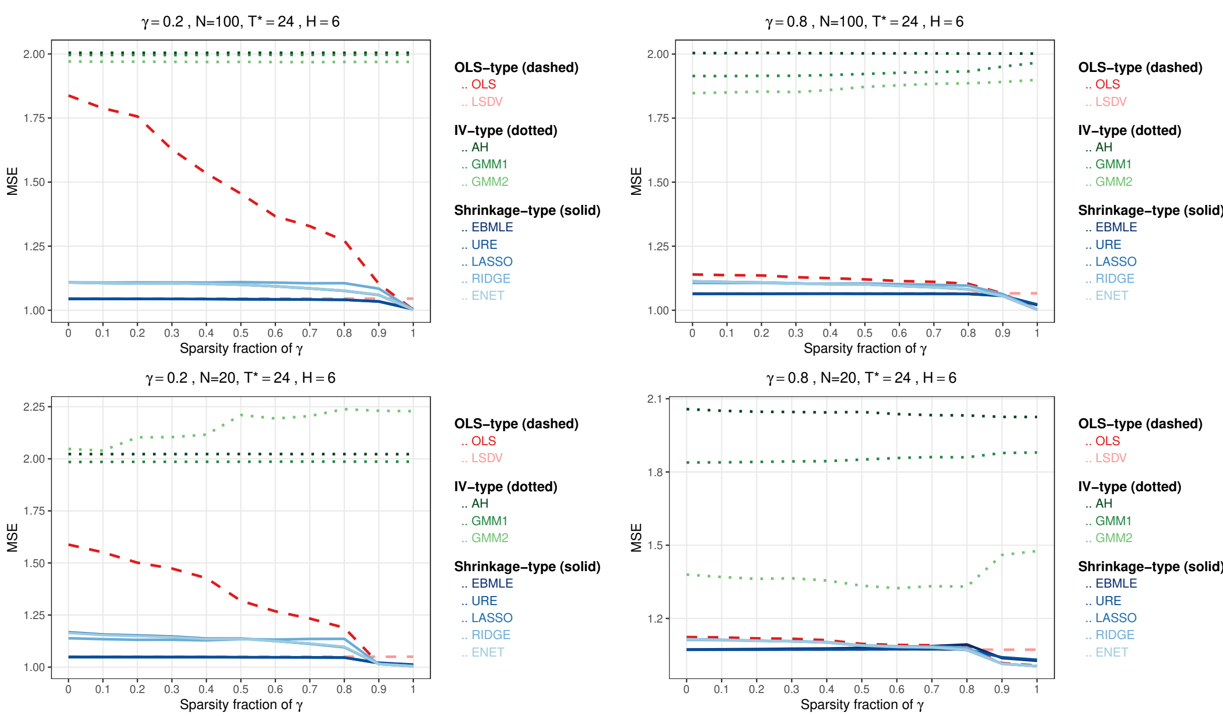

Figure 2: Out-of-sample mean squared error for $T = 30$

Figure 2 repeats the same exercise for a longer time dimension ($T^* = 24$), which allows us to assess how the results change when the dynamic panel bias becomes less severe. Overall, the qualitative patterns observed in Figure 1 remain largely unchanged, although the performance differences across estimators become smaller as the time dimension increases. In particular, the FE estimator becomes even more competitive, reflecting the well-known attenuation of the Nickell bias when T grows. As a consequence, the gap between FE and the regularized estimators narrows in the low-persistence case ($\gamma = 0.2$), where FE tends to deliver the lowest MSE for most sparsity levels, although EBMLE and URE remain very close competitors.

At the same time, the main conclusions regarding the other estimator groups remain broadly unchanged. IV-type estimators continue to display relatively large MSE across configurations, indicating that the variance cost associated with instrumentation remains substantial even when T increases. By contrast, the regularized estimators remain highly competitive. In particular, in the high-persistence case ($\gamma = 0.8$), the relative performance of regularized estimators depends on the degree of sparsity. LASSO, Ridge, and Elastic Net deliver the lowest MSE at high levels of sparsity, while EBMLE and URE tend to outperform them across the remaining configurations. Overall, these results reinforce the conclusion that shrinkage methods provide an effective forecasting strategy in highly persistent dynamic panel environments.

To further disentangle the performance within the class of shrinkage estimators, Figures A1 to A4 in the Appendix isolate the behavior of EBMLE, URE, LASSO, Ridge, and Elastic Net across different levels of sparsity and persistence. This representation highlights a clear pattern that is less visible in the full comparison: model-based shrinkage methods (EBMLE and URE) tend to dominate in low-to-moderate sparsity environments, while machine learning–based methods (LASSO, Ridge, and Elastic Net) deliver superior performance when sparsity is high.

To summarize, from the perspective of out-of-sample prediction, the results reveal a clear ranking across estimator groups. IV-based estimators are systematically dominated in terms of MSE, as the potential gains from bias reduction are consistently outweighed by their higher variance. By contrast, OLS-type and regularized estimators display substantially stronger predictive performance. The fixed-effects estimator performs particularly well when persistence is low and the time dimension increases, reflecting the attenuation of the Nickell bias as T grows.

At the same time, shrinkage-based estimators remain highly competitive across all configurations of the model. In the low-persistence case, EBMLE and URE typically emerge as the best-performing regularized alternatives—often close to FE—and become dominant as sparsity increases, while in the high-persistence case machine learning–based shrinkage methods (LASSO, Ridge, and Elastic Net) consistently deliver the lowest MSE.

A useful way to interpret these results is to view regularization as an adaptive estimator that lies between two extremes: pooled OLS, which imposes the restriction that all fixed effects are zero, and the fixed-effects estimator, which allows unrestricted heterogeneity across units. By shrinking individual effects toward zero, regularized estimators effectively select an intermediate degree of heterogeneity based on the data. This adaptive behavior helps explain their robust forecasting performance across environments with different levels of persistence and sparsity in the fixed effects.

5 An application to corporate capital structure

Capital structure decisions have long been modeled within a partial-adjustment framework in which firms adjust gradually toward target leverage ratios (e.g. Leary and Roberts, 2005; Flannery and Rangan, 2006; Lemmon et al., 2008). Empirically, this framework implies substantial persistence in firm leverage and motivates dynamic panel specifications that include lagged leverage as a predictor. Following this literature, we consider dynamic panel regressions of firm-level leverage,

$$lev_{it} = \alpha_i + \gamma lev_{i,t-1} + \beta_1 size_{it} + \beta_2 prof_{it} + \beta_3 tang_{it} + u_{it}, \tag{5}$$

lev_{it} denotes the book leverage ratio of firm i in year t, α_i captures firm-specific heterogeneity, and the remaining variables correspond to standard determinants of capital structure such as firm size, profitability, and asset tangibility (Rajan and Zingales, 1995).

Our empirical analysis uses firm-level data from the Compustat North America Fundamentals Annual database over the period 1991–2024. Following standard practice in the corporate finance literature, we exclude financial firms (SIC 6000–6999) and regulated utilities (SIC 4900–4999), whose capital structure decisions are shaped by industry-specific regulatory frameworks. The main variables are constructed using standard definitions: leverage is measured as the ratio of total debt (long-term debt plus debt in current liabilities) to total assets, profitability is defined as operating income over total assets, firm size as the logarithm of total assets, and asset tangibility as the ratio of property, plant and equipment to total assets.[2]

To ensure comparability across estimators that include firm-specific effects, we focus on a balanced panel of firms observed throughout the sample period. The final dataset contains 23,086 firm-year observations for 679 firms from 1991 to 2024 (34 years).

We evaluate the out-of-sample forecasting performance of the alternative estimators. The estimation period spans 1991–2010 ($T = 20$), while the remaining observations (2011–2024, $H = 14$) are reserved for out-of-sample evaluation. This split allows us to estimate the models using a moderately long time dimension while retaining a sufficiently long forecasting window to evaluate predictive accuracy.

For each estimator we generate one-step-ahead predictions of firm leverage and evaluate forecast performance using the same metrics employed in the Monte Carlo experiments: forecast bias, the variance of the predictions, and the root mean squared forecast error (RMSE). These metrics allow us to assess the bias–variance trade-off across estimators and to determine which methods deliver the best predictive performance in practice.

Table 1 reports the in-sample forecasting performance of the alternative estimators. Consistent with the Monte Carlo results, the fixed-effects estimator delivers the lowest in-sample RMSE, reflecting its ability to fully absorb persistent firm-specific heterogeneity. The URE estimator yields virtually identical performance, while EBMLE also performs very closely, achieving a noticeable reduction in variance relative to FE at the cost of a small increase in bias.

Pooled OLS and the machine learning shrinkage estimators (LASSO, Ridge, and Elastic Net) exhibit very similar in-sample performance, all lying slightly above FE in terms of RMSE. In contrast, the instrumental-variable estimators (AH and GMM) display substantially higher variance and correspondingly larger RMSE, despite exhibiting only modest bias.

Table 1: In-sample forecasting performance

Estimator	Bias	Variance	RMSE	MAE
POLS	0.000000	0.0267	0.0859	0.0516
FE	0.000000	0.0275	**0.0808**	**0.0491**
AH	-0.003275	0.0439	0.1136	0.0649
GMM1	-0.001098	0.0442	0.1144	0.0653
GMM2	-0.001221	0.0442	0.114243	0.0653
EBMLE	-0.000278	0.0256	0.0814	0.0503
URE	0.000002	0.0275	**0.0808**	**0.0491**
LASSO	0.000402	0.0269	0.0859	0.0513
Ridge	0.000407	0.0269	0.0859	0.0513
Elastic Net	0.000402	0.0269	0.0859	0.0513

Table 2 reports the corresponding out-of-sample results. Several patterns emerge. First, the machine learning shrinkage estimators perform very similarly to pooled OLS in terms of RMSE, while outperforming the alternative estimators in terms of out-of-sample MAE

Second, the fixed-effects estimator performs worse out-of-sample than in-sample, reflecting the well-known bias–variance trade-off when estimating a large number of firm-specific parameters. The URE estimator closely tracks the performance of FE, while EBMLE exhibits a modest improvement in RMSE driven by its lower variance, although it remains dominated by the regularized estimators.

Third, the IV-based estimators (AH and GMM) display substantially larger prediction variance and higher RMSE than the alternative estimators, despite exhibiting relatively small forecast bias.

Table 2: Out-of-sample forecasting performance

Estimator	Bias	Variance	RMSE	MAE
POLS	0.011695	0.0300	0.0749	0.0465
FE	0.014174	0.0283	0.0821	0.0547
AH	-0.000193	0.0460	0.0960	0.0565
GMM1	0.000961	0.0462	0.0968	0.0569
GMM2	0.000849	0.0462	0.0967	0.0568
EBMLE	0.013896	0.0267	0.0809	0.0539
URE	0.014176	0.0283	0.0821	0.0547
LASSO	0.011735	0.0302	**0.0749**	**0.0462**
Ridge	0.011752	0.0302	**0.0749**	**0.0462**
Elastic Net	0.011735	0.0302	**0.0749**	**0.0462**

Figure 3 complements these results by reporting relative differences in out-of-sample RMSE with respect to the fixed-effects estimator (as the natural benchmark). The figure makes clear that all machine learning shrinkage estimators achieve lower RMSE than FE, with Ridge delivering the largest improvement. EBMLE yields a modest gain relative to FE, while URE performs nearly identically. In contrast, IV-based estimators exhibit substantially worse predictive performance, with sizable increases in RMSE driven by their higher variance.

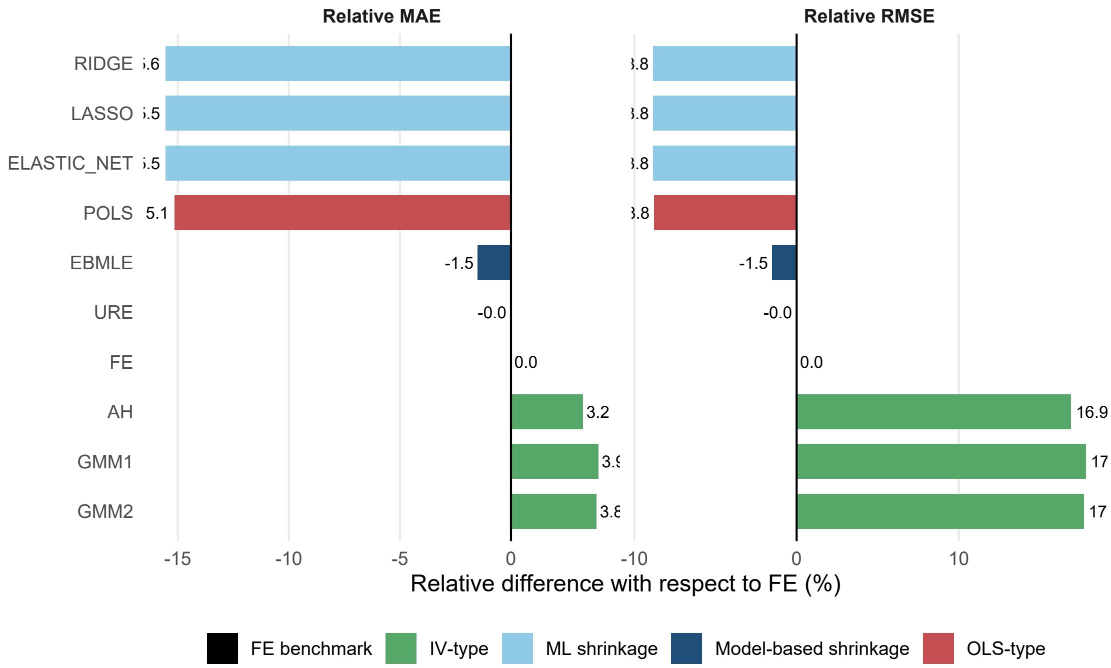

Figure 3: Relative Out-of-Sample Performance (Benchmark: Fixed Effects)

Overall, the empirical evidence closely mirrors the patterns documented in the Monte Carlo experiments. Estimators that rely on instrumental variables tend to produce highly variable forecasts, which deteriorates their predictive performance. In contrast, shrinkage-based approaches —including both model-based (EBMLE, URE) and machine learning methods— achieve competitive forecasting accuracy by reducing variance at the cost of introducing only limited bias. The gains relative to the fixed-effects estimator, summarized in Figure 3, highlight the benefits of regularization in high-dimensional dynamic panel settings. Among these approaches, machine learning estimators deliver the strongest out-of-sample performance, reflecting the advantages of more aggressive shrinkage.

6 Conclusion

This paper explores the performance of regularization methods in dynamic panel models when the goal is to predict the dependent variable rather than to obtain consistent estimates of structural parameters. The empirical evidence presented throughout the paper leads to two main conclusions.

First, conventional estimators exhibit a clear bias–variance trade-off. Naïve estimators such as pooled OLS may suffer from bias in dynamic panels, whereas bias-corrected alternatives such as Anderson–Hsiao or Arellano–Bond mitigate this problem at the cost of substantially increased variance. As a result, while consistent in principle, IV- and GMM-based methods often perform poorly from a forecasting perspective in samples of realistic dimensions, with prediction variance that remains large even for moderate time horizons. Regularized estimators—EBMLE, URE, LASSO, Ridge, and Elastic Net—offer an attractive alternative when the goal is forecasting. By construction, these methods introduce shrinkage that trades a certain amount of bias for a reduction in variance. The Monte Carlo results show that this strategy often leads to lower mean squared forecast errors. In particular, EBMLE and URE remain highly competitive across a wide range of configurations, while LASSO, Ridge, and Elastic Net deliver the largest gains in high-persistence and high-sparsity environments. In many configurations, shrinkage estimators match or outperform conventional estimators in terms of predictive accuracy.

Second, regularized estimators can be interpreted as adaptive procedures that bridge the gap between pooled and fully heterogeneous panel specifications. Pooled OLS implicitly assumes that all fixed effects are zero, whereas the fixed-effects estimator allows unrestricted heterogeneity across units. Regularization introduces an intermediate approach by shrinking individual effects toward zero, allowing the data to determine the effective degree of heterogeneity. From this perspective, shrinkage methods provide a flexible compromise between the two extremes, delivering robust predictive performance across environments where the true degree of heterogeneity is unknown.

The empirical application to firm-level leverage using Compustat illustrates the empirical relevance of these conclusions. In this setting, shrinkage estimators deliver forecasting performance comparable to the best-performing conventional estimators and slightly outperform them in terms of out-of-sample RMSE. At the same time, IV-based estimators produce substantially more variable predictions, resulting in poorer forecasting performance despite their desirable large-sample properties.

Overall, these results highlight the value of incorporating modern machine learning regularization tools into the econometric analysis of dynamic panels. Rather than replacing structural econometric frameworks, shrinkage methods complement them by improving predictive performance while preserving interpretability and economic structure. From this perspective, machine learning methods such as LASSO, Ridge, and Elastic Net can be viewed as practical forecasting tools for high-dimensional dynamic panel environments commonly encountered in financial econometrics.

More broadly, the evidence suggests that forecasting-oriented applications may benefit from moving beyond the traditional emphasis on asymptotic consistency toward approaches that explicitly balance bias and variance. In this setting, shrinkage-based methods provide a simple, transparent, and computationally tractable way to improve predictive accuracy in realistic samples.

References

Ahn, S. C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68*(1), 5–27.

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76*(375), 598–606.

Anderson, T. W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics 18*(1), 47–82.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies 58*(2), 277–297.

Babii, A., E. Ghysels, and J. Striaukas (2024). Econometrics of machine learning methods in economic forecasting. In *Handbook of Research Methods and Applications in Macroeconomic Forecasting*, pp. 246–278. Edward Elgar Publishing.

Baltagi, B. H. (2008). Forecasting with panel data. *Journal of Forecasting 27*(2), 153–173.

Baltagi, B. H. (2013). Panel data forecasting. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2B, pp. 995–1024. Elsevier.

Baltagi, B. H. (2021). *Econometric Analysis of Panel Data*. Springer Cham.

Baltagi, B. H., B. Fingleton, and A. Pirotte (2014). Estimating and forecasting with a dynamic spatial panel data model. *Oxford Bulletin of Economics and Statistics 76*(1), 112–138.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Brücker, H. and B. Siliverstovs (2006). On the estimation and forecasting of international migration: How relevant is heterogeneity across countries? *Empirical Economics 31*(3), 735–754.

Bun, M. J. and M. A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics 23*(2), 200–210.

Camehl, A. (2023). Penalized estimation of panel vector autoregressive models: A panel lasso approach. *International Journal of Forecasting 39*(3), 1185–1204.

Chernozhukov, V., I. Fernández-Val, C. Huang, and W. Wang (2024). Arellano-bond lasso estimator for dynamic linear panel models. *arXiv preprint arXiv:2402.00584*.

Flannery, M. J. and K. P. Rangan (2006). Partial adjustment toward target capital structures. *Journal of Financial Economics 79*(3), 469–506.

Frank, M. Z. and V. K. Goyal (2009). Capital structure decisions: which factors are reliably important? *Financial Management 38*(1), 1–37.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics 109*(1), 107–150.

Judson, R. A. and A. L. Owen (1999). Estimating dynamic panel data models: a guide for macroeconomists. *Economics Letters 65*(1), 9–15.

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics 68*(1), 53–78.

Kwon, S. (2026). Optimal shrinkage estimation of fixed effects in linear panel data models. *Econometrica 94*(2), 663–677.

Leary, M. T. and M. R. Roberts (2005). Do firms rebalance their capital structures? *The Journal of Finance 60*(6), 2575–2619.

Lemmon, M. L., M. R. Roberts, and J. F. Zender (2008). Back to the beginning: persistence and the cross-section of corporate capital structure. *The Journal of Finance 63*(4), 1575–1608.

Liu, L., H. R. Moon, and F. Schorfheide (2020). Forecasting with dynamic panel data models. *Econometrica 88*(1), 171–201.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, 1417–1426.

Pesaran, M. H., Y. Shin, and R. P. Smith (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association 94*(446), 621–634.

Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics 68*(1), 79–113.

Pick, A. and A. Timmermann (2024). Panel data forecasting. *Handbook of Research Methods and Applications in Macroeconomic Forecasting*, 68–89.

Rajan, R. G. and L. Zingales (1995). What do we know about capital structure? some evidence from international data. *The Journal of Finance 50*(5), 1421–1460.

Yang, B., W. Long, and Z. Cai (2024). Machine learning based panel data models. Technical Report 202402, University of Kansas, Department of Economics.

Zhang, L., Z. Zhu, X. Feng, and Y. He (2022). Shrinkage quantile regression for panel data with multiple structural breaks. *Canadian Journal of Statistics 50*(3), 820–851.

Notes

[1]When the training time dimension is very small (i.e., $T^* = 4$), the five-block temporal partition necessarily implies short validation windows. Although the rolling-origin procedure remains well defined, the validation loss is computed over a limited number of periods in this design.

[2]Because leverage ratios can become extremely large when book assets are close to zero, we cap leverage at one (e.g. Lemmon et al., 2008; Frank and Goyal, 2009).

A Additional results: shrinkage estimators

Figure A1: Out-of-sample MSE for shrinkage estimators ($T = 5$)

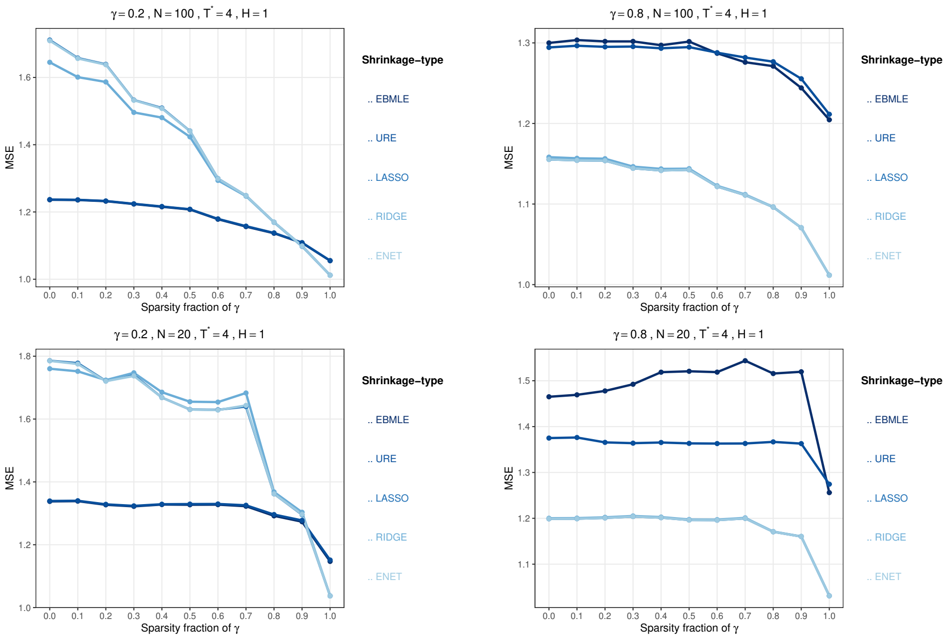

Figure A2: Out-of-sample MSE for shrinkage estimators ($T = 10$)

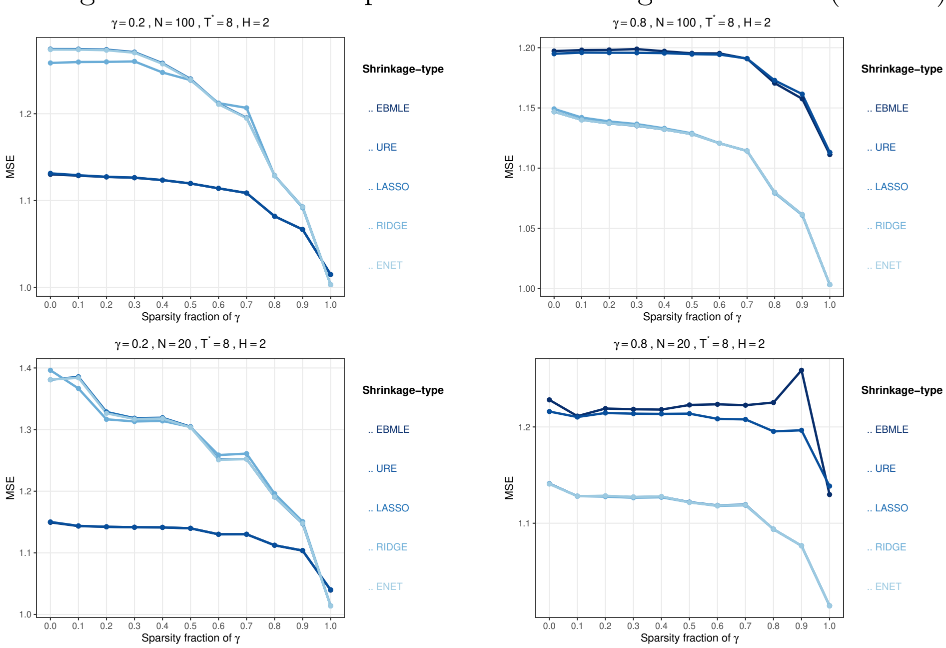

Figure A3: Out-of-sample MSE for shrinkage estimators ($T = 20$)

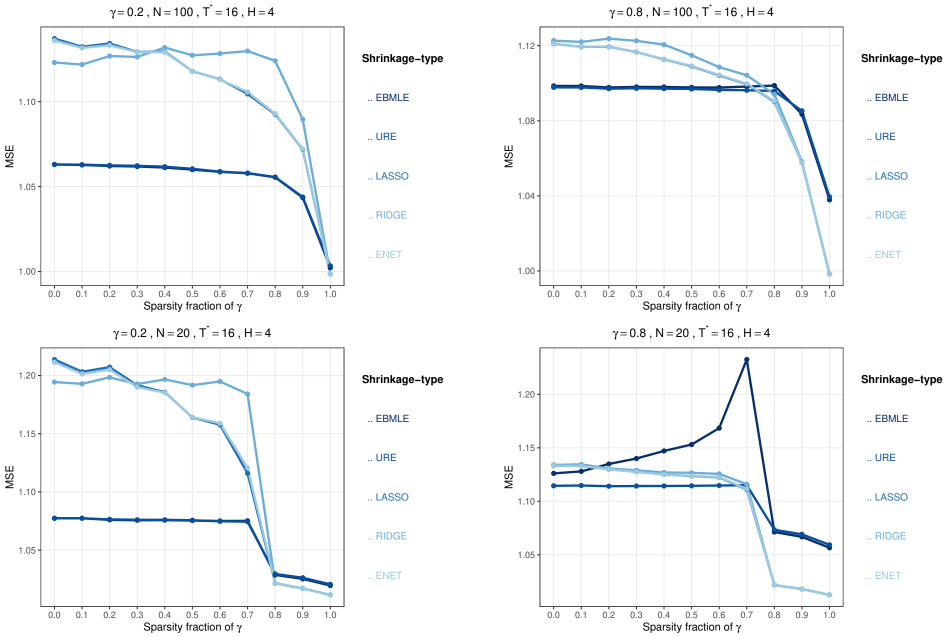

Figure A4: Out-of-sample MSE for shrinkage estimators ($T = 30$)

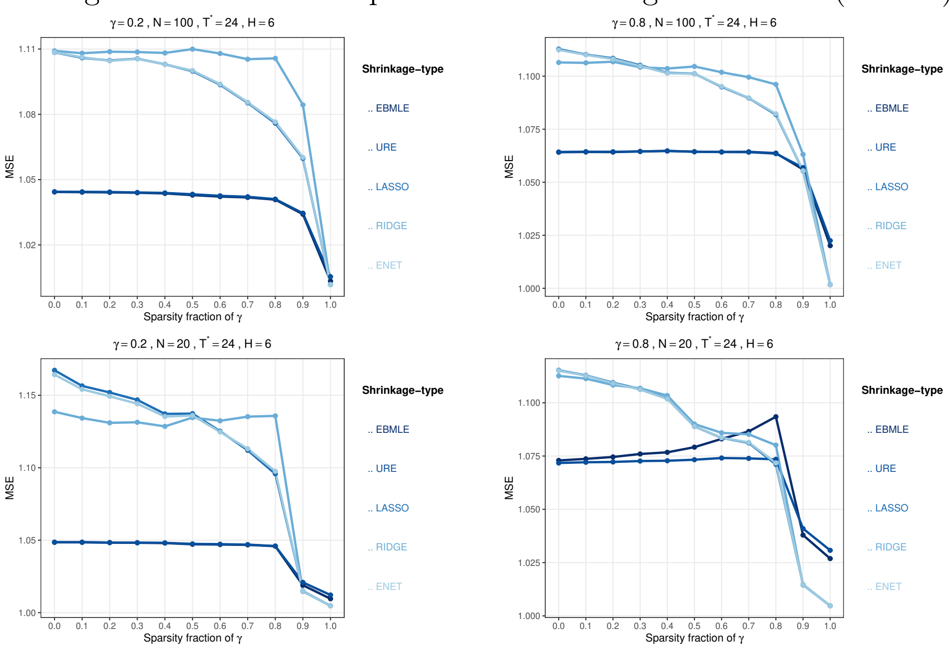

Table 3: In-sample γ estimation (with 0% sparsity in fixed effects)

N	T^*	γ	Measure	OLS	LSDV	AH	GMM1	GMM2	EBMLE	URE	LASSO	RIDGE	ENET
20	4	0.2	bias	-0.23014	0.10546	**-0.00410**	0.00952	0.01118	0.10546	0.10546	-0.20470	-0.20534	-0.20527
			var	**0.00302**	0.00591	0.01332	0.03875	0.04223	0.00591	0.00591	0.00499	0.00428	0.00485
			mse	0.05599	0.01703	**0.01334**	0.03884	0.04236	0.01703	0.01703	0.04689	0.04644	0.04699
		0.8	bias	**-0.18845**	0.45849	0.19831	0.65640	0.68808	0.45849	0.45849	-0.19377	-0.19449	-0.19380
			var	0.00047	0.01624	1.52782	2.33016	1.89010	0.01624	0.01624	0.00106	**0.00042**	0.00103
			mse	**0.03598**	0.22646	1.56714	2.76102	2.36356	0.22646	0.22646	0.03861	0.03825	0.03859
20	8	0.2	bias	-0.19985	0.04150	**-0.00179**	0.01646	0.01507	0.04150	0.04150	-0.12111	-0.13859	-0.12246
			var	**0.00134**	0.00195	0.00530	0.00367	0.00388	0.00195	0.00195	0.00464	0.00338	0.00453
			mse	0.04128	**0.00367**	0.00530	0.00394	0.00411	**0.00367**	**0.00367**	0.01931	0.02258	0.01952
		0.8	bias	-0.17154	0.21611	**-0.02053**	0.31591	0.32389	0.21611	0.21611	-0.15745	-0.16563	-0.15830
			var	**0.00017**	0.00501	0.53430	0.04759	0.05117	0.00501	0.00501	0.00441	0.00179	0.00424
			mse	0.02959	0.05172	0.53472	0.14739	0.15607	0.05172	0.05172	**0.02920**	0.02923	0.02930
20	16	0.2	bias	-0.25727	0.02013	**-0.00051**	0.01633	0.13817	0.02013	0.02013	-0.11081	-0.11653	-0.11174
			var	**0.00067**	0.00081	0.00260	0.00112	0.01685	0.00081	0.00081	0.00188	0.00162	0.00184
			mse	0.06685	**0.00122**	0.00260	0.00138	0.03594	**0.00122**	**0.00122**	0.01416	0.01520	0.01433
		0.8	bias	-0.18177	0.10176	**0.00778**	0.14395	0.23072	0.10176	0.10176	-0.16802	-0.16978	-0.16778
			var	**0.00004**	0.00159	0.28585	0.00542	0.02633	0.00159	0.00159	0.00171	0.00156	0.00173
			mse	0.03308	**0.01194**	0.28592	0.02614	0.07957	**0.01194**	**0.01194**	0.02994	0.03039	0.02988
20	24	0.2	bias	-0.18924	0.01223	**-0.00168**	0.01383	0.36692	0.01223	0.01223	-0.08295	-0.07933	-0.08289
			var	**0.00051**	0.00056	0.00189	0.00066	0.03925	0.00056	0.00056	0.00099	0.00083	0.00098
			mse	0.03632	**0.00071**	0.00190	0.00085	0.17388	**0.00071**	**0.00071**	0.00787	0.00713	0.00785
		0.8	bias	-0.17364	0.06588	**-0.00490**	0.09155	0.52405	0.06588	0.06588	-0.16059	-0.15698	-0.16040
			var	**0.00003**	0.00097	0.02707	0.00191	0.06035	0.00097	0.00097	0.00097	0.00115	0.00099
			mse	0.03018	**0.00531**	0.02710	0.01029	0.33498	**0.00531**	**0.00531**	0.02676	0.02579	0.02671
100	4	0.2	bias	-0.22330	0.09432	**0.00045**	0.00209	0.00163	0.09432	0.09432	-0.22131	-0.20936	-0.22101
			var	**0.00053**	0.00097	0.00263	0.00705	0.00707	0.00097	0.00097	0.00062	0.00075	0.00062
			mse	0.05039	0.00987	**0.00263**	0.00705	0.00707	0.00987	0.00987	0.04960	0.04458	0.04946
		0.8	bias	-0.18513	0.42439	**0.15326**	0.65436	0.68798	0.42439	0.42439	-0.18511	-0.18432	-0.18511
			var	**0.00009**	0.00273	1.85062	2.38191	3.37873	0.00273	0.00273	**0.00009**	**0.00009**	**0.00009**
			mse	0.03436	0.18284	1.87411	2.81010	3.85204	0.18284	0.18284	0.03435	**0.03406**	0.03435
100	8	0.2	bias	-0.24151	0.03949	**-0.00026**	0.00444	0.00195	0.03949	0.03949	-0.11422	-0.12120	-0.11479
			var	**0.00025**	0.00036	0.00114	0.00114	0.00137	0.00036	0.00036	0.00125	0.00109	0.00126
			mse	0.05858	0.00192	**0.00114**	0.00116	0.00137	0.00192	0.00192	0.01430	0.01578	0.01444
		0.8	bias	-0.18364	0.19392	**0.01249**	0.30372	0.36849	0.19392	0.19392	-0.17983	-0.18131	-0.18005
			var	**0.00003**	0.00094	0.65587	0.04346	0.07496	0.00094	0.00094	0.00099	0.00031	0.00094
			mse	0.03375	0.03854	0.65602	0.13571	0.21075	0.03854	0.03854	0.03333	**0.03318**	0.03336
100	16	0.2	bias	-0.24757	0.01923	**-0.00041**	0.00426	0.00377	0.01923	0.01923	-0.07308	-0.07964	-0.07346
			var	**0.00013**	0.00016	0.00056	0.00031	0.00032	0.00016	0.00016	0.00040	0.00038	0.00040
			mse	0.06142	0.00053	0.00056	**0.00032**	0.00033	0.00053	0.00053	0.00575	0.00672	0.00580
		0.8	bias	-0.18280	0.09582	**-0.00076**	0.07701	0.07712	0.09582	0.09582	-0.15587	-0.15829	-0.15595
			var	**0.00001**	0.00030	0.02231	0.00272	0.00283	0.00030	0.00030	0.00245	0.00242	0.00246
			mse	0.03342	0.00948	0.02231	**0.00865**	0.00878	0.00948	0.00948	0.02675	0.02748	0.02678
100	24	0.2	bias	-0.24582	0.01268	**0.00015**	0.00394	0.01847	0.01268	0.01268	-0.06282	-0.07363	-0.06350
			var	**0.00009**	0.00010	0.00034	0.00017	0.00056	0.00010	0.00010	0.00026	0.00026	0.00026
			mse	0.06052	0.00026	0.00034	**0.00018**	0.00090	0.00026	0.00026	0.00420	0.00568	0.00429
		0.8	bias	-0.17763	0.06034	**0.00297**	0.04619	0.08364	0.06034	0.06034	-0.13951	-0.13325	-0.13911
			var	**0.00000**	0.00017	0.00587	0.00082	0.00227	0.00017	0.00017	0.00178	0.00202	0.00180
			mse	0.03156	0.00381	0.00588	**0.00295**	0.00927	0.00381	0.00381	0.02124	0.01977	0.02115

Table 4: In-sample prediction evaluation (with 0% sparsity in fixed effects)

N	T^*	γ	Measure	OLS	LSDV	AH	GMM1	GMM2	EBMLE	URE	LASSO	RIDGE	ENET
20	4	0.2	bias	**0.00000**	**0.00000**	0.00074	-0.00020	0.00024	0.00018	-0.00016	0.35427	0.34363	0.35477
			var	3.93569	4.84627	6.00296	6.10308	6.10935	4.29981	4.33076	3.83774	**3.82239**	3.83438
			mse	1.59449	**0.69529**	1.95608	2.06060	2.07544	0.75289	0.74669	1.60256	1.57034	1.60345
		0.8	bias	**0.00000**	**0.00000**	-0.00152	-0.01984	-0.01627	0.05090	-0.02284	0.09135	0.09021	0.09135
			var	21.22721	21.76828	25.15564	24.84861	24.06272	**18.89464**	20.71272	21.49381	21.48963	21.49378
			mse	1.10996	**0.57565**	3.54434	4.34717	3.60709	0.71140	0.59871	1.12834	1.11977	1.12835
20	8	0.2	bias	**0.00000**	**0.00000**	0.00062	0.00121	0.00152	-0.00193	0.00165	-0.06670	-0.06499	-0.06696
			var	**5.23081**	5.95640	7.25112	7.18057	7.17473	5.55304	5.60121	5.40936	5.34301	5.40380
			mse	1.57956	**0.85851**	1.99181	1.96765	1.97163	0.87804	0.87394	1.20105	1.23827	1.20270
		0.8	bias	**0.00000**	**0.00000**	0.00045	-0.00384	-0.00372	-0.01855	0.00931	-0.01759	-0.01885	-0.01758
			var	23.60534	23.92640	27.11191	25.30547	25.28531	**22.03288**	23.27492	23.64241	23.63800	23.64177
			mse	1.11980	**0.80075**	2.66991	1.54085	1.53634	0.84931	0.80737	1.10640	1.11027	1.10756
20	16	0.2	bias	**0.00000**	**0.00000**	0.00063	0.00027	-0.00144	-0.00127	0.00101	-0.05212	-0.07766	-0.05411
			var	**4.95775**	5.79314	7.13671	7.06574	6.47277	5.61873	5.66193	5.26698	5.21938	5.25987
			mse	1.76228	**0.92951**	1.99234	1.96272	1.87556	0.93302	0.93181	1.14437	1.12762	1.14247
		0.8	bias	**0.00000**	**0.00000**	0.00058	-0.00316	-0.00507	-0.01908	0.00998	-0.04102	-0.04139	-0.04082
			var	31.13285	31.35332	34.17915	33.35278	33.07988	**30.15193**	30.96697	31.23187	31.23025	31.23154
			mse	1.12884	**0.90905**	2.30628	1.73858	1.63155	0.92453	0.91099	1.11065	1.11066	1.11028
20	24	0.2	bias	**0.00000**	**0.00000**	0.00005	0.00054	0.00040	-0.00033	0.00028	0.00850	-0.04977	0.00425
			var	**4.35517**	5.02708	6.38742	6.32882	5.15377	4.93750	4.94880	4.60004	4.57648	4.59668
			mse	1.62338	**0.95287**	1.99411	1.96619	2.01067	0.95480	0.95445	1.11081	1.08888	1.10832
		0.8	bias	**0.00000**	**0.00000**	0.00006	0.00018	-0.00277	-0.00906	0.00669	-0.02587	-0.02880	-0.02598
			var	21.72478	21.91607	24.50708	24.14764	23.05167	**21.38482**	21.63783	21.76662	21.76967	21.76679
			mse	1.13360	**0.94272**	2.02456	1.82388	1.36878	0.94700	0.94405	1.11416	1.10709	1.11389
100	4	0.2	bias	**0.00000**	**0.00000**	0.00098	0.00097	0.00098	0.00003	-0.00003	0.05643	0.04955	0.05669
			var	**4.76068**	5.69861	6.81883	6.84059	6.84270	5.06636	5.08530	4.78991	4.79318	4.79024
			mse	1.65755	**0.72196**	1.98726	2.00817	2.00951	0.76436	0.76178	1.62928	1.52200	1.62611
		0.8	bias	**0.00000**	**0.00000**	0.00156	0.00948	0.01029	-0.00326	0.00180	-0.00063	-0.00034	-0.00063
			var	23.70862	24.25301	28.06828	26.98208	28.08139	**23.08042**	23.86379	23.70743	23.70810	23.70743
			mse	1.14782	**0.60479**	4.00693	4.14373	5.31138	0.62092	0.60739	1.14994	1.13213	1.14994
100	8	0.2	bias	**0.00000**	**0.00000**	-0.00006	0.00010	0.00039	-0.00022	0.00019	0.00830	0.01948	0.00855
			var	**6.25410**	7.16882	8.44104	8.41799	8.40816	6.81597	6.85440	6.56601	6.50122	6.56205
			mse	1.78151	**0.86794**	2.00122	1.99417	2.00090	0.87939	0.87776	1.12143	1.11827	1.12157
		0.8	bias	**0.00000**	**0.00000**	0.00019	0.00849	0.01074	0.00016	0.00019	0.00049	0.00072	0.00054
			var	30.52169	30.85969	34.06460	32.07824	31.88088	**30.00376**	30.45513	30.52567	30.52384	30.52552
			mse	1.15232	**0.81475**	2.81532	1.56540	1.53289	0.82157	0.81669	1.14791	1.13888	1.14820
100	16	0.2	bias	**0.00000**	**0.00000**	0.00067	0.00064	0.00068	0.00014	-0.00012	0.01469	0.02313	0.01497
			var	**5.07267**	5.96998	7.30968	7.29040	7.28849	5.82979	5.85096	5.50444	5.46091	5.50161
			mse	1.83060	**0.93385**	1.99683	1.98803	1.98910	0.93661	0.93606	1.05644	1.04764	1.05557
		0.8	bias	**0.00000**	**0.00000**	0.00054	0.00100	0.00103	0.00160	-0.00170	0.01066	0.01104	0.01063
			var	31.17560	31.41000	33.95469	33.64121	33.64125	**30.96810**	31.13107	31.20561	31.20226	31.20548
			mse	1.14862	**0.91438**	2.02151	1.85257	1.85247	0.91661	0.91527	1.10934	1.10854	1.10948
100	24	0.2	bias	**0.00000**	**0.00000**	-0.00011	-0.00006	-0.00005	0.00004	-0.00004	0.03128	0.03084	0.03141
			var	**5.49352**	6.34160	7.69758	7.68433	7.58973	6.25076	6.25793	5.96971	5.90505	5.96568
			mse	1.80368	**0.95596**	1.99780	1.99071	1.96492	0.95712	0.95697	1.03786	1.04373	1.03813
		0.8	bias	**0.00000**	**0.00000**	-0.00004	0.00055	0.00099	0.00053	-0.00053	0.01372	0.01274	0.01365
			var	32.42800	32.62482	35.16592	35.00639	34.88218	**32.31026**	32.38871	32.47794	32.47543	32.47797
			mse	1.14339	**0.94665**	1.99764	1.90951	1.84260	0.94773	0.94725	1.08652	1.07590	1.08600

Table 5: Out-of-sample forecast evaluation (with 0% sparsity in fixed effects)

N	T^*	γ	Measure	OLS	LSDV	AH	GMM1	GMM2	EBMLE	URE	LASSO	RIDGE	ENET
20	4	0.2	bias	-0.14636	0.10326	-0.01471	-0.00333	**-0.00031**	0.10344	0.10310	0.24502	0.23907	0.24556
			var	4.50802	4.73100	6.09743	6.13648	6.14215	**4.27352**	4.29914	4.41604	4.39270	4.41163
			mse	1.83672	1.38875	2.11013	2.24897	2.26598	**1.33787**	1.33920	1.78566	1.76013	1.78516
		0.8	bias	**-0.02078**	0.14005	0.02952	0.19177	0.19562	0.19095	0.11721	0.07445	0.07240	0.07444
			var	23.02153	22.85561	26.77380	26.71168	25.86869	**19.85152**	21.74082	23.30698	23.32600	23.30701
			mse	1.20770	1.36772	3.81594	4.73227	4.10348	1.46505	1.37501	1.19899	1.20016	**1.19896**
20	8	0.2	bias	-0.10761	0.01072	-0.00540	-0.00507	**-0.00492**	0.00878	0.01236	-0.14110	-0.14672	-0.14201
			var	4.81469	5.00873	6.33592	6.26519	6.25667	**4.75736**	4.78674	4.93118	4.88554	4.92803
			mse	1.64719	1.16135	2.03689	1.99196	1.99554	**1.14950**	1.15016	1.38084	1.39631	1.38105
		0.8	bias	-0.03827	0.03985	**-0.00394**	0.01987	0.02065	0.02129	0.04915	-0.05410	-0.05603	-0.05419
			var	23.35429	23.51792	26.53763	25.20316	25.19982	**21.67262**	22.88691	23.39517	23.40367	23.39599
			mse	1.14790	1.22173	2.71422	1.56762	1.56336	1.22812	1.21599	1.14123	1.14090	**1.14072**
20	16	0.2	bias	-0.01265	-0.00271	-0.00206	-0.00153	**0.00090**	-0.00398	-0.00170	-0.06010	-0.08566	-0.06211
			var	**5.25198**	6.00071	7.36174	7.29955	6.76681	5.82568	5.86892	5.56141	5.49350	5.55426
			mse	1.68540	1.07814	2.01892	1.98079	1.88903	**1.07728**	1.07737	1.21376	1.19440	1.21142
		0.8	bias	0.00392	-0.00669	**-0.00092**	-0.00254	-0.00282	-0.02578	0.00328	-0.03752	-0.03769	-0.03746
			var	34.99648	34.12639	37.40505	36.68015	36.46682	**32.87512**	33.72203	35.06650	35.07745	35.06558
			mse	1.14018	**1.11304**	2.33906	1.75368	1.64774	1.12610	1.11449	1.13333	1.13408	1.13328
20	24	0.2	bias	0.00536	-0.00236	**0.00004**	-0.00064	0.01467	-0.00269	-0.00208	0.00639	-0.04940	0.00231
			var	**4.35827**	5.44648	6.81667	6.75366	5.52739	5.33770	5.35144	4.90868	4.86478	4.90096
			mse	1.58812	1.04996	2.02271	1.98480	2.04740	**1.04854**	1.04863	1.16727	1.13863	1.16429
		0.8	bias	0.01414	-0.00619	**-0.00038**	0.00217	0.02112	-0.01525	0.00050	-0.01358	-0.01669	-0.01371
			var	24.71107	24.35670	27.20562	26.90026	25.91370	**23.79127**	24.06154	24.76088	24.76843	24.76012
			mse	1.12375	1.07225	2.05725	1.83871	1.37978	1.07293	**1.07174**	1.11524	1.11262	1.11505
100	4	0.2	bias	-0.06093	0.02707	0.00481	0.00565	0.00547	0.02710	0.02704	-0.00297	-0.00553	**-0.00257**
			var	**4.29685**	5.05140	6.27979	6.29561	6.29625	4.57260	4.58659	4.34549	4.39510	4.34797
			mse	1.73718	1.29379	2.02127	2.04213	2.04328	**1.23625**	1.23688	1.71124	1.64511	1.70925
		0.8	bias	-0.01115	0.02423	**-0.00351**	0.02139	0.02271	0.02098	0.02604	-0.01189	-0.01132	-0.01189
			var	25.22887	24.70994	29.36094	28.93323	29.98411	**23.52725**	24.31718	25.22721	25.26421	25.22721
			mse	1.15879	1.29510	4.05634	4.22783	5.28443	1.29996	1.29436	**1.15541**	1.15819	**1.15541**
100	8	0.2	bias	-0.02713	0.00162	-0.00187	-0.00233	-0.00239	**0.00141**	0.00181	-0.00483	0.00539	-0.00464
			var	**5.42103**	6.13683	7.51224	7.48854	7.48891	5.81060	5.84615	5.96727	5.82125	5.96126
			mse	1.75622	1.14235	2.00828	1.99873	2.00557	**1.13024**	1.13140	1.27442	1.25846	1.27383
		0.8	bias	-0.03140	0.02721	-0.00907	**-0.00197**	-0.00209	0.02737	0.02741	-0.03021	-0.03032	-0.03028
			var	34.00314	32.87135	36.79401	35.14856	35.03291	**31.98871**	32.45410	33.99422	34.03593	33.99506
			mse	1.14890	1.19527	2.83590	1.56897	1.53633	1.19723	1.19499	1.14687	1.14922	**1.14686**
100	16	0.2	bias	-0.04207	0.00332	-0.00177	**-0.00172**	-0.00180	0.00347	0.00320	0.00178	0.00890	0.00198
			var	**6.29628**	7.52833	8.91502	8.88973	8.88805	7.35561	7.38179	7.05002	6.96257	7.04566
			mse	1.79391	1.06545	1.99734	1.98666	1.98775	**1.06298**	1.06305	1.13709	1.12302	1.13596
		0.8	bias	-0.02256	0.01261	-0.00172	**0.00000**	-0.00004	0.01421	0.01091	-0.00911	-0.00908	-0.00907
			var	35.13678	34.55006	37.49301	37.18003	37.17909	**34.08755**	34.25832	35.14391	35.16341	35.14357
			mse	1.13481	1.09842	2.02647	1.85137	1.85140	1.09859	**1.09779**	1.12105	1.12271	1.12105
100	24	0.2	bias	0.02369	-0.00178	0.00031	**0.00024**	0.00147	-0.00174	-0.00181	0.03677	0.03737	0.03696
			var	**5.34234**	6.22970	7.61107	7.59804	7.49794	6.14119	6.14815	5.87893	5.83382	5.87557
			mse	1.83789	1.04552	2.00434	1.99609	1.97061	**1.04434**	1.04445	1.10846	1.10918	1.10853
		0.8	bias	0.00352	-0.00182	0.00010	**-0.00001**	0.00054	-0.00129	-0.00234	0.01603	0.01499	0.01598
			var	32.64058	32.67004	35.40333	35.26361	35.14569	**32.35707**	32.43499	32.75428	32.75600	32.75407
			mse	1.13976	1.06519	2.00409	1.91439	1.84743	**1.06415**	1.06426	1.11276	1.10648	1.11239