Double Machine Learning with Interactive Fixed Effects

International Panel Data Conference, University of Exeter

Binzhi Chen Annalivia Polselli Paul Clarke

Institute for Social and Economic Research, University of Essex

07/07/2026

Motivation: Two Challenges for Panel Data

- **Challenge 1**: Causal effects under high-dimensional controls setting $\rightarrow$ solved by Double Machine Learning (DML) (Chernozhukov et al., 2018; Semenova et al., 2023; Ahrens et al., 2025; Clarke and Polselli, 2025; Avila Marquez, 2025; Baiardi et al., 2026)
  - Modern applications involve **many control variables** $(p \geqslant NT)$ entering through **unknown, nonlinear** functions
- **Challenge 2**: Latent common factor shocks $\rightarrow$ solved by modelling interactive fixed effects (IFE) (Pesaran, 2006; Bai, 2009; Chudik and Pesaran, 2015; Moon and Weidner, 2017, 2018; Miao et al., 2020; Juodis et al., 2021; Karavias et al., 2023; Ditzen et al., 2025; Ditzen and Karavias, 2025; Chen et al., 2025)
  - In panel data, N units share **unobserved time-varying factors** f_t, including global recessions, commodity cycles, pandemic effects. But each reacts with a unit-specific loading λ_i.

Building Block 1: Double Machine Learning (DML)

Partially Linear Model — estimate θ_0, partial out controls X (Robinson, 1988)

$$Y_{it} = \theta_0 V_{it} + \underbrace{l_0(\mathbf{X}_{it})}_{\text{nonlinear nuisance}} + U_{it}, \tag{1}$$

$$V_{it} = D_{it} - \underbrace{m_0(\mathbf{X}_{it})}_{\text{nonlinear nuisance}} \tag{2}$$

Where Y_{it} is the outcome variable, D_{it} is the treatment variable and θ_0 is treatment effect of interest.

DML procedure (Chernozhukov et al., 2018):

1. Use ML to estimate $\hat{l}(\mathbf{X}_{it})$ and $\hat{m}(\mathbf{X}_{it})$ on held-out data (cross-fitting)
2. Estimate θ_0 from moment condition of Neyman-orthogonal score function

Key Result

Works with **any** sufficiently accurate ML learner. $\sqrt{N}$-consistent estimator for θ_0 under weak assumptions.

Building Block 2: Interactive Fixed Effects (IFE)

Standard FE (additive): $\alpha_i + \gamma_t$

- Removes time-invariant differences
- Units react identically to time shocks

Interactive FE (Bai, 2009):

$$Y_{it} = D_{it}'\theta + \lambda_i' f_t + \epsilon_{it}$$

$\lambda_i' f_t$ (multiplicative)

- f_t: latent common factors (recession, pandemic); λ_i: unit-specific loadings
- Nests additive FE; **cannot be removed by demeaning**

$\lambda_i' f_t$ captures this heterogeneous exposure to common shocks, confounds $\hat{\theta}$ if ignored.
Ignoring IFE biases the treatment effect: the latent factor variation is correlated with D_{it}

The Gap We Fill

	High-dim X	Non-linear Panel	High-dim IFE	Inference
OLS	$\times$	$\times$	$\times$	$\checkmark$
IFE (Bai, 2009)	$\times$	$\times$	$\checkmark$	$\checkmark$
DML-FE (Clarke and Polselli, 2025)	$\checkmark$	$\checkmark$	$\times$	$\checkmark$
CCE-Lasso (Rücker et al., 2025)	$\checkmark$	$\times$	$\checkmark$	$\checkmark$
DML-IFE (this paper)	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$

This paper...

- Projection-based CCE factor removal + Neyman-orthogonal DML
- Any ML learner; p can exceed NT
- $\sqrt{NT}$-consistent; asymptotically normal

The Model: Partially Linear Panel with IFE

Outcome and treatment equations:

$$Y_{it} = \theta_0 V_{it} + \underbrace{l_0(\mathbf{X}_{it})}_{\text{nonlinear nuisance}} + \underbrace{\boldsymbol{\lambda}_i^\top \boldsymbol{f}_t}_{\text{low-dim IFE}} + U_{it} \tag{3}$$

$$V_{it} = D_{it} - m_0(\mathbf{X}_{it}) - \underbrace{\boldsymbol{\phi}_i^\top \boldsymbol{f}_t}_{\text{low-dim IFE}} \tag{4}$$

Covariate high-dimensional factor structure:

$$\mathbf{X}_{it} = \underbrace{\boldsymbol{\Gamma}_i^\top \boldsymbol{f}_t}_{\text{high-dim IFE}} + \mathbf{E}_{it} \tag{5}$$

Notation:

- θ_0: treatment effect of interest
- $\boldsymbol{f}_t \in \mathbb{R}^r$: latent common factors
- $\boldsymbol{\lambda}_i, \boldsymbol{\phi}_i, \boldsymbol{\Gamma}_i$: unit-specific loadings
- p can exceed NT (high-dimensional)

Factor Removal: The CCE-based Projection

Based on (Pesaran, 2006; Rücker et al., 2025): cross-sectional averages $\bar{X}_t = N^{-1}\sum_i X_{it}$ approximately span the factor space:

$$\bar{X}_t \approx \bar{\Gamma}' f_t \quad \Rightarrow \quad \text{project onto } \operatorname{col}(\bar{X}) \text{ to remove } f_t$$

Projection matrix construction:

1. Threshold eigenvalues of $T^{-1}\bar{X}'\bar{X}$ to estimate $\hat{r}$
2. High-dim case $(p > T)$: extract top $\hat{r}$ eigenvectors $\hat{\phi}$; set $\hat{W} = \bar{X}\hat{\phi}$
3. Annihilator: $\hat{\Pi} = I_T - \hat{W}(\hat{W}'\hat{W})^{-1}\hat{W}'$

Advantage: $\Pi F = 0$ $(\Pi = I - F\left(F^\top F\right)^{-1} F^\top)$: projecting Y and D through $\hat{\Pi}$ removes the interactive FE. No non-convex optimisation; no iterative PCA.

The Panel DML-IFE Algorithm

- **Step 1: Rücker et al. (2025) Projection**
  Compute $\hat{\Pi}$ from $I_T - \hat{W}(\hat{W}'\hat{W})^{-1}\hat{W}'$ remove $\boldsymbol{\lambda}_i^\top \boldsymbol{f}_t$, $\boldsymbol{\phi}_i^\top \boldsymbol{f}_t$ and $\boldsymbol{\Gamma}_i^\top \boldsymbol{f}_t$

- **Step 2: Transform Data**
  $\tilde{\boldsymbol{Y}} = \hat{\Pi}\boldsymbol{Y}$, $\tilde{\boldsymbol{D}} = \hat{\Pi}\boldsymbol{D}$ and $\tilde{\mathbf{X}} = \hat{\Pi}\mathbf{X}$ (IFE-free data) as:

$$\tilde{\boldsymbol{Y}}_i = \tilde{\boldsymbol{V}}\theta_0 + \widehat{\Pi}\boldsymbol{l}(\tilde{\mathbf{X}}_i) + \tilde{\boldsymbol{U}}_i, \tag{6}$$

$$\tilde{\boldsymbol{V}}_i = \tilde{\boldsymbol{D}}_i - \widehat{\Pi}\boldsymbol{m}(\tilde{\mathbf{X}}_i). \tag{7}$$

- **Step 3: Cross-fit DML** Apply ML learners $\Rightarrow$ Neyman score $\Rightarrow$ $\hat{\theta}$
  ▶ Neyman Score Function Derivation

- **Block cross-fitting:** K folds, nuisance estimated on complement set
- **Any ML algorithm:** Lasso, gradient boosting, neural networks

Assumption

Assumption (A.1 Panel process)

Let ξ_{it} *represent the entire set of omitted confounders for individual* i *at time* t. *Then* $\{Y_{it}, D_{it}, \mathbf{X}_{it}, \xi_{it} : t \in \mathbb{N}\}$ *satisfies the conditions below.*

1. $\mathbf{X}_{it}, \xi_{it} \perp\!\!\!\perp \{Y_{is}, D_{is} : s < t\} \mid \{\mathbf{X}_{is}, \xi_{is} : s < t\}$,
2. $Y_{it}, D_{it} \perp\!\!\!\perp \{Y_{is}, D_{is}, \mathbf{X}_{is}, \xi_{is} : s < t\} \mid X_{it}, \xi_{it}$,

where $\perp\!\!\!\perp$ *indicates conditional (stochastic) independence.*

Assumption (A.2 Additive separability)

1. $Y_{it} = \theta_0 V_{it} + \widetilde{l}_0\left(\mathbf{X}_{it}, \xi_{it}\right) + U_{it}$, *where* $\mathbb{E}[U_{it} \mid D_{it}, \mathbf{X}_{it}, \xi_{it}] = 0$, $\widetilde{l}_0(\mathbf{X}_{it}, \xi_{it}) = l_0(\mathbf{X}_{it}) + \alpha_{it}$ *and fixed effect* $\alpha_{it} = \alpha(\xi_{it})$,
2. $V_{it} = D_{it} - \widetilde{m}_0\left(\mathbf{X}_{it}, \xi_{it}\right)$, *where* $\mathbb{E}[V_{it} \mid \mathbf{X}_{it}, \xi_{it}] = 0$, $\widetilde{m}_0(\mathbf{X}_{it}, \xi_{it}) = m_0(\mathbf{X}_{it}) + \gamma_{it}$ *and fixed effect* $\gamma_{it} = \gamma(\xi_{it})$.

Inference

Lemma (Projection Matrix Convergence)

let $\widehat{\Pi}$ *be the feasible high-dimensional projection matrix,* $\widehat{\Pi} = \mathbf{I}_T - \widehat{\mathbf{W}}(\widehat{\mathbf{W}}'\widehat{\mathbf{W}})^{-1}\widehat{\mathbf{W}}'$ *with* $\widehat{\mathbf{W}} = \overline{\mathbf{X}}\widehat{\mathbf{\Phi}}$*, where* $\widehat{\mathbf{\Phi}} = (\widehat{\varphi}_1, \ldots, \widehat{\varphi}_r)$ *are the top-*r *eigenvectors of* $\widehat{\mathbf{\Sigma}} = T^{-1}\overline{\mathbf{X}}'\overline{\mathbf{X}}$ *and* $\overline{\mathbf{X}} = N^{-1}\sum_i \mathbf{X}_i$*.*

$$\widehat{\Pi} = \Pi - \widehat{R}$$

where

$$\|\widehat{R}\| = O_p\left(\frac{\sigma_E}{c_\Gamma}\sqrt{(T+p)\log(T+p)/(NTp)}\right)$$

.

Inference

Theorem (Asymptotic normality of the panel DML-IFE estimator)

Let $\widehat{\theta}$ denote the K-fold panel DML-IFE estimator with $K \geq 2$. As $(N, T) \to \infty$,

$$\sqrt{NT}\,(\widehat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

Where

$$\sigma^2 = \frac{\mathcal{V}}{\left(\mathbb{E}[\widetilde{V}_{it}^2]\right)^2}, \qquad \mathcal{V} := \sum_{k=-\infty}^{\infty} \mathrm{Cov}\left(\widetilde{V}_{it}\widetilde{U}_{it}, \widetilde{V}_{i,t+k}\widetilde{U}_{i,t+k}\right).$$

where $\widetilde{V}_{it} = [\Pi(\boldsymbol{D}_i - \boldsymbol{m}_0(\mathbf{X}_i))]_t$, $\widetilde{U}_{it} = [\Pi(\mathbf{Y}_i - \theta_0\boldsymbol{D}_i - \boldsymbol{l}_0(\mathbf{X}_i))]_t$, and the long-run variance $\mathcal{V}$ reduces to $\mathbb{E}[\widetilde{V}_{it}^2\widetilde{U}_{it}^2]$ when $\widetilde{V}_{it}\widetilde{U}_{it}$ is serially uncorrelated.

Monte Carlo Simulations

Data-generating process:

$$Y_{it} = V_{it}\,\theta_0 + l_0(\mathbf{X}_{it}) + \boldsymbol{\lambda}_i^\top \boldsymbol{f}_t + U_{it}, \tag{8}$$

$$V_{it} = D_{it} - m_0(\mathbf{X}_{it}) - \boldsymbol{\phi}_i^\top \boldsymbol{f}_t, \tag{9}$$

$$\mathbf{X}_{it} = \boldsymbol{\Gamma}_i^\top \boldsymbol{f}_t + \boldsymbol{E}_{it}, \tag{10}$$

$N : \{20,\ 100,\ 500\}$ $T : \{30,\ 50,\ 100\}$ $p : \{50,\ 100,\ 300\}$

Data-generating process:

- latent factors $r = 2$; $\theta_0 = 1$;
- 100 replications
- Three nuisance designs (see next slide):
  - DGP1: linear nuisance functions
  - DGP2: nonlinear, non-smooth
  - **DGP3: nonlinear, discontinuous**

Estimators:

- IFE (Bai, 2009) (benchmark)
- Panel DML-IFE (Lasso)
- Panel DML-IFE (XGBoost)
- Panel DML-IFE (NNet)

Monte Carlo: DGP Specifications

Only $X_{it,1}$ and $X_{it,2}$ (of p covariates) enter the nuisance functions. Parameters: $a_1 = b_2 = 0.25$, $a_2 = b_1 = 0.5$.

DGP1:

$$
\begin{aligned}
l_0 &= a_1 X_{it,1} + a_2 X_{it,2} \\
m_0 &= b_1 X_{it,1} + b_2 X_{it,2}
\end{aligned}
$$

DGP2:

$$
\begin{aligned}
l_0 &= a_1 \max(X_{it,1}, 0) + a_2 |X_{it,2}| \\
m_0 &= b_1 \mathbb{1}\{X_{it,1} > 0\} + b_2 \log(1 + |X_{it,2}|)
\end{aligned}
$$

DGP3:

$$
\begin{aligned}
l_0\left(\boldsymbol{X}_{it}\right) &= a_1\left(X_{it,1} \cdot X_{it,2}\right) + a_2\left(X_{it,2} \cdot \mathbb{1}\{X_{it,2} > 0\}\right) \\
m_0\left(\boldsymbol{X}_{it}\right) &= b_1\left(X_{it,1} \cdot \mathbb{1}\{X_{it,1} > 0\}\right) + b_2\left(X_{it,1} \cdot X_{it,2}\right)
\end{aligned}
$$

Results: DGP3, $N = 20$ (Small Panels)

Bias $= \hat{\theta} - 1$; **columns:** $p = \{50, 100, 300\} \times T = \{30, 100\}$:

	$p = 50$		$p = 100$		$p = 300$	
Estimator	$T = 30$	$T = 100$	$T = 30$	$T = 100$	$T = 30$	$T = 100$
IFE	0.245	0.207	0.293	0.215	0.295	0.243
infeasible bias (IFE)	0.199	0.203	0.204	0.206	0.202	0.206
DML-IFE (Lasso)	0.126	0.105	0.083	0.061	0.063	0.033
infeasible bias (Lasso)	0.050	0.016	0.045	0.017	0.051	0.017
DML-IFE (XGBoost)	0.039	0.038	0.032	-0.027	0.044	-0.024
infeasible bias (XGBoost)	0.001	-0.049	0.009	-0.054	0.051	-0.031
DML-IFE (NNet)	-0.151	0.005	-0.357	-0.192	-0.057	-0.627
infeasible bias (NNet)	-0.126	-0.055	-0.256	-0.195	-0.082	-0.577

Results: DGP3, $N = 100$ (Medium Panels)

Bias $= \hat{\theta} - 1$; **columns:** $p = \{50, 100, 300\} \times T = \{30, 100\}$:

	$p = 50$		$p = 100$		$p = 300$	
Estimator	$T = 30$	$T = 100$	$T = 30$	$T = 100$	$T = 30$	$T = 100$
IFE	0.202	0.201	0.200	0.199	0.202	0.201
infeasible bias (IFE)	0.206	0.206	0.205	0.207	0.207	0.206
DML-IFE (Lasso)	0.118	0.094	0.084	0.058	0.058	0.030
infeasible bias (Lasso)	0.043	0.014	0.044	0.016	0.044	0.016
DML-IFE (XGBoost)	0.101	0.064	0.038	0.020	-0.003	0.000
infeasible bias (XGBoost)	0.010	-0.026	0.002	-0.020	0.009	-0.013
DML-IFE (NNet)	0.071	0.086	-0.055	0.031	-0.459	-0.296
infeasible bias (NNet)	0.010	-0.012	-0.121	-0.006	-0.525	-0.294

Results: DGP3, $N = 500$ (Large Panels)

Bias $= \hat{\theta} - 1$; **columns:** $p = \{50, 100, 300\} \times T = \{30, 100\}$:

	$p = 50$		$p = 100$		$p = 300$	
Estimator	$T = 30$	$T = 100$	$T = 30$	$T = 100$	$T = 30$	$T = 100$
IFE	0.198	0.199	0.199	0.199	0.198	0.199
infeasible bias (IFE)	0.205	0.205	0.204	0.205	0.207	0.205
DML-IFE (Lasso)	0.116	0.092	0.079	0.057	0.055	0.030
infeasible bias (Lasso)	0.041	0.014	0.040	0.015	0.041	0.015
DML-IFE (XGBoost)	0.084	0.072	0.047	0.033	0.026	0.003
infeasible bias (XGBoost)	0.004	-0.004	0.006	-0.006	0.009	-0.009
DML-IFE (NNet)	0.105	0.084	0.061	0.039	-0.107	-0.009
infeasible bias (NNet)	0.033	-0.015	0.023	-0.018	-0.103	-0.028

Empirical Application: Asset Pricing Application by Rücker et al. (2025)

Research question:
Which firm characteristics have a **causal** effect on monthly excess stock returns?

Data:

- $N = 29$ Dow Jones large-cap stocks
- $T = 72$ months (April 2016 – March 2022)
- $p = 89$ lagged firm characteristics as controls
- Treatment variable: Panel A: Bid-ask spread, Panel B: Inventory change
  Panel C: Dividend omission, Panel F: Scaled earning forecast

Estimators: OLS, IFE, DML-FE (Lasso, Boosting, NNet), **DML-IFE (Lasso, Boosting, NNet)**

Economic relevance: latent factors include global risk, liquidity cycles, macro shocks, where IFE is needed.

Empirical Application: Asset Pricing Application by Rücker et al. (2025)

Results: ▶ Empirical Application

1. OLS estimates are systematically unreliable: they either overstate effects or fail to detect genuine effects (Panel A,B,F).
2. The IFE estimator relies on the assumption of linearity in the control variables, making it vulnerable to functional form misspecification (Panel A,B,F).
3. Third, the DML-IFE estimator consistently delivers estimates that are robust to both sources of confounding simultaneously (Panel B,C,F).

Summary and Contributions

This paper...

1. **We proposed a DML framework for panel data with interactive fixed effects**

   Combines CCE-based projection with Neyman-orthogonal score construction

2. **Monte Carlo confirmed finite-sample properties**

   Any ML learner; no iterative PCA; $\sqrt{NT}$-consistent

3. **R Package: xtifedml**

   Easy to implement, available on CRAN soon

Thank You

Questions and Comments Welcome

Binzhi Chen

bc25911@essex.ac.uk

Annalivia Polselli

annalivia.polselli@essex.ac.uk

Paul Clarke

pclarke@essex.ac.uk

Institute for Social and Economic Research
University of Essex, Colchester, UK

References I

Ahrens, A., Chernozhukov, V., Hansen, C., Kozbur, D., Schaffer, M., and Wiemann, T. (2025). An introduction to double/debiased machine learning. *arXiv preprint arXiv:2504.08324*.

Avila Marquez, M. (2025). Weak instrumental variables due to nonlinearities in panel data: A super learner control function estimator. *arXiv preprint arXiv:2504.03228*. Fift version; Last revised 12 November 2025.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Baiardi, A., Clarke, P. S., Naghi, A. A., and Polselli, A. (2026). Double machine learning for static panel data with instrumental variables: New method and applications. *arXiv preprint arXiv:2603.20464*.

Chen, J., Cui, G., Sarafidis, V., and Yamagata, T. (2025). Iv estimation of heterogeneous spatial dynamic panel models with interactive effects. *arXiv preprint arXiv:2501.18467*.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

Chudik, A. and Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, 188(2):393–420.

References II

Clarke, P. S. and Polselli, A. (2025). Double machine learning for static panel models with fixed effects. *Econometrics Journal*, page utaf011.

Ditzen, J. and Karavias, Y. (2025). Interactive, grouped and non-separable fixed effects: A practitioner's guide to the new panel data econometrics. *arXiv preprint arXiv:2507.19099*.

Ditzen, J., Karavias, Y., and Westerlund, J. (2025). Multiple structural breaks in interactive effects panel data models. *Journal of Applied Econometrics*, 40(1):74–88.

Juodis, A., Karabiyik, H., and Westerlund, J. (2021). On the robustness of the pooled cce estimator. *Journal of Econometrics*, 220(2):325–348.

Karavias, Y., Narayan, P. K., and Westerlund, J. (2023). Structural breaks in interactive effects panels and the stock market reaction to covid-19. *Journal of Business & Economic Statistics*, 41(3):653–666.

Miao, K., Li, K., and Su, L. (2020). Panel threshold models with interactive fixed effects. *Journal of Econometrics*, 219(1):137–170.

Moon, H. R. and Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1):158–195.

References III

Moon, H. R. and Weidner, M. (2018). Nuclear norm regularized estimation of panel regression models. *arXiv preprint arXiv:1810.10987*.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: journal of the Econometric Society*, pages 931–954.

Rücker, M., Vogt, M., Linton, O., and Walsh, C. (2025). Estimation and inference in high-dimensional panel data models with interactive fixed effects. *Quantitative Economics*, 16(4):1457–1509.

Semenova, V., Goldman, M., Chernozhukov, V., and Taddy, M. (2023). Inference on heterogeneous treatment effects in high-dimensional dynamic panels under weak dependence. *Quantitative Economics*, 14(2):471–510.

Assumption

Assumption (A.3 Machine learning rate conditions)

Let $\widehat{l}_k$ *and* $\widehat{m}_k$ *denote the predictions of the true population parameters* l_0 *and* m_0 *trained on the complement of fold* k. *There exists a sequence* $\delta_{NT} \to 0$ *such that, for each* $k = 1, \ldots, K$,

$$\left\|\widehat{l}_k - l_0\right\|_{L_2} = O_p(\delta_{NT}), \qquad \left\|\widehat{m}_k - m_0\right\|_{L_2} = O_p(\delta_{NT}),$$

where $\|\cdot\|_{L_2}$ *denotes the* $L_2(P)$*-norm. Moreover,* $\delta_{NT} = o\big((NT)^{-1/4}\big)$, *so that* $\delta_{NT}^2 = o\big((NT)^{-1/2}\big)$.

The Neyman orthogonal score

For the partialled out representation, we follow Section 2.2.4 of Chernozhukov et al. (2018) for the projected moment condition at the unit level

$$\mathbb{E}\big[\Pi_0\{\boldsymbol{Y}_i - (\boldsymbol{D}_i - \boldsymbol{m}_0(\mathbf{X}_i))\theta_0 - \boldsymbol{l}_0(\mathbf{X}_i)\} \mid \mathbf{X}_i, \boldsymbol{D}_i\big] = \mathbf{0}_T. \tag{11}$$

Set $W_i = (\mathbf{Y}_i, \mathbf{D}_i, \mathbf{X}_i)$, $R_i = (\mathbf{D}_i, \mathbf{X}_i)$, $Z_i = \mathbf{X}_i$, $\gamma(W_i; \theta, l) = \Pi_0(\boldsymbol{Y}_i - \boldsymbol{D}_i\theta - \boldsymbol{l}(\mathbf{X}_i))$. The Neyman orthogonal score takes the form

$$\psi_i = \mu_0(R_i)\,\gamma(W_i; \theta_0, l_0) = -\sum_{t=1}^{T} \widetilde{V}_{it}\,\widetilde{U}_{it}, \tag{12}$$

where

$$\mu_0(R_i) = A(R_i)'\Omega(R_i)^{-} - G(Z_i)\,J_l(R_i)'\,\Omega(R_i)^{-} \tag{13}$$

Empirical Results: Selected Firm Characteristics

Table: The Effect of Firm Characteristics on Stock Returns

Estimator:	OLS	IFE	DML-FE			DML-IFE		
Learner:	–	–	Lasso	Boosting	NNet	Lasso	Boosting	NNet
Panel A: Bid-Ask Spread								
	0.946***	-0.207	0.161***	0.941**	0.04	-0.007	-0.159	-0.006
	(0.222)	(0.343)	(0.042)	(0.46)	(0.03)	(0.084)	(0.28)	(0.344)
Model RMSE			0.564	0.602	0.590	0.542	0.569	0.570
RMSE_l			0.070	0.074	0.074	0.069	0.070	0.069
RMSE_m			0.024	0.008	0.080	0.014	0.006	0.008
Panel B: Change in inventory								
	-0.144	-0.07	-0.107	-0.032	0.041	-0.037***	-0.387*	-0.195**
	(0.299)	(0.159)	(0.081)	(0.084)	(0.095)	(0.013)	(0.234)	(0.097)
Model RMSE			0.564	0.599	0.597	0.538	0.571	0.569
RMSE_l			0.069	0.074	0.074	0.069	0.070	0.069
RMSE_m			0.035	0.036	0.054	0.126	0.018	0.023

$N = 29$ firms, $T = 72$ months. Significance: $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$. SE: asymptotic (OLS); Robust (IFE); cluster-robust at firm level (DML).

Empirical Results: Selected Firm Characteristics

Table: The Effect of Firm Characteristics on Stock Returns

Estimator:	OLS	IFE	DML-FE			DML-IFE		
Learner:	–	–	Lasso	Boosting	NNet	Lasso	Boosting	NNet
Panel C: Dividend omission								
	-0.057	-0.041*	0.000	-0.024	-0.047***	-0.024***	-0.038**	-0.021
	(0.089)	(0.023)	(0.003)	(0.015)	(0.018)	(0.006)	(0.017)	(0.02)
Model RMSE			0.564	0.628	0.593	0.542	0.620	0.571
RMSE_l			0.069	0.075	0.073	0.069	0.074	0.069
RMSE_m			0.477	0.167	0.093	0.104	0.143	0.087
Panel F: Scaled earnings forecast								
	-0.099	-0.004	-0.067	-0.193***	-0.144***	-0.092***	-0.216***	0.003
	(0.114)	(0.081)	(0.061)	(0.023)	(0.04)	(0.013)	(0.074)	(0.077)
Model RMSE			0.565	0.648	0.593	0.541	0.573	0.569
RMSE_l			0.069	0.075	0.072	0.069	0.070	0.069
RMSE_m			0.058	0.049	0.059	0.074	0.028	0.031

$N = 29$ firms, $T = 72$ months. Significance: $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$. SE: asymptotic (OLS); Robust (IFE); cluster-robust at firm level (DML).