Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas

Eureka Journal of Business, Economics & Innovation Studies (EJBEIS)

ISSN 2760-4950 (Online) Volume 2, Issue 5, May 2026

https://eurekaoa.com/index.php/6

APPLIED ECONOMETRICS: THEORETICAL FOUNDATIONS, ESTIMATION TECHNIQUES, AND EMPIRICAL APPLICATIONS

Djumanazarova Zamira Kojaboyevna
Senior Lecturer, Department of Business and Management,
Oriental University, Tashkent, Uzbekistan

ABSTRACT

Applied econometrics bridges economic theory and empirical data, equipping researchers with rigorous statistical tools for causal inference, forecasting, and policy evaluation. This article provides a systematic overview of the core methods constituting the applied econometrics toolkit: Ordinary Least Squares (OLS) regression and its classical assumptions, panel data estimators (pooled OLS, fixed effects, and random effects), instrumental variable and two-stage least squares (IV/2SLS) estimation, time-series analysis including unit-root and cointegration tests, robust and cluster-robust inference, and dynamic panel GMM methods. For each technique the theoretical motivation, identifying assumptions, diagnostic procedures, and remedies for assumption violations are presented. The article further synthesises selected empirical applications drawn from development economics, household finance, and transition economy contexts — with particular reference to the determinants of household savings in Uzbekistan (Djumanazarova, 2025) — illustrating how methodological choices shape substantive conclusions. Cross-tabular comparisons of estimator properties and a practical implementation guide facilitate both classroom instruction and independent research.[1] The discussion concludes with a forward-looking

---

[1] Applied econometrics is defined here as the use of statistical and mathematical methods to test economic theory against empirical data. For a comprehensive treatment, see Wooldridge (2010) and Greene (2018).

assessment of emerging frontiers: machine learning augmentation of econometric models, causal forests, and synthetic control methods.

Keywords: Applied econometrics, OLS regression, panel data, fixed effects, instrumental variables, cointegration, GMM, robust inference, household savings, transition economy.

1. INTRODUCTION

Econometrics — literally, "economic measurement" — occupies a distinctive methodological position within social science. Unlike pure statistical analysis, which may be agnostic about causality, econometrics is explicitly concerned with the identification of causal relationships implied by economic theory (Heckman, 2000). Unlike theoretical economics, it disciplines abstract models with the constraints imposed by real data. The result is a discipline that is simultaneously a branch of mathematics, statistics, and empirical social science.

The evolution of applied econometrics over the past half-century has been dramatic. Early cross-sectional regression studies, relying on strong assumptions about random sampling and exogeneity, gave way to panel data methods as micro-level longitudinal datasets became available (Mundlak, 1978; Hausman, 1978). The credibility revolution of the 1990s — associated with Card, Krueger, Angrist, and others — placed causal identification at the centre of empirical practice, popularising quasi-experimental designs such as instrumental variables, regression discontinuity, and difference-in-differences (Angrist & Pischke, 2009). Most recently, the machine learning literature has begun to offer complementary tools for high-dimensional covariate selection, heterogeneous treatment effects, and prediction (Athey & Imbens, 2019).

This article provides a structured overview of the principal techniques within the applied econometrics toolkit, situating each within its theoretical context, laying out its core assumptions, and illustrating its application with reference to current empirical research. The motivating empirical case throughout is the study of household savings determinants in Uzbekistan conducted by Djumanazarova (2025), whose methodological choices — panel fixed effects estimation, the Hausman specification test, and cluster-robust standard errors — exemplify applied econometric practice in a transition economy setting. Supplementary illustrations are drawn from the canonical literature on development economics and macroeconometrics.

The remainder of the article is structured as follows. Section 2 reviews the classical OLS framework. Section 3 extends the discussion to panel data methods. Section 4 addresses endogeneity and instrumental variable estimation. Section 5 covers time-series econometrics including unit-root and cointegration analysis. Section 6 discusses robust inference. Section 7 introduces dynamic panel GMM. Section 8 synthesises empirical applications. Section 9 concludes.

2. CLASSICAL OLS: FOUNDATIONS AND DIAGNOSTICS

2.1 The Classical Linear Regression Model

The point of departure for applied econometrics is the Classical Linear Regression Model (CLRM). Let y be an $(n \times 1)$ vector of observations on the dependent variable, X an $(n \times k)$ matrix of observations on k explanatory variables (including a unit vector for the intercept), β a $(k \times 1)$ vector of unknown parameters, and ε an $(n \times 1)$ vector of unobservable disturbances. The model asserts:

$$y = X\beta + \varepsilon$$

Under the six Gauss–Markov assumptions (Table 1), the OLS estimator $\hat{b} = (X'X)^{-1}X'y$ is the Best Linear Unbiased Estimator (BLUE) of β, meaning it achieves the lowest variance among all linear unbiased estimators (Gauss–Markov theorem).[2] When errors are additionally normally distributed, the OLS

---

[2] OLS requires the Gauss–Markov conditions: linearity in parameters, random sampling, no perfect multicollinearity, zero conditional mean of errors, and homoskedasticity. Violation of any condition necessitates corrective techniques discussed in Section 3.

estimator is also the Maximum Likelihood Estimator, enabling exact finite-sample inference via t and F distributions.

Table 1. Gauss–Markov Assumptions, Diagnostic Tests, and Remedies

Assumption	Formal Statement	Diagnostic Test	Remedy if Violated
A1. Linearity	$E(y\|X) = X\beta$	RESET test; residual plots	Non-linear transformation; polynomial terms
A2. Random sampling	i.i.d. observations	Study design review	Clustered sampling; survey weights
A3. No multicollinearity	rank(X) = k (full column rank)	VIF > 10 signals concern	Drop / combine collinear predictors; Ridge
A4. Zero cond. mean	$E(\varepsilon\|X) = 0$	Endogeneity tests (Hausman)	IV/2SLS; control functions
A5. Homoskedasticity	$Var(\varepsilon\|X) = \sigma^2$	Breusch–Pagan; White test	HC-robust SE; WLS; FGLS
A6. Normality (finite samples)	$\varepsilon\|X \sim N(0, \sigma^2 I)$	Jarque–Bera; Q-Q plot	Large-sample CLT; bootstrap inference

Note: Sources: Wooldridge (2010, Ch. 3); Greene (2018, Ch. 4). VIF = Variance Inflation Factor. HC = Heteroskedasticity-Consistent. WLS = Weighted Least Squares. FGLS = Feasible Generalised Least Squares.

2.2 Heteroskedasticity

Heteroskedasticity — non-constant variance of the error term — is endemic in cross-sectional microeconomic data. Its presence does not bias OLS coefficient estimates but renders the conventional OLS standard errors inconsistent, invalidating t and F tests.[3] The standard remedy is the use of Heteroskedasticity-Consistent (HC) standard errors proposed by White (1980), which are valid

---

[3] The Breusch–Pagan (1979) test regresses squared OLS residuals on the explanatory variables. Rejection of H_0 (constant variance) signals heteroskedasticity and mandates the use of HC-robust or cluster-robust standard errors.

asymptotically under unknown forms of heteroskedasticity. In grouped or clustered data — such as students within schools or households within regions — cluster-robust standard errors account for within-cluster error correlation as well as heteroskedasticity (Cameron & Trivedi, 2010).

2.3 Serial Correlation

In time-series regressions, the error terms εt and εt−1 are frequently correlated — a phenomenon known as autocorrelation or serial correlation.[4] The Durbin–Watson statistic and the Breusch–Godfrey test are standard diagnostics. Newey and West (1987) propose heteroskedasticity- and autocorrelation-consistent (HAC) standard errors that account for both problems simultaneously, making them the workhorse of applied macroeconometrics.

3. PANEL DATA METHODS

3.1 The One-Way Error Components Model

Panel (or longitudinal) data track the same cross-sectional units (individuals, firms, regions, countries) across multiple time periods. The canonical one-way error components model decomposes the disturbance into a time-invariant individual effect μi and an idiosyncratic shock εit:

$$y\ it = x'it\ \beta + \mu i + \varepsilon it$$

The individual effect μi captures unobserved time-invariant heterogeneity — cultural propensities to save, geographic endowments, institutional quality — that would otherwise contaminate cross-sectional estimates. Whether μi is treated as a fixed unknown parameter (fixed effects) or a random variable (random effects)

---

[4] The Durbin–Watson statistic ranges from 0 to 4. Values near 2 indicate no autocorrelation; values substantially below 2 suggest positive serial correlation. For panel data, the Wooldridge (2002) test for serial correlation in the idiosyncratic error is preferred.

is the central methodological question in panel econometrics, resolved empirically by the Hausman (1978) specification test.[5]

3.2 Estimator Comparison

Table 2 compares the principal panel estimators along the dimensions most relevant to applied researchers. The fixed effects (FE) or within estimator eliminates μi by time-demeaning the data, yielding consistent estimates whenever the regressors are correlated with the individual effect — the typical scenario in economic applications. The random effects (RE) estimator is more efficient under the stricter assumption that μi is uncorrelated with all regressors; when this assumption fails, RE estimates are biased and inconsistent. The Hausman test formally discriminates between the two by examining whether the difference in coefficient vectors is statistically significant.

In Djumanazarova (2025), the Hausman test statistic $\chi^2 = 34.71$ ($p < 0.001$) strongly rejects the random effects specification, confirming that household-level and regional unobservables are correlated with the explanatory variables and that fixed effects estimation is required for consistent inference on the saving determinants.[6]

---

[5] The Hausman (1978) test evaluates whether the difference between FE and RE coefficients is systematic. A significant test statistic ($p < 0.05$) indicates that RE estimates are inconsistent and FE is preferred. Djumanazarova (2025) applies this test in the household savings context, obtaining $\chi^2 = 34.71$ ($p < 0.001$).

ISSN 2760-4950 (Online) Volume 2, Issue 5, May 2026

https://eurekaoa.com/index.php/6

Table 2. Comparison of Panel Data Estimators

Estimator	Notation	Key Assumption	Consistency	Efficiency	Best Used When
Pooled OLS	POLS	No unobserved heterogeneity	Only if no RE/FE	High	No individual effects
Random Effects	RE/GLS	Individual effects ⊥ regressors	If RE valid	Highest	RE uncorrelated with X
Fixed Effects	FE/LSDV	Individual effects may correlate with X	Always (within var)	Lower	RE correlated with X; n large
First Differences	FD	Strict exogeneity after differencing	Under strict exog.	Lower	T = 2 or strong serial corr.
Hausman–Taylor	HT	Some X correlated with effects	If instruments valid	Moderate	Mixed endogeneity

Note: Sources: Baltagi (2013, Ch. 2–4); Wooldridge (2010, Ch. 10–11). FE = Fixed Effects. RE = Random Effects. FD = First Differences. HT = Hausman–Taylor. LSDV = Least Squares Dummy Variable.

3.3 The Mundlak Device

An elegant alternative to the binary FE/RE choice is the Mundlak (1978) device, which augments the RE model with the within-group means of all time-varying regressors: x̄i. If the coefficients on x̄i are jointly zero, the RE and FE estimates coincide and RE is preferred for efficiency. Otherwise, the augmented RE model captures the correlation between individual effects and regressors without discarding the between-group variation exploited by RE. This device is particularly useful when the researcher wishes to estimate the effects of time-invariant variables (e.g., gender, nationality) that are swept out by the FE transformation.

4. ENDOGENEITY AND INSTRUMENTAL VARIABLE ESTIMATION

4.1 Sources of Endogeneity

The zero conditional mean assumption — $E(\varepsilon|X) = 0$ — is arguably the most consequential in applied econometrics, since its violation renders OLS estimates biased and inconsistent. Endogeneity arises from three principal sources: (i) omitted variables that are correlated with both the regressor and the outcome; (ii) measurement error in the explanatory variable (classical attenuation bias); and (iii) simultaneous causality, where the outcome and regressor jointly determine each other (Wooldridge, 2010, Ch. 15).

A canonical example in the savings literature is the relationship between income and saving. While OLS estimates the conditional correlation between the two, income is itself partly determined by saving decisions (through capital accumulation) and is correlated with unobserved time preference parameters that independently drive both saving and labour supply. The direction of OLS bias is ambiguous a priori and must be addressed through instrumental variable methods.[7]

4.2 The IV/2SLS Estimator

Let z denote a $(k \times 1)$ vector of instruments satisfying (i) $\mathrm{Cov}(z, \varepsilon) = 0$ (exogeneity) and (ii) $\mathrm{Cov}(z, x) \neq 0$ (relevance).[8] The Two-Stage Least Squares (2SLS) estimator proceeds in two steps. In the first stage, the endogenous regressor x is regressed on all exogenous variables and the instruments z, yielding fitted values $\hat{x}$. In the second stage, y is regressed on $\hat{x}$ and the remaining exogenous regressors. The 2SLS estimator is consistent under the IV assumptions

---

[7] In the context of Uzbekistan's household saving study (Djumanazarova, 2025), the financial inclusion variable (% adults with a formal account) rose from 22% in 2014 to 57% in 2021 (World Bank Global Findex, 2022), making it a natural candidate for instrumental variable analysis.

[8] Instrumental Variables (IV) estimation requires instruments that are (i) correlated with the endogenous regressor (relevance) and (ii) uncorrelated with the error term (exogeneity). Weak instruments — those with low first-stage F-statistics (typically < 10) — yield biased IV estimates.

and collapses to OLS when the instruments are the explanatory variables themselves.

Instrument validity can be partially tested: the first-stage F-statistic (should exceed 10 for a single endogenous variable; Staiger & Stock, 1997) assesses relevance, while the Sargan–Hansen J-test assesses overidentifying restrictions when the model is overidentified. Exogeneity of excluded instruments cannot be directly tested; it rests on theoretical and contextual argumentation.

5. TIME-SERIES ECONOMETRICS

5.1 Stationarity and Unit Root Tests

A time series yt is (weakly) stationary if its mean, variance, and autocovariances are time-invariant. Regressions involving non-stationary (integrated) series risk spurious regression — a phenomenon in which high R^2 and significant t-statistics arise purely from shared trends rather than genuine causal relationships (Granger & Newbold, 1974). The standard diagnostic is the Augmented Dickey–Fuller (ADF) test, which tests the null hypothesis of a unit root against the alternative of stationarity.[9]

Table 3 summarises the principal time-series diagnostic tests. A series is said to be integrated of order d — denoted I(d) — if it must be differenced d times to achieve stationarity. Most macroeconomic variables (GDP, price levels, interest rates) are I(1), meaning the first difference is stationary. Differencing removes the unit root but eliminates long-run information, motivating the cointegration approach.

---

[9] The ADF test null hypothesis is non-stationarity (unit root). Rejection requires a sufficiently negative test statistic. Phillips and Perron (1988) offer a nonparametric alternative that is robust to heteroskedastic and autocorrelated errors.

Table 3. Time-Series Diagnostic Tests: Hypotheses, Statistics, and Decision Rules

Test	Null Hypothesis (H_0)	Test Statistic	Decision Rule
Augmented Dickey–Fuller (ADF)	Series has a unit root	τ (t-ratio on ρ)	Reject H_0 if $\tau <$ critical value
Phillips–Perron (PP)	Series has a unit root	Z(t) nonparametric	Same as ADF; robust to serial corr.
KPSS	Series is stationary	LM statistic	Reject H_0 if LM > critical value
Johansen Trace	At most r cointegrating vectors	Trace statistic	Reject if trace > $\chi^2(p-r)$ critical
Engle–Granger	No cointegration (residuals I(1))	ADF on residuals	Reject H_0 if ADF rejects unit root
Granger Causality	X does not Granger-cause Y	F-statistic on lagged X	Reject H_0 if F > critical value

Note: Sources: Engle & Granger (1987); Johansen (1991); Dickey & Fuller (1979); Phillips & Perron (1988); Kwiatkowski et al. (1992).

5.2 Cointegration and Error Correction Models

Two or more I(1) series are cointegrated if a linear combination of them is I(0) — stationary. Cointegration implies the existence of a long-run equilibrium relationship towards which the series converge, even though each individual series wanders non-stationarily in the short run (Engle & Granger, 1987). This is the formal econometric representation of concepts such as purchasing power parity, the Fisher equation, and long-run consumption functions.

The Johansen (1991) maximum likelihood procedure allows simultaneous testing for multiple cointegrating vectors and estimation of the cointegrating space.[10]

---

[10] The Johansen (1991) trace test sequentially tests H_0: at most r cointegrating vectors. The procedure starts at $r = 0$ and stops when H_0 cannot be rejected, giving the cointegration rank. Normalisation on one variable allows interpretation of long-run elasticities.

Once cointegration is established, the dynamic relationship is modelled by a Vector Error Correction Model (VECM), in which the short-run dynamics of each variable are partly driven by the deviation from the long-run equilibrium (the error correction term). The coefficient on the error correction term — the speed of adjustment parameter γ — must be negative and significant for genuine cointegration.

6. ROBUST AND CLUSTER-ROBUST INFERENCE

Modern applied econometrics routinely reports robust standard errors as a matter of best practice, regardless of whether formal tests reject homoskedasticity.[11] The logic is precautionary: if the true variance is homoskedastic, HC standard errors are slightly less efficient than conventional standard errors but remain valid; if heteroskedasticity is present, conventional standard errors are inconsistent and inference is unreliable.

When observations are clustered — meaning that within-cluster errors are correlated even after conditioning on covariates — cluster-robust standard errors are required. Cameron and Miller (2015) provide a comprehensive review, emphasising that the number of clusters drives asymptotic validity: fewer than 20–30 clusters renders cluster-robust inference unreliable, motivating wild bootstrap procedures as an alternative.[12]

In Djumanazarova (2025), standard errors are clustered at the regional (viloyat) level to account for within-region error correlation across the thirteen years of the panel. With fourteen regions, the cluster count is at the lower boundary of reliable cluster-robust inference; robustness to bootstrapped standard errors is therefore an important validation exercise.

---

[11] Heteroskedasticity-Consistent (HC) standard errors — due to White (1980) — are asymptotically valid under unknown forms of heteroskedasticity. Extensions HC1–HC4 offer finite-sample corrections; MacKinnon and White (1985) recommend HC3 for small samples.

[12] Bootstrapping resamples the data (with replacement) B times (typically B = 999 or 1,999) and re-estimates the model each time. The resulting empirical distribution of the statistic provides standard errors and confidence intervals without distributional assumptions.

7. DYNAMIC PANEL GMM

The Arellano–Bond (1991) Generalised Method of Moments (GMM) estimator addresses two simultaneous challenges in panel data: (i) dynamic misspecification arising when a lagged dependent variable appears as a regressor (violating strict exogeneity required by FE), and (ii) endogeneity of other regressors. The estimator first-differences the model to eliminate fixed effects and then instruments the differenced lagged dependent variable (and other endogenous variables) with their own levels lagged two or more periods.

The validity of GMM depends on two conditions: instrument exogeneity (tested via the Sargan–Hansen J-test) and second-order serial correlation in the differenced residuals (the AR(2) test must fail to reject the null of no autocorrelation). Blundell and Bond (1998) extend the approach with a system GMM estimator that adds the levels equation to the first-differenced equation, using lagged differences as additional instruments and substantially reducing the finite-sample bias from instrument proliferation — a problem that arises when T is large relative to N.

8. EMPIRICAL APPLICATIONS: COMPARATIVE SURVEY

Table 4 synthesises a representative cross-section of empirical studies that deploy the methods reviewed above, illustrating how methodological choice is conditioned by the research question, data structure, and identification challenge.

Table 4. Selected Empirical Applications of Applied Econometric Methods

Study	Country / Sample	Methodology	Key Finding
Loayza et al. (2000)	69 developing countries, 1965–1994	Panel OLS, FE/RE	Income elasticity of savings $\approx$ 0.3–0.5
Djumanazarova (2025)	Uzbekistan, 14 regions, 2010–2022	Panel FE-OLS, Hausman test	Income $\beta=0.312^{***}$; financial inclusion $\beta=0.196^{***}$
Allen et al. (2016)	148 countries, cross-section	Probit; IV	Account ownership raises formal saving probability
Engle & Granger (1987)	US macro time series	Cointegration, ECM	Long-run equilibrium identified between I(1) series
Arellano & Bond (1991)	UK firm-level panel	GMM first differences	Dynamic panel bias corrected; persistent labour demand
Wooldridge (2002)	Methodological (textbook)	N/A	Cluster-robust SE; Mundlak device; serial corr. test

Note: Sources: as cited. Djumanazarova (2025) results reproduced from Table 3 of that study: β = standardised OLS coefficient, *** $p < 0.01$, ** $p < 0.05$.

The Djumanazarova (2025) study is particularly instructive as an exemplar of applied econometric practice in a transition economy context. The study's use of panel fixed effects controls for time-invariant regional heterogeneity (cultural propensities to save, geographic remoteness, historical institutional quality) that would confound cross-sectional estimates. The Hausman test provides a principled basis for preferring FE over RE. Cluster-robust standard errors account for within-region correlation across years. The sequential robustness checks — urban/rural subsamples, alternative inflation measure, RE comparison — follow the standard credibility-enhancing protocol recommended by Angrist and Pischke (2009).

The study's findings — that per capita income ($\beta = 0.312$, $p < 0.01$), educational attainment ($\beta = 0.187$, $p < 0.05$), and financial inclusion ($\beta = 0.196$, $p < 0.01$) are the dominant positive determinants of household saving rates, while household size ($\beta = -0.214$, $p < 0.01$) and inflation ($\beta = -0.163$, $p < 0.05$) are the principal negative drivers — are substantively consistent with the broader cross-country literature (Loayza et al., 2000; Allen et al., 2016) while providing country-specific quantification relevant to Uzbekistan's policymakers.

9. PRACTICAL IMPLEMENTATION

Applied econometric research follows a disciplined sequence of steps. First, the researcher specifies the estimating equation, derived from economic theory, identifying the dependent variable, regressors, and — where endogeneity is suspected — potential instruments. Second, the data are assembled and cleaned, with careful attention to measurement consistency, missing data mechanisms, and outlier treatment. Third, the baseline model is estimated and the coefficient estimates are subjected to economic and statistical scrutiny. Fourth, a battery of specification tests is conducted to assess assumption validity. Fifth, robustness exercises — alternative specifications, subsamples, estimation methods — validate or qualify the baseline findings.[13]

A practical principle often emphasised in graduate econometrics instruction is that the choice of estimator should be driven by the data-generating process, not by the estimator that produces the most statistically significant results. Data mining — the practice of searching across specifications until desired results are obtained — inflates false discovery rates and produces non-replicable findings. Pre-registration of research designs, routine reporting of all specification choices, and transparent presentation of both significant and null results are increasingly recognised as essential components of credible empirical practice.

---

[13] Stata 18 commands: regress y x1 x2, robust (OLS with robust SE); xtreg y x1 x2, fe (panel FE); ivreg2 y (x1 = z1) x2, robust (IV/2SLS); vecm y x1, lags(2) (VECM). R equivalents: lm(), plm(model='within'), ivreg() [AER package], ca() [urca package].

10. EMERGING FRONTIERS

The boundary between econometrics and machine learning is dissolving. Regularisation methods such as LASSO and Ridge regression address the challenge of covariate selection in high-dimensional settings; the Double LASSO of Belloni, Chernozhukov, and Hansen (2014) combines machine learning variable selection with valid post-selection inference for low-dimensional causal parameters. Causal forests, developed by Wager and Athey (2018), estimate heterogeneous treatment effects non-parametrically, allowing the researcher to identify subgroups for whom a policy intervention is most effective. The synthetic control method of Abadie, Diamond, and Hainmueller (2010) constructs a weighted counterfactual for treated units from donor pool controls, providing a credible approach to comparative case studies that avoids the extrapolation problems of conventional regression-based difference-in-differences.

These developments do not supersede the classical econometric toolkit; they complement it. The Gauss–Markov theorem, the Frisch–Waugh–Lovell theorem, and the principles of causal identification remain foundational. The applied econometrician of the 21st century commands both the classical methods reviewed in this article and the emerging machine learning augmentations, deploying each as the research question demands.

11. CONCLUSION

Applied econometrics provides the methodological infrastructure for empirical economic research. This article has reviewed the core techniques — OLS and its classical assumptions, panel data fixed and random effects estimators, instrumental variable methods, time-series and cointegration analysis, robust inference, and dynamic panel GMM — situating each within its theoretical context and illustrating its application through the lens of recent empirical research on household savings in Uzbekistan.

The central lesson that emerges from both methodological exposition and empirical application is that credible causal inference requires transparency about identifying assumptions. An estimator's validity is only as strong as the assumptions that underlie it; diagnostic tests can detect violations but cannot substitute for theoretical discipline in specification. The applied researcher's task is to select the estimator whose assumptions are most defensible given the available data and theoretical knowledge, communicate those assumptions clearly, and conduct systematic robustness exercises to assess their sensitivity.

The integration of machine learning methods into the econometric toolkit offers exciting prospects for improved covariate selection, non-parametric estimation of heterogeneous effects, and causal discovery from observational data. The coming decade is likely to see the most significant methodological evolution in applied econometrics since the credibility revolution of the 1990s, with potentially transformative implications for policy evaluation and empirical social science more broadly.

REFERENCES

1. Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies. Journal of the American Statistical Association, 105(490), 493–505.
2. Allen, F., Demirguc-Kunt, A., Klapper, L., & Peria, M. S. M. (2016). The foundations of financial inclusion. Journal of Financial Intermediation, 27, 1–30.
3. Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton University Press.
4. Arellano, M., & Bond, S. (1991). Some tests of specification for panel data. Review of Economic Studies, 58(2), 277–297.
5. Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. Annual Review of Economics, 11, 685–725.

6. Baltagi, B. H. (2013). Econometric analysis of panel data (5th ed.). Wiley.
7. Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. Review of Economic Studies, 81(2), 608–650.
8. Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, 87(1), 115–143.
9. Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroskedasticity and random coefficient variation. Econometrica, 47(5), 1287–1294.
10. Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. Journal of Human Resources, 50(2), 317–372.
11. Cameron, A. C., & Trivedi, P. K. (2010). Microeconometrics using Stata (Rev. ed.). Stata Press.
12. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366), 427–431.
13. Djumanazarova, Z. K. (2025). Determinants of household savings in developing economies: An econometric approach with evidence from Uzbekistan. Economics, Management and Money, 3, 131–148.
14. Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. Econometrica, 55(2), 251–276.
15. Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. Journal of Econometrics, 2(2), 111–120.
16. Greene, W. H. (2018). Econometric analysis (8th ed.). Pearson.
17. Hausman, J. A. (1978). Specification tests in econometrics. Econometrica, 46(6), 1251–1271.
18. Heckman, J. J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. Quarterly Journal of Economics, 115(1), 45–97.

19. Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica, 59(6), 1551–1580.
20. Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics, 54(1–3), 159–178.
21. Loayza, N., Schmidt-Hebbel, K., & Servén, L. (2000). What drives private saving across the world? Review of Economics and Statistics, 82(2), 165–181.
22. Mundlak, Y. (1978). On the pooling of time series and cross section data. Econometrica, 46(1), 69–85.
23. Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3), 703–708.
24. Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. Biometrika, 75(2), 335–346.
25. Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. Econometrica, 65(3), 557–586.
26. Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228–1242.
27. White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48(4), 817–838.
28. Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data. MIT Press.
29. Wooldridge, J. M. (2010). Introductory econometrics: A modern approach (4th ed.). South-Western Cengage.