UHERO

THE ECONOMIC RESEARCH ORGANIZATION
AT THE UNIVERSITY OF HAWAIʻI

A Brief History of Time Series Econometrics

by

Peter Fuleky

Working Paper No. 2026-2

October 19, 2026

UNIVERSITY OF HAWAIʻI AT MANOA
2424 MAILE WAY, ROOM 540 • HONOLULU, HAWAIʻI 96822
WWW.UHERO.HAWAII.EDU

WORKING PAPERS ARE PRELIMINARY MATERIALS CIRCULATED TO STIMULATE DISCUSSION AND CRITICAL COMMENT. WHILE WORKING PAPERS BENEFIT FROM ACTIVE UHERO DISCUSSION, THEY HAVE NOT UNDERGONE FORMAL ACADEMIC PEER REVIEW. **THE FINDINGS AND RECOMMENDATIONS ARE THOSE OF THE AUTHOR(S) AND DO NOT REFLECT THE VIEWS OR OFFICIAL POSITIONS OF UHERO, THE UNIVERSITY OF HAWAIʻI, OR ANY FUNDING OR SUPPORTING ORGANIZATION.**

A Brief History of Time Series Econometrics*

Peter Fuleky†

October 19, 2026

Abstract

This chapter surveys a century of time series econometrics, from its statistical foundations to modern, data-rich practice. It traces the field's major turning points: univariate ARIMA modeling, the rise of vector autoregressions and the identification debates that followed, and the treatment of nonstationarity through unit roots, cointegration, and error correction. Later sections take up volatility and nonlinear modeling, state-space signal extraction, and Bayesian estimation. The chapter then follows the shift to high-dimensional forecasting and real-time nowcasting, and the spread of machine learning into time series workflows. It compares the major approaches on forecasting performance, structural interpretability, and the measurement of uncertainty. The methods that endured answered a concrete empirical need, whether for better out-of-sample forecasts, credible policy analysis, or an honest measure of risk. The chapter closes with lessons from the field's evolution and an agenda for future research as data and computation expand.

1 Introduction

Time series econometrics develops and applies quantitative methods for analyzing data observed over time. It is central to modern economics because many objects economists care about are dynamic: output, inflation, employment, asset prices, exchange rates, and interest rates. These series show persistence, cycles, volatility clustering, and structural change. Unlike cross-sectional analysis, time series work must confront temporal dependence and evolving distributions. If those features are ignored, standard regression tools can generate misleading inference, including what Yule (1926) memorably described as "nonsense correlations".

Over the past century, the field has grown from early descriptive studies of business cycles into a body of methods spanning univariate and multivariate modeling, long-run equilibrium analysis, volatility and state-space methods, and, more recently, data-rich and machine-learning approaches. A recurring interaction between practical needs and theoretical innovation has driven this growth. The demands came from forecasting, risk measurement, and policy analysis; the theory came from work on probability foundations, identification, and asymptotic theory. Progress has usually started from a concrete empirical problem or a new kind of data, not from methodological novelty alone.

The chapter proceeds chronologically. The first group of sections traces the core arc of the discipline: statistical foundations and the stochastic view of cycles (Section 2), econometric modeling under the Cowles Commission program (Section 3), Box–Jenkins ARIMA modeling (Section 4), vector autoregressions and the debate over identification (Section 5), and unit roots, cointegration, and error correction (Sections 6 and 7). Subsequent sections cover extensions of this core framework: volatility and nonlinear dynamics (Section 8), filtering and signal extraction via state-space methods (Section 9), Bayesian methods and computation (Section 10), the data-rich era of factor models, shrinkage, and nowcasting (Section 11), and the integration of machine learning into time series workflows (Section 12). The chapter concludes with testing and forecast evaluation (Section 13), a comparative assessment of the major approaches (Section 14), an agenda for future research (Section 15), and enduring lessons for remaining relevant (Section 16). Table 1 lists the milestones the chapter is built around.

*Manuscript prepared for the book *A History of Econometrics: The Challenge of Remaining Relevant*, edited by László Mátyás and Alain Pirotte.
†University of Hawaii, Honolulu, USA, fuleky@hawaii.edu

Table 1: Selected milestones in the development of time series econometrics.

Period	Development
1926–1927	Nonsense correlations; autoregressive models (Yule, 1926, 1927)
1927–1938	Stochastic view of cycles; impulse–propagation; Wold decomposition (Slutsky, 1937; Frisch, 1933; Wold, 1938)
1939–1944	Tinbergen's macroeconometric models; the probability approach (Tinbergen, 1939; Haavelmo, 1944)
1950s	Cowles Commission identification and structural modeling (Koopmans, 1950; Klein, 1950)
1960	Kalman filter (Kalman, 1960)
1970	Box–Jenkins ARIMA methodology (Box & Jenkins, 1970)
1974–1979	Spurious regression; unit-root testing (Granger & Newbold, 1974; Dickey & Fuller, 1979)
1980	Vector autoregressions (Sims, 1980)
1982	Autoregressive conditional heteroskedasticity (Engle, 1982)
1987	Cointegration and error correction (Engle & Granger, 1987)
1989	Markov switching; long-run SVAR identification (Hamilton, 1989; Blanchard & Quah, 1989)
1990s	Bayesian computation; stochastic volatility (Gelfand & Smith, 1990; S. Kim, Shephard, & Chib, 1998)
1995	Predictive accuracy testing (Diebold & Mariano, 1995)
2002–2005	Dynamic factor forecasting; FAVAR; local projections; TVP-VAR (Stock & Watson, 2002; Bernanke, Boivin, & Eliasz, 2005; Jordà, 2005; Primiceri, 2005)
2008–2010	Nowcasting; large Bayesian VARs (Giannone, Reichlin, & Small, 2008; Banbura, Giannone, & Reichlin, 2010)
2010s	Machine learning enters macro forecasting (Mullainathan & Spiess, 2017; Medeiros, Vasconcelos, Veiga, & Zilberman, 2021)
2018–2020	Forecasting competitions; hybrid deep learning (Makridakis, Spiliotis, & Assimakopoulos, 2020; Smyl, 2020)
2023–	Foundation models for time series (Garza, Challu, & Mergenthaler-Canseco, 2023; Ansari et al., 2024)

2 Statistical Foundations (Pre-1940)

The quantitative study of economic time series began in the early twentieth century, and many of the ideas that still shape modern econometrics date from that period. One question stood out early on. Are observed cycles driven by deterministic periodic forces, or can they arise from the propagation of random shocks? Researchers tackled it with time plots, early dynamic models, and demonstrations of where naive correlation analysis goes wrong.

Early answers leaned toward the deterministic view. Periodogram analysis, imported from physics, searched for hidden periodicities in series such as wheat prices (W. H. Beveridge, 1922), while the business 'barometers' of the Harvard school decomposed series into trend, cyclical, seasonal, and irregular components for forecasting (Persons, 1919). The language of that decomposition survives in seasonal adjustment practice to this day (Section 9). Both traditions treated cycles as regular structures to be uncovered rather than as products of randomness.

The break with this view came when researchers recognized serial dependence. Yule (1927) introduced autoregressive (AR) models as a parsimonious way to represent persistence. Fitting a low-order AR to sunspot data, a series previously studied by periodogram methods, he showed that current values can be statistically related to their own past. Closely related was a famous warning about misleading inference in trended data. In his discussion of "nonsense correlations," Yule (1926) showed that high correlations can arise between unrelated series simply because both drift over time. Trending time series can suggest spurious relationships if treated with static tools, a lesson that would recur in later econometric debates.

At roughly the same time, the stochastic interpretation of cycles gained prominence. Slutsky (1937), in work first published in Russian in 1927, showed that simple transformations of random shocks can generate smooth oscillations that resemble business cycles. This complemented the 'impulse and propagation' view of Frisch (1933), in which shocks ('impulses') are transmitted through dynamic economic mechanisms ('propagation'), producing sustained fluctuations even without deterministic periodicity. Between them, Slutsky and Frisch moved attention away from purely deterministic descriptions of cycles and toward models in which randomness and dynamics together shape what we observe.

A key statistical advance came with the Wold decomposition. Wold (1938) showed that any zero-mean covariance-stationary process can be written uniquely as the sum of (i) a deterministic component that is perfectly predictable from the infinite past and (ii) a purely nondeterministic component that admits a one-sided infinite moving-average representation in orthogonal innovations. This result explains why autoregressive and moving average terms belong in time series models. Under broad conditions, a stationary process can be represented (or approximated) as a weighted sum of current and past shocks, which is the theoretical basis for later ARMA modeling.

The mathematical theory of stochastic processes was taking shape at the same time. Kolmogorov's work on the axiomatic foundations of probability and on the theory of stationary processes gave time series analysis a formal foundation (Kolmogoroff, 1933). His results on optimal linear prediction for stationary processes (Kolmogorov, 1941), developed alongside the wartime work of Wiener (1949), underpinned the filtering and forecasting methods that economics later adopted widely. Working in the frequency domain instead, Cramér (1942) proved the spectral representation theorem, showing that any stationary process can be decomposed into contributions at different frequencies. This gave economists a clean way to talk about cyclical behavior, though they did not apply it systematically to business-cycle measurement and band-pass filtering until decades later.

Researchers at the National Bureau of Economic Research (NBER) were also assembling long historical series on production, employment, and other aggregates, and developing descriptive methods for identifying expansions, recessions, and turning points. That research program began well before 1940 and culminated in *Measuring Business Cycles* (Burns & Mitchell, 1946). Although statistically informal, this empirical taxonomy set a benchmark chronology that later econometric models tried to explain and to forecast. The Burns–Mitchell tradition also encouraged the study of comovements across many series, foreshadowing the factor models formalized decades later.

Economists ended the 1930s with workable statistical representations of persistence, a sharper understanding of spurious correlation in trending data, an emerging stochastic view of cyclical fluctuations, and a general representation theorem for stationary series. No framework yet tied probability-based inference, identification, and dynamic specification together. That synthesis began in the 1940s.

3 The Rise of Econometric Modeling (1940s–1950s)

The 1930s closed with the first full-scale attempt to put business-cycle theory to statistical test. In work for the League of Nations, Tinbergen (1939) built multi-equation models of investment and of the U.S. economy, estimating behavioral equations for entire economies and asking whether competing cycle theories were consistent with the data. The effort provoked a famous methodological controversy. Keynes (1939) questioned whether regression methods could ever uncover reliable economic relations from nonexperimental, time-dependent data. He raised issues that anticipated much of the later econometric agenda, among them omitted factors, time-varying coefficients, and the direction of causality. The debate sharpened the case for explicit probabilistic foundations, and Frisch and Tinbergen would share the first Nobel Memorial Prize in Economic Sciences in 1969 for their pioneering work on dynamic macroeconomic modeling. Detailed accounts of this formative period are given by Morgan (1990) and Hendry and Morgan (1995).

The 1940s then redirected econometric methodology, in what is often called the 'probabilistic revolution'. In *The Probability Approach in Econometrics*, Haavelmo (1944) argued that empirical economic relations should be modeled as stochastic and analyzed with the tools of statistical inference. For the study of economic time series, this was more than a philosophical shift. It encouraged economists to treat observed macroeconomic histories as realizations of underlying stochastic processes rather than as unique deterministic trajectories. The disturbance term then became a substantive part of the model, capturing shocks and measurement error rather than serving as a residual nuisance. The credibility of inference came to depend on explicitly stated probabilistic assumptions. Haavelmo's contribution was recognized with the 1989 Nobel Memorial Prize.

Haavelmo's program was developed most systematically within the Cowles Commission. Building on the probability approach, Cowles researchers emphasized that estimating dynamic relationships in systems of simultaneous equations required careful attention to identification, that is, to which theoretical restrictions and exogeneity assumptions are needed to pin down structural parameters from time series data (Koopmans, 1950). The resulting structural tradition placed economic theory at the center of empirical work. Model builders specified behavioral equations and accounting identities, imposed exclusion restrictions, and fitted the resulting systems to macroeconomic time series with newly developed estimation machinery, first limited-information maximum likelihood (Anderson & Rubin, 1949) and later two-stage least squares (Theil, 1953; Basmann, 1957). The same theory-first stance clashed with the NBER's descriptive business-cycle program in the "measurement without theory" debate (Koopmans, 1947; Vining, 1949). That was an early round of a tension between theory-driven and data-driven empirical work that resurfaces throughout this history.

In the 1950s, the Cowles approach took concrete form in the first generation of large-scale macroeconometric models. A prominent example is Klein (1950), whose Cowles Commission work combined multi-equation structure with dynamic adjustment via lagged variables. Klein would receive the 1980 Nobel Memorial Prize for this modeling program. The Klein–Goldberger model of the U.S. economy (Klein & Goldberger, 1955) scaled this approach up and became the prototype for the large forecasting models of subsequent decades. One related response to dynamics in this period was the distributed-lag model, which offered a parsimonious representation of gradual adjustment over time (Koyck, 1954). These models were intended for both explanation and forecasting, yet by the standards of later time series work their treatment of stochastic dynamics was limited. Disturbances were usually assumed to be serially independent (or treated as such for estimation), and when residual autocorrelation was detected it was common to apply a correction rather than to revisit the dynamic specification. The best-known example is the transformation proposed by D. Cochrane and Orcutt (1949), alongside diagnostic tools such as the Durbin–Watson tests for serial correlation in regression residuals (Durbin & Watson, 1950, 1951). These developments foreshadowed a central lesson of time series econometrics: dynamic modeling, identification, and inference cannot be cleanly separated from the serial dependence properties of the data.

Meanwhile, a partly separate stream of work was advancing frequency-domain approaches to stochastic processes, drawing on statistics, signal processing, and electrical engineering. Wiener (1949) framed optimal linear prediction and filtering in spectral terms, while Whittle (1953) developed likelihood-based methods for stationary (including multivariate) time series that helped place spectral estimation and inference on firmer statistical foundations. Spectral tools offered a direct way to discuss business-cycle periodicities and noise reduction, but they spread into mainstream econometric practice more slowly than structural equation modeling. Later researchers drew on them nonetheless; Granger and Hatanaka (1964), for example, provided an early systematic treatment of spectral methods for economic time series, showing how the frequency domain could illuminate cyclical comovements and lead–lag relationships.

Two changes at the end of the 1950s pushed in the same direction. Policy institutions and businesses wanted more reliable forecasts, and digital computing was making iterative estimation feasible. Both favored a more data-oriented approach to time series modeling and forecasting, which Box and Jenkins (1970) would popularize in the following

decade. The postwar era delivered the probabilistic foundations and the first dynamic econometric models, and it exposed a tension between theory-driven structural specification and empirically driven time series modeling that would shape later work.

4 The Box–Jenkins Approach to Univariate ARIMA Modeling (1960s–1970s)

The publication of *Time Series Analysis: Forecasting and Control* by Box and Jenkins (1970) was a landmark in time series econometrics. Written from an engineering and applied-statistics perspective, the book codified a practical approach to univariate stochastic modeling and forecasting that quickly spread into economics and policy work. Its central premise was that observed series are noisy realizations of an underlying data-generating mechanism and that modeling should begin from serial dependence in the data rather than from a fully specified economic structure. The approach implements the Wold representation in practice: a covariance-stationary process admits a moving-average representation in innovations, which motivates approximating it with a parsimonious autoregressive–moving-average specification (Wold, 1938).

The core contribution was the family of autoregressive integrated moving-average (ARIMA) models and a systematic three-stage cycle of *identification*, *estimation*, and *diagnostic checking* (Box & Jenkins, 1970). Here 'identification' means selecting the model orders, a different usage from the structural identification of the Cowles tradition. In compact form, an ARIMA(p,d,q) model can be written as $\phi(L)(1-L)^d y_t = c + \theta(L)\varepsilon_t$, where L is the lag operator, $(1-L)^d$ differences the series to achieve approximate stationarity, and $\phi(\cdot)$ and $\theta(\cdot)$ collect the AR and MA dynamics. Seasonal extensions, often written as $(p,d,q)\times(P,D,Q)_s$, provided a compact way to represent periodic dynamics in economic series. Model identification relied heavily on the autocorrelation function (ACF) and the partial autocorrelation function (PACF): in low-order cases, the contrast between 'cut-off' and 'tailing-off' patterns helped distinguish autoregressive from moving-average structure. Estimation relied on (conditional) least squares or maximum likelihood, after which residuals were checked for remaining structure; a model was considered adequate only if the errors resembled white noise. Formal portmanteau checks such as the Q-statistic of Box and Pierce (1970) and its small-sample refinement by Ljung and Box (1978) became standard diagnostics. Parsimony was the complementary selection principle, later formalized as information criteria, most prominently Akaike's information criterion (AIC) (Akaike, 1974) and the Schwarz Bayesian information criterion (BIC) (Schwarz, 1978).

For economists, the appeal of Box–Jenkins was methodological as much as technical. The approach offered a coherent data-oriented alternative to the Cowles Commission tradition of theory-driven simultaneous equations. Rather than treating serial correlation as an inconvenience to be corrected after estimation, Box–Jenkins treated dynamic dependence as the very object to be modeled. This shift mattered particularly for forecasting, where the criterion of success is predictive performance rather than structural interpretation. C. R. Nelson (1972) then showed empirically that simple ARIMA models often out-predicted the large FRB–MIT–PENN structural model of the U.S. economy, a result that unsettled confidence in the prevailing approach. The insistence on iterative diagnostics helped change econometric practice. It encouraged routine checks for residual autocorrelation, explicit attention to stationarity and invertibility conditions, and a culture in which model adequacy was judged by whether additional serial structure remained in the errors. Box–Jenkins thus broadened time series modeling by supplying a procedure that worked without a full structural theory. It also raised the standard for empirical credibility, since stochastic assumptions and diagnostics now had to be stated.

Texts such as Granger and Newbold (1977) spread these ideas further into economics, presenting univariate ARIMA methods alongside forecasting theory and forecast evaluation, and helping make Box–Jenkins procedures routine in economic applications. The same modeling logic also pointed toward richer specifications that combined regression-type relationships with ARIMA error structure. Box and Jenkins discussed transfer-function (dynamic regression) models for relating an output series to one or more input series while accounting for serially correlated disturbances (Box & Jenkins, 1970). Closely related, Box and Tiao (1975) developed intervention analysis to quantify the dynamic effects of discrete events such as policy shifts, strikes, and regulatory changes, when the background evolution of the series follows an ARIMA process. These extensions let economists study policy-relevant questions with time series data without abandoning the diagnostics that made ARIMA models attractive in the first place. A further extension automated model selection itself. Algorithms such as that of Hyndman and Khandakar (2008) replace manual inspection of ACF and PACF plots with information criteria and unit-root pre-tests, choosing the differencing order and lag structure from the data. This made univariate forecasting practical at the scale of hundreds or thousands of series.

The Box–Jenkins approach also exposed what univariate time series methods could not deliver. ARIMA models

provide flexible reduced-form descriptions, but they are largely silent on the identification of structural parameters and on the causal interpretation of coefficients within an economic system. Moreover, treating each series in isolation leaves open the question of how shocks transmit across variables, a limitation that became more pressing as researchers moved toward multivariate descriptions of macroeconomic dynamics. Finally, the Box–Jenkins practice of differencing to achieve stationarity raised, without resolving, the question whether trends in macroeconomic data are deterministic or stochastic. Nonetheless, by the late 1970s the Box–Jenkins methods had become part of the econometric baseline, supplying a common language of ARMA dynamics, residual diagnostics, and forecasting benchmarks that shaped subsequent developments in multivariate modeling and structural interpretation. They also shaped official statistics, where seasonal ARIMA modeling became a core ingredient of seasonal adjustment and trend-cycle decomposition workflows, a theme taken up in Section 9.

5 Vector Autoregressions and the VAR vs. Structural Debate (1980s)

Dissatisfaction with the prevailing large-scale macroeconometric models had grown by the late 1970s. Their heavy theoretical structure invited criticism of the identifying restrictions needed for estimation; their forecasting record was mixed in an era marked by stagflation; and Lucas (1976) argued that relationships estimated under one policy regime need not survive a change in regime, undermining the models' use for policy evaluation. Univariate ARIMA models, for their part, offered reliable forecasting tools but were ill suited to questions that are inherently multivariate, e.g., how monetary policy shocks propagate jointly through output, inflation, and interest rates.

Sims (1980) responded with vector autoregressions (VARs), a deliberately modest statistical description of the joint dynamics of macroeconomic time series. Sims argued that many structural systems achieved identification only through restrictions he dismissed as literally "incredible". A reduced-form VAR instead treats the main variables as jointly endogenous and allows each to depend on its own lags and the lags of the others. In an N-variable VAR(p), $y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$, so dynamic interactions are learned from the data rather than imposed a priori, with contemporaneous dependence summarized by the covariance matrix of u_t.

This shift in emphasis also changed what economists reported and compared. Because a stable VAR admits a moving-average representation, it delivers impulse response functions (IRFs) that trace the dynamic effect of a one-time innovation through the entire system. It also delivers forecast error variance decompositions (FEVDs), which summarize how much each innovation contributes to forecast uncertainty at different horizons. IRFs and FEVDs gave economists a common language for describing 'how shocks propagate' in macro data. VARs also provided a testing ground for predictability-based causality: the notion that X 'Granger-causes' Y if past values of X improve forecasts of Y conditional on past Y (Granger, 1969). Granger-causality tests became one of the most widely applied concepts in empirical economics, used routinely to assess lead–lag relationships in finance, monetary economics, and international trade. Yet the concept captures predictive precedence rather than structural causation, a distinction Granger himself emphasized. Economists used these tools above all to re-examine the monetarist view that money has systematic effects on real activity (M. Friedman & Schwartz, 1963); the study of money, income, and causality by Sims (1972) became the best-known application of this approach.

The very features that made VARs attractive, however, also sharpened an identification problem. Reduced-form innovations are typically contemporaneously correlated, so interpreting a particular innovation as a 'monetary policy shock' or a 'supply shock' requires additional restrictions that map reduced-form residuals into orthogonal structural shocks. Early practice often relied on recursive (Cholesky) identification, which is simple to implement but can be sensitive to the assumed ordering of variables (Sims, 1980). The subsequent structural VAR (SVAR) literature tried to restore economic interpretation while retaining the empirical flexibility of VAR dynamics. A prominent example is the long-run restriction approach of Blanchard and Quah (1989), which separated demand and supply disturbances by assuming that one class of shocks has no permanent effect on output. An alternative route to identification imposed short-run restrictions through contemporaneous exclusion patterns, often motivated by institutional arguments about timing (e.g., that output and prices do not respond to a monetary policy shock within the same month). This preserves the VAR's data-driven dynamics while adding just enough theory to isolate specific shocks, and it became the workhorse of the empirical monetary policy literature (Christiano, Eichenbaum, & Evans, 1999).

Later identification strategies expanded this menu considerably. Sign restrictions constrain the direction of impulse responses without fixing their magnitude, providing set-identified rather than point-identified conclusions (Faust, 1998; Canova & De Nicoló, 2002; Uhlig, 2005). External-instrument ('proxy') approaches bring instruments constructed outside the model into the VAR (Mertens & Ravn, 2013), such as high-frequency movements around policy

announcements (Gertler & Karadi, 2015); see Stock and Watson (2018) for a survey. Narrative identification builds measures of exogenous policy shifts from historical records. These can enter dynamic regressions directly (Romer & Romer, 2004) or, under additional assumptions, serve as external instruments. Rigobon (2003) showed that changes in the variance of structural shocks across regimes can also provide identifying information, an approach termed 'identification through heteroskedasticity'. A comprehensive methodological treatment of structural VAR identification and inference is provided by Kilian and Lütkepohl (2017).

Alongside identification, applied work had to quantify the uncertainty around impulse responses. Because IRFs are nonlinear functions of estimated coefficients, delta-method approximations can perform poorly in small samples, and bootstrap methods became the standard tool for constructing confidence bands in applied VAR work, particularly the bias-corrected bootstrap developed by Kilian (1998). In Bayesian settings, posterior credible sets for IRFs come straight from the MCMC output, and reporting them is now routine.

Jordà (2005) introduced local projections (LPs), an important alternative to VAR-based impulse response analysis. Rather than estimating a full VAR system, LPs recover impulse responses by running direct regressions at each forecast horizon. This approach is less dependent on a particular finite-order VAR specification and extends readily to nonlinearities and state-dependent effects, for example by interacting shocks with business-cycle indicators to estimate how policy effects differ across recessions and expansions. The trade-off is that LPs estimate each horizon separately, so they can be less efficient than a correctly specified VAR and may produce less smooth impulse response paths. Subsequent work sharpened this comparison. Plagborg-Møller and Wolf (2021) showed that, with unrestricted lag structures, LPs and VARs estimate the same impulse responses in population, so the choice between them is a finite-sample bias–variance trade-off rather than a matter of principle. Large-scale simulations confirm that LPs have lower bias but substantially higher variance at intermediate and long horizons (Li, Plagborg-Møller, & Wolf, 2024). Jordà and Taylor (2025) provide a practitioner-oriented survey. In practice, many applied researchers report both VAR-based and LP-based impulse responses as a robustness check, treating agreement between the two as evidence that results are not driven by model-specific assumptions.

Practical limitations also shaped how VARs developed. Unrestricted VARs quickly become parameter rich as the number of variables and lags grows, raising the risk of overfitting in the relatively short postwar macro samples available to early users. A response, developed alongside the VAR movement, was to impose shrinkage through Bayesian priors, most famously the 'Minnesota prior' used in Bayesian VAR forecasting (Litterman, 1986). The VAR program therefore did not eliminate structural modeling so much as reframe the division of labor: reduced-form VARs provided empirical benchmarks and dynamic summaries, while theory was increasingly invoked to justify identification and to address the regime-change concerns raised by the Lucas critique. The parallel 'general-to-specific' tradition associated with the LSE also shared this commitment to statistical adequacy and dynamic specification, though it typically reintroduced economic structure through sequential simplification and theory-based long-run relationships (Hendry, 1995).

VARs had become central tools in empirical macroeconomics and policy institutions by the end of the 1980s. They were valued for their forecasting performance, competitive with large structural systems in direct comparisons, and for the interpretive power of IRFs and related decompositions (for a comprehensive survey, see Stock & Watson, 2001). Sims shared the 2011 Nobel Memorial Prize with Thomas Sargent for this line of empirical research on cause and effect in the macroeconomy. The decade also made clear that multivariate dynamics could not be separated from the nonstationary features of many macroeconomic series.

6 Dealing with Unit Roots and Spurious Regressions (1970s–1980s)

In the late 1970s and 1980s a major challenge became how to do credible inference with economic time series that exhibit strong persistence, pronounced trends, and slow mean reversion. In such settings, classical regression logic built on covariance stationarity can fail. Standard t-tests and R^2 values may look persuasive even when relationships are illusory.

The canonical warning example is spurious regression. Using simulations and empirical illustrations, Granger and Newbold (1974) showed that regressing one random walk on another can produce a high R^2 and apparently significant coefficients even when the series are independent. This modern result echoes the earlier concern about "nonsense correlations" in trended series (Yule, 1926). Two independent series that each carry a trend, whether deterministic or stochastic, tend to appear statistically related. What drives the spurious association is the presence of unrelated trends, not the unit root specifically.

Unit roots required rethinking standard tools. A series with a unit root has a variance that grows without bound over

time, and a regression of one such series on another typically produces nonstationary residuals unless the variables share a common stochastic trend, the case later formalized as cointegration. When the trends are unrelated, Phillips (1986) showed that regressing an $I(1)$ series on another independent $I(1)$ series yields an R^2 that does not vanish asymptotically and t-statistics that diverge, often indicating significance when none exists. Beyond the spurious-regression case, both estimation and inference can have nonstandard limiting behavior under integration, so classical hypothesis tests can over-reject dramatically (Phillips, 1987a).

These problems focused attention on *unit roots* and the concept of integration. In autoregressive terms, a unit root corresponds to a characteristic root equal to one, implying that shocks accumulate over time and the process is nonstationary. A series is integrated of order one, $I(1)$, if its first difference is $I(0)$ while its level is not; higher-order integration requires further differencing. Formal testing procedures emerged rapidly. Dickey and Fuller (1979) introduced the Dickey–Fuller framework, emphasizing that the test statistic has a nonstandard distribution under the unit-root null. Because economic series typically have richer short-run dynamics than an AR(1), the augmented Dickey–Fuller (ADF) approach became standard, adding lagged differences to control residual serial correlation (Said & Dickey, 1984). Related procedures relaxed parametric assumptions about the short-run dynamics by using nonparametric corrections, as in the Phillips–Perron tests (Phillips & Perron, 1988). Subsequent work focused on improving the size and power properties of unit-root tests (Elliott, Rothenberg, & Stock, 1996; Ng & Perron, 2001). The unit-root framework was also extended to seasonal frequencies by Hylleberg, Engle, Granger, and Yoo (1990).

In applied work, conclusions depend on practical specification choices, notably whether to include an intercept or deterministic trend and how many lags to add. A key limitation is that unit-root tests can have low power against highly persistent but stationary alternatives, so-called near-unit-root behavior (Schwert, 1989). This limitation encouraged diagnostics that swap the null and alternative hypotheses, most notably the KPSS test, which takes stationarity as the null and a unit root as the alternative (Kwiatkowski, Phillips, Schmidt, & Shin, 1992). Phillips (1987b) supplied the deeper theoretical basis for the power problem, developing the local-to-unity asymptotic framework, in which the autoregressive coefficient is parameterized as $\rho = 1 + c/T$ where c is fixed and T is sample size. The framework explains why conventional tests struggle to distinguish exact unit roots from highly persistent processes, and it guided the design of tests with better power against such alternatives (Elliott, 1999). Nonrejection of a unit root is therefore not definitive evidence of a random walk. Economic theory, long-term behavior, and stationarity tests should all inform the modeling decision.

C. R. Nelson and Plosser (1982) gave these issues empirical weight, examining long U.S. macroeconomic time series and finding that the unit-root null could not be rejected for many aggregates. This evidence helped shift practice toward modeling growth rates and differences rather than levels: if levels are plausibly $I(1)$, working with changes removes the stochastic trends and reduces the risk of spurious findings. The precaution became routine enough that applied work often defaulted to differencing even when long-run relations were plausible and many inferential targets did not require it (Sims, Stock, & Watson, 1990). This practice arose in part because levels regressions were hard to interpret safely without an explicit cointegration framework. However, differencing discards low-frequency information, and economists often care precisely about long-run co-movement in levels.

A related modeling response was to include deterministic trends when variables appear trend-stationary. If a series is stationary around a deterministic trend (i.e., $I(0)$ after removing a linear trend), then detrending or including t in the regression can absorb the deterministic component. This choice fed into a central debate: are macroeconomic series better treated as difference-stationary (unit-root processes with drift) or as trend-stationary (fluctuations around a deterministic growth path)? The distinction matters. For example, if output is trend-stationary, recessions are deviations below a stable trend and output eventually returns to that path; if output is difference-stationary, large downturns can permanently lower the level of output, implying persistent 'scars' rather than temporary 'gaps'. Measuring the size of the permanent component became an empirical literature of its own (Campbell & Mankiw, 1987; J. H. Cochrane, 1988).

The unit-root debate also drew renewed attention to structural change. Apparent nonstationarity may reflect regime shifts or rare breaks rather than a genuine unit root. Perron (1989) showed that failing to model a one-time break in level or trend can bias standard unit-root tests toward nonrejection. Later work developed tests that allow for a break date determined from the data rather than fixed a priori (Zivot & Andrews, 1992). Empirical macroeconomics closed the 1980s having absorbed the lesson that trends, persistence, and breaks are central features of time series, and that credible long-run inference requires tools designed for nonstationary data.

The unit-root framework also prompted work on long memory and fractional integration, which sits between stationary $I(0)$ dynamics and exact unit roots. In this approach, fractional differencing captures persistence, so that shocks decay slowly and autocorrelations decline hyperbolically rather than geometrically. Granger and Joyeux (1980) and Hosking (1981) developed fractionally integrated ARIMA (ARFIMA) models that formalize this idea, and

subsequent work proposed semiparametric estimation and testing procedures based on low-frequency behavior of the spectrum, including the log-periodogram estimator of Geweke and Porter-Hudak (1983) and refinements surveyed by Baillie (1996). Robinson (1995) developed semiparametric estimation and inference for long-memory processes. Long-memory ideas also found application in volatility modeling, particularly to capture the slow autocorrelation decay observed in squared returns (Baillie, Bollerslev, & Mikkelsen, 1996). In applied macro-finance work, fractional integration remained a useful reminder that persistence may be richer than the simple $I(0)/I(1)$ dichotomy, even though practical model choice often continued to hinge on unit-root and break diagnostics.

Unit-root testing was later extended in the opposite direction as well. Conventional unit-root tests are one-sided against stationary alternatives, with the autoregressive root below one, but during asset-price booms the relevant alternative is mildly explosive behavior, with a root above one. Prices then grow faster than a random walk with drift would imply. Phillips, Wu, and Yu (2011) developed right-tailed tests computed over recursively expanding samples, so that explosive episodes emerging partway through the sample can be detected, and Phillips, Shi, and Yu (2015) generalized the procedure to accommodate multiple episodes of exuberance and collapse and to estimate their origination and termination dates. These procedures gave empirical content to the notion of a speculative bubble and are used in real-time monitoring of housing and financial markets.

As unit-root testing became routine in univariate settings, researchers recognized the need to extend these methods to panel data. Panel unit-root tests pool information across cross-sectional units, improving power relative to separate unit-root tests and allowing researchers to characterize persistence while accommodating heterogeneity. Levin, Lin, and Chu (2002) developed the Levin–Lin–Chu test, which assumes a common autoregressive coefficient across units. Im, Pesaran, and Shin (2003) relaxed this assumption, proposing a test that permits unit-specific autoregressive dynamics. A distinctive concern in panel settings is cross-sectional dependence, which can invalidate procedures that treat units as independent and can itself be an object of interest, reflecting common shocks or spillovers. Pesaran (2007) addressed this by augmenting unit-level Dickey–Fuller regressions with cross-sectional averages to control for common factors and pooling the resulting statistics into a panel (CIPS) test, providing a practical way to test for unit roots in panels with pervasive dependence.

A parallel development addressed the estimation of dynamic panel models with short time dimensions and persistent dependent variables. GMM estimators proposed by Arellano and Bond (1991) and extended by Arellano and Bover (1995) and Blundell and Bond (1998) provided consistent estimation of autoregressive panel models by using lagged levels or differences as instruments, and became standard tools in applied microeconometrics and growth empirics where time-series dependence at the unit level matters.

These tools became especially important as researchers worked with international macro and firm-level panels, many of which display strong common components akin to the factor structures discussed in Section 11. The differencing dilemma had a constructive resolution, found in the same decade: some nonstationary series move together in the long run, so that certain linear combinations are stationary. Engle and Granger (1987) formalized the idea as cointegration.

7 Long-Run Equilibria in Dynamic Models (1980s–1990s)

Cointegration offers a middle way between spurious levels regressions and differencing that throws information away: variables may be individually $I(1)$, but if some linear combination is $I(0)$, the variables share a common stochastic trend and the combination represents a stable long-run restriction. Long-run relationships can then be studied in levels without treating trending behavior as an econometric nuisance.

The economic content of cointegration is that theory often constrains *spreads*, *ratios*, or *accounting identities* to remain bounded over long horizons. A familiar illustration is consumption and income. Each aggregate may contain a stochastic trend, yet standard models suggest that their long-run linkage is stable. Cointegration makes this idea testable. If C_t and Y_t are $I(1)$ but $C_t - \beta Y_t$ is stationary for some β, then deviations from the long-run relation are transitory rather than drifting without bound. The levels regression is therefore not spurious in the sense highlighted by Granger and Newbold (1974); instead, the residual has an interpretable role as an equilibrium error.

The error-correction model (ECM) at the heart of this approach was not itself new. Its single-equation form originated with Sargan (1964), in a study of UK wages and prices often regarded as the first application of what became the LSE approach to dynamic econometric modeling. The consumption-function study of Davidson, Hendry, Srba, and Yeo (1978) popularized it, and Hendry (1995) later set it out as part of a general modeling strategy. A central contribution of Engle and Granger (1987) was to tie this representation directly to nonstationarity. Their Granger Representation Theorem implies that cointegrated variables admit an error-correction model in which short-run changes respond to

lagged disequilibrium. In a simple bivariate case, $\Delta y_t = \alpha\,(y_{t-1} - \beta x_{t-1}) + \sum_{j=1}^{p} \gamma_j\,\Delta y_{t-j} + \sum_{j=0}^{p} \delta_j\,\Delta x_{t-j} + \varepsilon_t$, where $(y_{t-1} - \beta x_{t-1})$ is the lagged equilibrium error and α measures the speed of adjustment. The ECM reconciles two objectives that differencing alone cannot: it yields a stationary specification for inference while retaining the long-run information encoded in the levels relation. Economically, α quantifies how quickly deviations from equilibrium feed back into subsequent changes.

For empirical work, Engle and Granger (1987) proposed a practical two-step approach: estimate the long-run relation in levels, test whether the residual is stationary, and, if so, include the lagged residual as the error-correction term in a dynamic model for differences. Residual-based tests and associated asymptotic theory were developed further, including procedures that remain valid under general short-run dynamics (Phillips & Ouliaris, 1990). A key practical implication is that while cointegrating coefficients can be estimated very precisely (i.e., super-consistently) in large samples (Stock, 1987), conventional t- and F-approximations do not automatically carry over, so inference requires cointegration-specific asymptotics or corrected procedures.

These concerns motivated estimators tailored to reliable inference in cointegrating regressions. Phillips and Hansen (1990) proposed fully modified OLS (FM–OLS), which corrects for endogeneity and serial correlation using nonparametric long-run variance estimates. Stock and Watson (1993) developed dynamic OLS (DOLS), which augments the cointegrating regression with leads and lags of first differences to improve finite-sample performance and deliver efficient inference under common conditions. Both methods helped standardize long-run estimation in cointegrated settings where feedback and serial dependence matter.

A second major strand treated cointegration as a system property rather than a single-equation restriction. Johansen (1988) embedded cointegration in a vector autoregression and showed that the number of cointegrating relations equals the rank of the long-run coefficient matrix in the system's error-correction form. In a k-variable setting, a VAR in levels can be written as a vector error-correction model (VECM), $\Delta y_t = \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j\,\Delta y_{t-j} + \varepsilon_t$, where $\Pi = \alpha\beta'$ is the long-run coefficient matrix and the rank of Π equals the number of linearly independent cointegrating vectors. Equivalently, k variables with r cointegrating relations are driven by $k - r$ common stochastic trends (Stock & Watson, 1988). The Johansen framework delivers likelihood-based tests of the cointegration rank and enables joint estimation and hypothesis testing on long-run relations and adjustment dynamics (Johansen, 1991, 1995). System methods proved especially valuable in macroeconomic applications where multiple variables share trends and more than one equilibrium restriction is plausible.

With these tools in place, empirical applications proliferated in the late 1980s and 1990s. Early work often used bivariate or small-system settings, testing purchasing power parity via exchange rates and relative prices, examining term-structure and other present-value restrictions (Campbell & Shiller, 1987), or assessing fiscal sustainability by relating revenues and expenditures. As system methods matured, researchers turned to multivariate VECMs, which allow several common trends and multiple long-run restrictions at once. Examples include open-economy systems combining exchange rates, domestic and foreign prices, and interest differentials, and monetary systems in which money, prices, income, and interest rates may support more than one cointegrating relation (Johansen, 1995).

Cointegration also deepened the interplay between economic theory and time series analysis. Whereas reduced-form VARs were often presented as deliberately modest statistical descriptions (Sims, 1980), a cointegrated VAR (or VECM) invites structural interpretation. The cointegrating combinations $\beta' y_t$ can be read as long-run equilibrium conditions or budget constraints, while the adjustment coefficients α indicate how each variable responds to disequilibrium and which variables do the 'adjusting' versus the 'driving' of long-run trends. A close complement was the concept of common cycles. Just as cointegration implies linear combinations that remove the common stochastic trends, Vahid and Engle (1993) showed that cointegrated variables may also share common cyclical features, so that a separate set of combinations removes the transitory dynamics. When the cointegrating and cofeature vectors together span the system, the two sets deliver a unique decomposition into common trends and common cycles.

An important extension dispensed with prior knowledge of the integration order of the regressors. Pesaran, Shin, and Smith (2001) developed the Autoregressive Distributed Lag (ARDL) bounds testing approach, which tests for long-run relationships without pretesting whether variables are $I(0)$, $I(1)$, or mutually cointegrated. The approach reformulates the model as an ARDL specification and derives bounds on a test statistic so that inference is valid whether the regressors are $I(0)$ or $I(1)$, though not for $I(2)$ or higher. This flexibility made the approach widely popular across macroeconomics, finance, and applied microeconomics, especially where integration status is ambiguous or sample sizes make pretesting unreliable. A statistic falling between the critical bounds leaves the result inconclusive.

Cointegration methods were also extended to panel settings. Pedroni (1999, 2004) developed panel cointegration tests for long-run relationships across cross-sectional units such as countries or regions. These tests pool information across units while allowing heterogeneous cointegrating slopes and short-run dynamics, and they became particularly

important in growth empirics and international macroeconomics.

Error correction need not be linear either. Balke and Fomby (1997) and others developed threshold cointegration models that allow adjustment toward a stable long-run relationship to vary with the size of the equilibrium error. Transaction costs or bands of inaction motivate such behavior economically: small deviations are tolerated and large ones trigger strong correction.

Although cointegration methods were adopted rapidly, the literature also emphasized practical caveats. Results can be sensitive to deterministic components (intercepts and trends) and lag-length choices, and structural change can shift the long-run relation in ways that standard tests may miss. Allowing for regime shifts in the cointegrating vector can materially change empirical conclusions, motivating tests that accommodate breaks in the long-run relation (Gregory & Hansen, 1996). Specification matters as well. A levels regression that omits a relevant cointegrating variable will generally fail to be cointegrating, since its residual remains $I(1)$, even when the included variables are genuinely related. Such a regression is therefore misspecified but not spurious (Banerjee, Dolado, Galbraith, & Hendry, 1993; Martin, Hurn, & Harris, 2013). Cointegration and error-correction modeling became standard equipment for nonstationary data during the 1990s, providing a coherent way to combine short-run dynamics with long-run equilibrium restrictions. The 2003 Nobel Memorial Prize recognized the impact of these ideas on empirical practice; Granger shared it for cointegration and Engle for volatility models.

The rapid spread of these dynamic-modeling methods into routine practice owed much to dedicated time series software. From the 1980s onward, packages such as TSP, RATS, and EViews in North America, and GIVE/PcGive, STAMP, and Microfit in the United Kingdom, handled dates, lags and leads, and time series plots far more gracefully than general-purpose statistical systems. They made visual inspection of the data and routine diagnostic testing a normal part of dynamic modeling in central banks and other institutions. These programs were built around the classical methods of their era: ARIMA, VAR, and unit-root and cointegration analysis. They translated academic advances into accessible procedures.

8 Modeling Volatility and Nonlinear Dynamics (1980s–1990s)

Up to the early 1980s, much applied time series econometrics focused on modeling the conditional mean under linear dynamics and approximately homoskedastic disturbances. Yet macroeconomic and financial time series often show patterns that fit uneasily with this baseline. Volatility tends to cluster, extreme outcomes matter for risk and policy, and the propagation of shocks can depend on the state of the economy. Two strands of work in the 1980s and 1990s responded to these observations, one modeling time-varying conditional variance, the other bringing nonlinear dynamics and regime dependence into routine empirical modeling.

The canonical starting point for modeling volatility is the autoregressive conditional heteroskedasticity (ARCH) framework of Engle (1982). ARCH formalizes the idea that volatility can be predictable even when mean shocks are not. In an ARCH(q) specification, the conditional variance depends on lagged squared innovations, $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2$, so recent large shocks raise current conditional variance. This structure captures volatility clustering and yields forecast intervals whose width adapts to prevailing uncertainty; constant-variance models, by contrast, understate uncertainty in turbulent periods and overstate it in calm ones, even if the conditional mean is specified sensibly.

Bollerslev (1986) generalized ARCH by allowing lagged conditional variance to enter the variance equation. The parsimonious GARCH(1,1), $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, often fits asset-return data well with few parameters because it permits volatility shocks to decay slowly. This parsimony helped make conditional variance models practical for forecasting and risk applications where the goal is to predict uncertainty as well as means. Supporting inference theory advanced quickly. Quasi-maximum likelihood arguments explained why Gaussian likelihood-based estimation can deliver consistent parameter estimates under weak conditions and why robust standard errors can be used when conditional distributions are misspecified (Bollerslev & Wooldridge, 1992). Because financial and macroeconomic series frequently have heavier tails than the Gaussian distribution implies, Student-t and other heavy-tailed innovations improved density forecasts and risk measures, and robust estimation methods offered further protection against outliers (Huber, 1964).

A large literature then extended ARCH and GARCH in directions suggested by the data. One theme was asymmetry. Negative return shocks often increase volatility more than equally sized positive shocks, a pattern consistent with leverage effects and feedback mechanisms in financial markets. Models such as EGARCH (D. B. Nelson, 1991) and the GJR specification (Glosten, Jagannathan, & Runkle, 1993) incorporate this behavior directly into the conditional

variance equation. Another theme was multivariate volatility, motivated by time-varying comovement and portfolio risk. Multivariate GARCH formulations allowed researchers to model conditional covariances jointly with conditional variances (Engle & Kroner, 1995), initially under constant conditional correlation (Bollerslev, 1990) and subsequently through dynamic conditional correlation specifications that let comovements evolve through time (Engle, 2002). Although these models were most influential in finance, they also informed work in macroeconomics where time-varying uncertainty and comovement matter.

Extending the GARCH framework to capture long-memory persistence, Baillie et al. (1996) introduced fractionally integrated GARCH (FIGARCH). FIGARCH models capture persistent volatility through hyperbolically decaying weights on past squared innovations. Related specifications such as IGARCH allow volatility shocks to have permanent effects. These long-memory volatility models helped reconcile high-frequency return behavior with theoretical models of long-run risk.

An alternative paradigm models volatility as an unobserved latent process rather than as a deterministic function of past shocks. In stochastic volatility (SV) models, the log conditional variance follows its own stochastic law of motion, typically a first-order autoregression driven by innovations that are distinct from the mean-equation shocks (Taylor, 1986). This parameter-driven structure contrasts with the observation-driven GARCH family. Because the variance is latent, SV models require simulation-based or state-space estimation methods rather than straightforward maximum likelihood, a point that long limited their uptake. The development of efficient MCMC samplers removed this barrier, making Bayesian SV estimation practical (Jacquier, Polson, & Rossi, 1994; S. Kim et al., 1998). SV specifications have since become standard components of macroeconomic models with time-varying volatility, most notably in the TVP-VAR framework of Primiceri (2005), where allowing shock variances to evolve stochastically proved important for capturing the changing volatility of the postwar U.S. economy. The observation-driven versus parameter-driven distinction remains a useful way to organize this literature: GARCH is computationally convenient and dominates financial risk applications, whereas SV integrates cleanly into Bayesian state-space systems and is now the usual choice in structural macroeconometrics.

A related advance was the development of realized volatility measures, which exploit high-frequency intraday data to construct nonparametric estimates of integrated variance. Andersen, Bollerslev, Diebold, and Labys (2003) showed that realized volatility, computed as the sum of squared intraday returns, provides a directly observable proxy for the latent variance that SV models seek to estimate. Barndorff-Nielsen and Shephard (2002) established the econometric theory linking realized measures to the underlying continuous-time volatility process and showed how they can be used to estimate SV model parameters. Realized measures also supported simple forecasting models of their own, most notably the heterogeneous autoregression of Corsi (2009), which became a standard tool for volatility prediction. By making volatility (approximately) observable, realized measures transformed evaluation: parametric volatility models, whether GARCH or SV, could now be compared against a high-frequency benchmark rather than solely against each other.

Beyond the ARCH/GARCH family, score-driven (or Generalized Autoregressive Score, GAS) dynamics unified many time-varying parameter models. Creal, Koopman, and Lucas (2013) let model parameters evolve through time based on the scaled score of the likelihood function. GAS encompasses GARCH and other observation-driven models as special cases, and it supplies a common set of inference tools and simulation-based methods for forecasting.

Econometricians also developed nonlinear time series models to capture state dependence in economic dynamics. Linear AR and VAR specifications are often useful first approximations, but they impose symmetric and time-invariant propagation, so the response to a shock is the same whether the economy is near a constraint, in a recession, or in a high-volatility period. Nonlinear models relax these restrictions, allowing persistence and mean reversion to vary with the level of the series or with other state variables.

A leading example is threshold autoregression. Building on earlier work, Tong (1983) formalized threshold autoregressive (TAR) models in which parameters depend on whether an observed transition variable crosses a threshold. In a simple two-regime TAR, the process follows one autoregression when $y_{t-d} \leq \tau$ and another when $y_{t-d} > \tau$, permitting sharp changes in persistence or mean reversion across regimes. Such models offered an interpretable way to represent nonlinear adjustment, for instance when behavior differs in booms and busts. Related smooth-transition autoregressive models allow parameters to change gradually rather than discontinuously, which can be useful when dynamics evolve smoothly as conditions change (Granger & Teräsvirta, 1993; Teräsvirta, 1994).

Another prominent approach treated regimes as latent rather than defined by an observed threshold. Hamilton (1989) introduced Markov-switching models for U.S. output growth, in which the economy alternates between unobserved states governed by a Markov transition matrix. This framework captures the clustering of downturns and provides an econometric mechanism for dating business-cycle turning points using the estimated probability of being in each regime. Beyond business-cycle dating, regime-switching models offered a way to represent recurrent changes in mean growth

and volatility without fixing break dates in advance, and they became a common tool for studying nonlinear propagation and time variation in macroeconomic dynamics.

A complementary strand focused on structural breaks and parameter stability. While nonlinear models allow dynamics to change with the state, break tests formalize the possibility that coefficients shift at unknown dates. The classical Chow test addresses a known break date (Chow, 1960), while the Quandt likelihood ratio test enabled testing over a range of candidate break dates (Quandt, 1960). Later contributions provided asymptotically justified procedures for breaks at unknown dates (Andrews, 1993) and for multiple breaks (J. Bai & Perron, 1998). Such tools were central to empirical debates about stability in macroeconomics and finance, including evidence of reduced macroeconomic volatility in the mid-1980s that is often associated with the 'Great Moderation' (McConnell & Pérez-Quirós, 2000). In practice, break testing, threshold models, and regime switching were often used together, depending on whether the data suggested a one-time shift, a small number of recurring regimes, or state-dependent adjustment.

Quantile regression (Koenker & Bassett, 1978), extended to time series settings (Koenker & Xiao, 2006), offered a further departure from classical conditional-mean modeling. Rather than modeling the conditional variance parametrically, quantile regression estimates the conditional quantile function directly, providing a flexible way to characterize the entire conditional distribution without strong distributional assumptions. In financial risk management, Value-at-Risk had become the benchmark measure of tail risk (Jorion, 2007). Quantile regression found a natural application in the conditional autoregressive Value-at-Risk (CAViaR) framework, which models tail quantiles as functions of past returns and past quantile estimates, a semiparametric alternative to GARCH-based Value-at-Risk (Engle & Manganelli, 2004). Beyond risk management, quantile regression methods have been applied to study how macroeconomic and financial predictors affect the tails of growth and return distributions differently from the center, an approach that has gained renewed attention in the 'growth-at-risk' literature (Adrian, Boyarchenko, & Giannone, 2019).

Volatility models and nonlinear dynamics broadened the scope of time series econometrics, making conditional second moments routine inputs to empirical work and providing structured departures from linear propagation. They also increased the demand for computationally intensive estimation and for probabilistic representations of latent states. Once volatility, regimes, and other forms of instability were treated as time-varying objects, it made sense to represent them through unobserved components and to estimate them with filtering and smoothing methods. This signal extraction perspective had long been embedded in practical work such as seasonal adjustment and business-cycle measurement, but state-space representations and the Kalman filter tied it much more closely to econometric modeling.

9 Filtering, Seasonal Adjustment, and Signal Extraction (1960s–2000s)

In the 1960s, while some econometricians were developing stochastic models for forecasting and inference, others took up the problem of separating 'signal' from 'noise'. In macroeconomics and official statistics, this meant seasonal adjustment, trend–cycle decomposition, and related filtering procedures that aim to extract interpretable components from noisy and persistent data. These methods answer a practical need. Many economic decisions require real-time assessment of underlying conditions (trend growth, cyclical slack, inflation momentum) rather than raw observations that are dominated by seasonal patterns, measurement error, and short-lived disturbances.

Seasonal adjustment provides a leading example. The U.S. Census Bureau's X-11 program systematized moving-average based seasonal adjustment procedures and quickly became influential in statistical agencies (Shiskin, Young, & Musgrave, 1967). Subsequent developments integrated seasonal ARIMA models into adjustment workflows, as in X-12-ARIMA (Findley, Monsell, Bell, Otto, & Chen, 1998), reflecting the broader diffusion of Box–Jenkins style dynamics into empirical practice. Further refinements led to X-13ARIMA-SEATS, now the standard seasonal adjustment tool at many statistical agencies. X-13ARIMA-SEATS combines the signal-extraction principles of both the X-11 tradition and the model-based SEATS (Signal Extraction in ARIMA Time Series) approach (Gómez & Maravall, 1996), allowing flexible decomposition of series into seasonal, trend, and irregular components through automatic or user-specified ARIMA models. Related decomposition methods, including STL, provided flexible ways to represent seasonality and trend components (Cleveland, Cleveland, McRae, & Terpenning, 1990). The component language of these programs (trend, seasonal, irregular) descends directly from the interwar business barometers encountered in Section 2.

A closely related applied tradition is exponential smoothing, a family of adaptive filters designed for routine forecasting where transparency and computational simplicity matter. Early formulations were developed for inventory and production planning (Brown, 1959) and extended to accommodate trends and seasonality by Holt (1957) and Winters (1960). Although exponential smoothing was long treated as an effective algorithmic recipe, Muth (1960)

showed early on that simple exponential smoothing is optimal for a random-walk-plus-noise process. Later work generalized the point by showing that widely used smoothing schemes correspond to particular innovations state-space models. In this formulation, exponential smoothing becomes a model-based signal extraction method with an explicit likelihood, prediction intervals, and systematic model selection (Ord, Koehler, & Snyder, 1997; Hyndman, Koehler, Snyder, & Grose, 2002). This connection further blurred the line between 'filtering' and 'modeling' by showing that even seemingly simple smoothing rules embed substantive assumptions about component dynamics and forecast uncertainty.

Trend–cycle decomposition in macroeconomics developed along similar lines. The Beveridge–Nelson decomposition interprets the permanent component of a series through its stochastic trend representation (S. Beveridge & Nelson, 1981). The Hodrick–Prescott filter, circulated in working-paper form since 1981 (Hodrick & Prescott, 1997), became the standard pragmatic trend extractor. It has also drawn forceful criticism, most prominently from Hamilton (2018), who points to end-point bias, induced dynamics, and the risk of attributing economic meaning to artifacts of the filter. Band-pass filters, such as those of Baxter and King (1999) and Christiano and Fitzgerald (2003), formalized the notion of isolating fluctuations at business-cycle frequencies and remain popular because they are simple and deliver readily interpretable objects. Filtering choices matter, and they belong to the modeling strategy rather than to harmless preprocessing.

How do these decomposition approaches relate to one another? The Beveridge–Nelson trend represents the permanent component as the long-run forecast of the series once accumulated drift is subtracted, while unobserved-components (UC) models treat trends and cycles as explicit latent components evolving according to specified stochastic laws of motion. These approaches can yield quite different trend estimates. Morley, Nelson, and Zivot (2003) resolved the apparent tension by showing that when the correlation between trend and cycle shocks is left unrestricted, the UC model and the Beveridge–Nelson decomposition deliver equivalent permanent-transitory decompositions. This insight reconciled long-standing debates about which decomposition is more appropriate and showed that specification choices, particularly restrictions on shock correlations, determine the trends and cycles extracted from the data.

Model-based signal extraction offers an alternative to ad hoc filters, using state-space representations with the Kalman filter as the core computational apparatus. The Kalman filter is a recursive algorithm for optimal updating in linear Gaussian state-space systems (Kalman, 1960). It earns its place in economics because so many empirical problems can be expressed in measurement-and-transition form: observed data are linked to a vector of latent states through a measurement equation, while the states follow a stochastic law of motion. Once a model is written in this form, filtering and smoothing deliver estimates of the unobserved components together with measures of uncertainty, so that the decomposition is not imposed externally but implied by an econometric model. Durbin and Koopman (2001) provide a systematic treatment. Several widely used filters can then be interpreted as special cases of unobserved-components state-space models. The Hodrick–Prescott trend, for example, corresponds to the smoothed estimate from a local-linear-trend model under particular assumptions about the relative variances of trend and irregular shocks (A. C. Harvey & Jaeger, 1993), making explicit that 'filtering' embeds a probabilistic model for the components.

The structural time series approach popularized by A. C. Harvey (1989) made these ideas accessible to applied users. In a basic unobserved-components model, a series is written as the sum of a stochastic trend, a cycle, a seasonal component, and an irregular term, each governed by its own evolution equation. The Kalman filter and smoother then produce filtered and smoothed estimates of the latent components together with prediction and smoothing variances. This framework proved useful for tasks such as trend-cycle decomposition, real-time output gap estimation, and handling missing observations or irregular sampling, all of which arise frequently in macroeconomic monitoring and policy work. Real-time use also exposed the limits of signal extraction, as Orphanides and van Norden (2002) showed that output-gap estimates are heavily revised as new data arrive, echoing the end-point concerns raised about mechanical filters.

A closely related extension combines state-space structure with discrete regime changes. In these models, some parameters (and often shock variances) depend on an unobserved state that evolves as a Markov chain, so the data are generated by a small number of recurring regimes. C.-J. Kim (1994) extended Hamilton-style Markov switching to general state-space systems and developed filtering and smoothing methods that make likelihood-based estimation feasible in practice. C.-J. Kim and Nelson (1999b) provided a systematic treatment of regime-switching state-space models under both classical likelihood approximations and Gibbs-sampling approaches, with applications that study changes in macroeconomic dynamics and stability (C.-J. Kim & Nelson, 1999a).

State-space methods also supported time-varying parameter (TVP) models, in which coefficients evolve as latent processes rather than remaining fixed. A common specification treats a coefficient vector as drifting slowly, for instance $\beta_t = \beta_{t-1} + \eta_t$, which captures gradual changes in relationships arising from evolving policy regimes, changes in the private sector's behavior, or persistent shifts in volatility. TVP specifications became a natural way to represent concerns

about structural instability and regime change in macroeconomic data. A landmark application combined TVP dynamics with stochastic volatility in a Bayesian framework. Primiceri (2005) estimated a time-varying parameter VAR with stochastic volatility using MCMC methods, allowing both the transmission mechanisms and the shock volatilities to evolve over time. This framework became a mainstay of macroeconomic research for decomposing historical changes in propagation mechanisms and shock importance. More generally, empirical work using TVP-VARs often found evidence of time variation in shock propagation and in the volatility of innovations over the postwar period (Cogley & Sargent, 2005), reinforcing the case for models that can adapt as the economy evolves. A prominent univariate counterpart is the unobserved-components model with stochastic volatility that Stock and Watson (2007) used to characterize trend inflation, which became a benchmark in inflation forecasting.

Filtering methods were also extended to settings where the linear-Gaussian assumptions of the Kalman filter are inappropriate. Sequential Monte Carlo methods, often called particle filters, approximate filtering distributions in nonlinear or non-Gaussian state-space models by propagating weighted samples through time (Doucet, de Freitas, & Gordon, 2001). Particle methods broadened the set of models that could be brought to data and provided practical likelihood approximations for models with discrete regime switches, nonlinear measurement equations, or heavy-tailed shocks. Although computationally intensive, they further strengthened the idea that latent-state representations coupled with simulation could handle empirically realistic dynamics.

Seasonal adjustment, trend–cycle decomposition, and state-space signal extraction became central to applied time series work in macroeconomic monitoring and official statistics. They also supplied a practical language of 'real-time' assessment and component interpretation that carried over to mixed-frequency forecasting and nowcasting.

10 Bayesian Methods in Time Series Econometrics (1990s–2000s)

By the 1990s, applied time series econometrics had accumulated models that were attractive in principle but costly to estimate. VARs were flexible but parameter-heavy, and many nonlinear specifications were not amenable to standard likelihood calculations or conventional asymptotics. Bayesian inference itself was hardly new to econometrics, since Zellner (1971) had championed it decades earlier, but its reach had been limited by computation. That constraint eased as improvements in processing power and numerical methods made intensive optimization and simulation feasible in routine empirical work. These developments produced a practical Bayesian reorientation in econometrics. Shrinkage priors made large dynamic systems estimable by regularizing high-dimensional parameterizations. Simulation-based methods made latent-state and nonlinear models tractable, providing coherent ways to integrate over unobserved states and to characterize uncertainty in complex dynamic settings. This section focuses on Bayesian methods as they bear on time series.

A central Bayesian contribution in macroeconomics was the Bayesian vector autoregression (BVAR) built around shrinkage priors, especially the Minnesota prior associated with Litterman (1986). The motivation was pragmatic. An unrestricted VAR with many variables and lags can fit in-sample fluctuations well, yet with the sample sizes typical in macroeconomics it often yields imprecise coefficients, unstable impulse responses, and weak out-of-sample forecasts. Shrinkage priors address this by pulling many coefficients toward economically sensible benchmarks (often near zero), thereby reducing estimation variance and improving predictive performance.

The Minnesota prior amounts to a small set of prior moments. Coefficients on own lags are allowed to be relatively large, with the first own lag sometimes centered near one when a near random-walk component is plausible. Cross-variable effects and longer lags are shrunk toward zero, typically with prior variances that decline with lag length. The resulting estimates resemble ridge-type regularization, but the Bayesian formulation has a practical advantage. It delivers a predictive density and a coherent way to quantify uncertainty about coefficients, impulse responses, and forecasts.

Two practical features made BVARs attractive for applied work. The prior can be tuned through a small set of hyperparameters that govern overall tightness and lag decay. Those hyperparameters can be chosen using forecast performance, marginal likelihood criteria, or related data-driven rules, which makes the approach adaptable across applications and sample sizes (Giannone, Lenza, & Primiceri, 2015; Carriero, Clark, & Marcellino, 2015). The prior is also transparent enough to be explained in economic terms, such as the belief that most variables are well approximated by low-order autoregressive dynamics and that cross-variable spillovers are typically modest unless the data strongly suggest otherwise.

Shrinkage became even more valuable as empirical work moved from small macro systems to larger information sets. Banbura et al. (2010) showed that, with suitable shrinkage, large BVARs can forecast well even when the number

of parameters is large relative to the sample size. Their analysis established a practical point: the forecasting gains from adding information can dominate the estimation noise induced by high dimensionality, provided that the prior controls overfitting effectively. This logic helped legitimize large-scale forecasting systems in central banks and policy institutions where the goal is not necessarily to interpret parameters but to predict reliably in real time and to analyze scenarios.

While the Kalman filter is computationally efficient in linear-Gaussian systems with known parameters, many models of interest require integration over latent states and parameters in settings where exact Gaussian updating is unavailable. Posterior simulation methods made this feasible. Markov chain Monte Carlo (MCMC) provides a general way to approximate posterior distributions by constructing a Markov chain whose stationary distribution is the target posterior. The Gibbs sampler and related algorithms entered econometrics from the statistics literature (Gelfand & Smith, 1990) and quickly proved useful for dynamic models with latent states. In the state-space context, methods for sampling entire state trajectories, often referred to as simulation smoothing, were central. A standard approach is due to Carter and Kohn (1994), which enabled practical Gibbs sampling schemes that alternate between draws of parameters and draws of the latent state vector.

Subsequent advances in MCMC and related computational methods further widened the range of models that Bayesian estimation could tackle. Hamiltonian Monte Carlo (HMC), which exploits gradient information to explore posterior distributions more efficiently than random-walk Metropolis algorithms, has become increasingly practical for time series models with complex posterior geometries (Neal, 2011). Variational inference methods, which approximate the posterior by optimizing over a family of tractable distributions, offer a computationally cheaper alternative when full MCMC is prohibitively expensive, though with less precise uncertainty quantification (Blei, Kucukelbir, & McAuliffe, 2017).

Stochastic volatility models, introduced in Section 8, illustrate the payoff from combining state-space structure with MCMC. Because the latent log-variance cannot be integrated out analytically, Bayesian estimation via Gibbs sampling and related algorithms was central to making SV models practical (Jacquier et al., 1994; S. Kim et al., 1998). In applied macro-finance work, the resulting density forecasts and credible interval statements, which reflect time-varying uncertainty, became a standard output of Bayesian time series analysis.

Bayesian computation also transformed structural identification in VARs by making it feasible to explore the set of admissible structural models rather than selecting a single point-identified specification. Sign restrictions, introduced in Section 5, impose constraints on the sign (positive or negative) of impulse responses in the periods immediately following shocks, allowing researchers to set-identify structural shocks through economic theory without requiring fully parametrized structural models (Uhlig, 2005; Rubio-Ramírez, Waggoner, & Zha, 2010). This approach proved particularly useful for policy applications where the direction of transmission is known but the magnitude is uncertain. Later work showed, however, that the seemingly agnostic priors underlying these procedures can themselves shape the conclusions, prompting researchers to state and defend their prior assumptions more explicitly (Baumeister & Hamilton, 2015). A complementary approach uses narrative restrictions (Antolín-Díaz & Rubio-Ramírez, 2018), where researchers condition on specific historical episodes to pin down the contributions of particular shocks. Bayesian computation made both approaches practical in large systems, and they are now routine in policy analysis and historical shock decomposition. But it does not supply identification, which still rests on the restrictions and the priors.

The same computational advances reshaped the empirical role of dynamic stochastic general equilibrium (DSGE) models. DSGE frameworks were prominent well before the 1990s, but empirical work often relied on calibration or on limited-information matching of selected moments. The combination of linearized DSGE solutions, state-space representations, and MCMC made likelihood-based estimation more practical. In seminal contributions, Smets and Wouters (2003) and Smets and Wouters (2007) estimated medium-scale DSGE models using Bayesian methods and showed that these models could fit and forecast key macro aggregates competitively, while delivering a coherent decomposition of fluctuations into structural shocks and an explicit accounting of uncertainty.

For models where the likelihood function is intractable but simulation is feasible, approximate Bayesian computation (ABC) offers an alternative approach. ABC methods, including the regression-adjustment approach of Beaumont, Zhang, and Balding (2002), forgo likelihood evaluation and instead use simulations from the model to approximate the posterior distribution. In recent years, ABC methods have attracted interest for complex dynamic models such as agent-based models (Grazzini, Richiardi, & Tsionas, 2017) and large-scale DSGE specifications where likelihood evaluation is prohibitively expensive. ABC trades some precision in the posterior approximation for tractability, and remains valuable where likelihood-free inference is the only option.

Explicit treatment of model uncertainty was another Bayesian contribution to forecasting. Bayesian model averaging (BMA) replaces model selection by weighting models according to posterior probabilities, which can improve predictive

performance when the data-generating process is uncertain and many specifications fit comparably well (Hoeting, Madigan, Raftery, & Volinsky, 1999). It is a probabilistic version of the older case for forecast combination, namely that averaging can stabilize forecasts and outperform selection when models are misspecified or when relative performance shifts over time (Bates & Granger, 1969). The difference is that the weights come from posterior model probabilities rather than from ad hoc rules. In time series settings, BMA also fits the idea that different models may be locally useful at different horizons or under different volatility regimes. A complementary perspective evaluates the forecasts themselves through predictive scoring rules (Geweke & Amisano, 2010), connecting Bayesian analysis to the forecast-evaluation tradition treated in Section 13.

Bayesian methods also changed how uncertainty is reported in applied work. Posterior distributions deliver probability statements about objects such as impulse responses, forecasts, and counterfactual policy experiments, and they make the role of identifying assumptions and prior restrictions explicit. In many macro applications, priors are chosen to be weakly informative, encoding broad beliefs such as stability, modest lag effects, or plausible long-run growth rates. Yet Bayesian analysis also forces researchers to confront sensitivity: posterior conclusions can depend on prior tightness, on identification restrictions, and on how the likelihood treats nonstationarity, measurement error, or structural breaks.

Finally, Bayesian methods made nonparametric dynamic models possible, with the number of latent regimes or states learned from the data rather than fixed in advance. Drawing on Dirichlet process priors and related constructions from Bayesian nonparametric statistics (Ferguson, 1973), these approaches allow flexible regime structure while preserving an explicit probabilistic treatment of dependence and uncertainty. In econometrics, Song (2014) applied an infinite hidden Markov model to integrate regime switching and structural breaks in a unified framework, showing the value of letting the data determine the number and character of regimes.

By the early 2000s, Bayesian VARs and posterior simulation were standard practice, and empirical work could handle large parameterizations, latent states, time variation, and structural models whose likelihoods resist analytical treatment.

11 High-Dimensional Models and the Era of Big Data (2000s)

Entering the 2000s, larger and more varied datasets reshaped empirical macroeconomics and finance. Researchers gained access to hundreds of macroeconomic indicators, firm- and security-level panels, transaction-level financial data, and more timely soft data such as surveys and internet-based measures. The econometric challenge was that classical time series tools were designed for small systems. If one simply adds variables and lags to a VAR or a predictive regression, the number of parameters grows quickly, degrees of freedom vanish, and estimates become unstable. Multicollinearity becomes more severe as similar indicators accumulate, and in finite samples the resulting overfitting can produce deceptively good in-sample fit alongside poor real-time forecasting. The answer was to exploit information in large datasets while imposing enough structure to keep estimation feasible and forecasts reliable. This section concentrates on the time-series response to large datasets.

Two complementary strategies became standard. Dimensionality reduction came first, especially factor models that replace a large panel of observables with a small number of latent components summarizing pervasive comovement. Shrinkage or variable selection came second, implemented either through Bayesian priors or through frequentist penalization, which controls overfitting by pushing many coefficients toward zero and, in some cases, exactly to zero. Closely connected to both was *nowcasting*. Nowcasting treats forecasting as a real-time filtering problem in which indicators arrive at different dates and frequencies, so that the information set evolves within a quarter or month as new releases become available.

The leading approach to dimensionality reduction is the dynamic factor model (DFM) and its close relatives. The basic idea is that a large panel of series $\{x_{it}\}_{i=1}^{N}$ shares a small number of common forces, so that most comovement can be captured by a few latent factors. In a standard static representation one writes $x_{it} = \lambda_i' f_t + e_{it}$, $i = 1, \ldots, N$, where f_t is an $r \times 1$ vector of common factors, λ_i is a vector of factor loadings, and e_{it} is an idiosyncratic component. The representation is termed *static* because the factors enter only contemporaneously; the model is nonetheless *dynamic* in that the factors evolve over time according to their own law of motion, made explicit in the state-space form below. The factors are unobserved and identified only up to rotation, but in macroeconomic panels they often have interpretable content ex post, such as a broad business-cycle factor or a credit and financial conditions factor. The idea of extracting common movements has deep roots in business-cycle measurement (Burns & Mitchell, 1946). 'Diffusion index' methods turned it into a practical forecasting device, summarizing a large, collinear predictor set by a few estimated

factors and using those as predictors for macroeconomic targets (Stock & Watson, 2002). Jurado, Ludvigson, and Ng (2015) took factor-based measurement a step further, constructing an index of macroeconomic uncertainty not from the common movements themselves but from the volatility of the unforecastable variation that remains across a large panel of series. Doz and Fuleky (2020) provide a comprehensive survey of how dynamic factor models evolved from their early origins to the large-scale specifications that became central to macroeconomic forecasting.

The econometric foundations for factor-based forecasting rely on large-N panel asymptotics. Under conditions that make the factors 'pervasive' in the cross section, principal-component estimators recover the factor space as N and T grow, up to an invertible rotation. The early 2000s also brought practical guidance for selecting the number of factors. J. Bai and Ng (2002) proposed information criteria that balance goodness of fit against the complexity induced by adding factors, and J. Bai (2003) developed asymptotic theory for principal-component estimation with large N and T. Such information criteria identify the number of static factors r. Because those factors are typically current and lagged values of a smaller number $q < r$ of dynamic factors, counting the dynamic factors themselves became a separate problem, addressed in later work (J. Bai & Ng, 2007; Hallin & Liška, 2007). When comovement is strong and pervasive, a few factors can summarize much of the signal, while the remaining idiosyncratic components capture measurement error, sector-specific shocks, and short-lived noise. Boivin and Ng (2006) cautioned, however, that bigger panels are not automatically better, since adding noisy or strongly correlated series can worsen factor estimates and forecasts. The subsequent standardization of large macroeconomic datasets, notably the FRED-MD monthly database of McCracken and Ng (2016), made these methods easy to benchmark and replicate.

A second line of work emphasized the dynamic structure of the common component through spectral representations. Like the diffusion-index approach above, the generalized dynamic factor model of Forni, Hallin, Lippi, and Reichlin (2000) rests on an approximate factor structure, allowing idiosyncratic components to be serially and mildly cross-sectionally correlated; what sets it apart is how comovement enters and how it is estimated. Each series responds to a small number of common dynamic shocks through filters, so that a shock reaches different variables with different lags rather than only contemporaneously. Estimation is carried out in the frequency domain, with dynamic principal components replacing the ordinary principal components of the static approach. The separation of common from idiosyncratic variation is again asymptotic: as the cross section expands, only the eigenvalues of the spectral density matrix that correspond to the common shocks diverge, while the rest remain bounded. Later work within the spectral framework developed one-sided estimators suited to real-time forecasting, retaining its dynamic information while avoiding the end-of-sample problems created by two-sided filtering (Forni, Hallin, Lippi, & Reichlin, 2005).

DFMs also map directly onto state-space representations. When factors follow dynamic laws of motion, a convenient formulation is $f_t = A f_{t-1} + u_t$, $x_t = \Lambda f_t + e_t$, where x_t is the $N \times 1$ vector of observables and (u_t, e_t) are innovations. In this form, the Kalman filter delivers optimal updating for linear-Gaussian systems and can handle ragged-edge datasets in which some entries of x_t are missing at the end of the sample. This state-space perspective is well suited to institutions that monitor the economy continuously in real time. It also motivated estimation methods that combine principal components with Kalman smoothing. Doz, Giannone, and Reichlin (2011) and Doz, Giannone, and Reichlin (2012) developed procedures that treat principal-component estimates as noisy proxies for the latent factors and then refine inference using state-space likelihood methods. Once the factor structure was accepted, many high-dimensional forecasting tasks reduced to estimating a low-dimensional state and updating it as new information arrives.

The real-time dimension of large datasets motivated formal nowcasting models built on the same state-space machinery. Because key aggregates such as GDP are released with delay while many indicators arrive monthly, weekly, or even daily, the information set is both incomplete at the end of the sample and uneven in its timing. Giannone et al. (2008) showed how DFMs can be used to extract the real-time information contained in a large panel and translate it into a nowcast of the current quarter. As each release arrives, the Kalman filter updates the nowcast mechanically. The update decomposes into a 'news' component, the surprise in the release relative to what the model predicted, and a weight reflecting how informative that release is for the target (see also Bañbura, Giannone, Modugno, & Reichlin, 2013). This decomposition provides both a forecast and an interpretable account of why the forecast moved, valuable in policy communication and real-time monitoring.

Beyond forecasting, factor methods were integrated into multivariate time-series analysis aimed at policy and structural interpretation. The factor-augmented VAR (FAVAR) framework of Bernanke et al. (2005) embeds estimated factors into a VAR to represent a much larger information set than the analyst could include directly. This addressed a criticism of small VARs, namely that a handful of observed variables may not capture the broader state of the economy that is relevant for identifying policy shocks and describing transmission mechanisms. In a FAVAR, impulse responses can be computed not only for the variables included in the VAR, but also for the full panel through the factor loadings. The approach thus combines the interpretability of VAR analysis with the informational richness of large datasets, while

keeping the core dynamic system low dimensional.

Mixed-frequency and ragged-edge settings motivated additional tools that complement factor models. Mixed data sampling (MIDAS) regressions relate a low-frequency target to higher-frequency predictors through parameterized distributed-lag weight functions, avoiding the need to aggregate high-frequency information to the low frequency (Ghysels, Santa-Clara, & Valkanov, 2004, 2006). A basic MIDAS specification can be written as $y_t = \alpha + \beta \sum_{j=0}^{K} w(j;\theta)\, x_{t-j/m} + u_t$, where y_t is observed at a low frequency, $x_{t-j/m}$ is a higher-frequency regressor sampled m times within the low-frequency period, and $w(j;\theta)$ is a parsimonious weight function governed by a small parameter vector θ. The weight function replaces a long unrestricted distributed lag with a low-dimensional representation that can be estimated with limited T. Ghysels, Sinko, and Valkanov (2007) surveyed and extended MIDAS methods, emphasizing practical issues such as horizon dependence, choice of lag length, and robustness to alternative weight parameterizations. MIDAS thus sat between simple aggregation rules and full state-space mixed-frequency systems (Mariano & Murasawa, 2003; Schorfheide & Song, 2015), particularly in cases where high-frequency predictors are plentiful and timeliness is critical.

Alongside factor models, the 2000s saw rapid spread of regularization methods that handle large predictor sets directly. Penalized regression replaces ordinary least squares with an estimator that trades off fit against complexity, thereby stabilizing estimation when predictors are numerous or strongly correlated. The lasso introduced by Tibshirani (1996) adds an ℓ_1 penalty that can set many coefficients exactly to zero, so it performs variable selection in high-dimensional predictive regressions. This is attractive in macro-finance forecasting problems in which only a subset of a large set of predictors may have incremental predictive content once persistence and multicollinearity are accounted for. The adaptive lasso refines this idea by applying data-dependent weights to the penalty, so that relevant predictors receive less shrinkage and the estimator can achieve oracle properties, performing asymptotically as well as if the true set of active predictors were known in advance (Zou, 2006). A closely related development is the elastic net, which combines ℓ_1 and ℓ_2 penalties to improve performance when predictors are highly correlated and selection of a single series from a tight group is undesirable (Zou & Hastie, 2005). Penalization also pushed practice toward tuning procedures explicitly linked to out-of-sample performance, such as cross-validation and rolling-window forecast evaluation. Whether macroeconomic prediction problems are genuinely sparse is contested, however. Giannone, Lenza, and Primiceri (2021) found that dense specifications, which shrink many small coefficients rather than zeroing most of them out, often predict better, a finding they summarized as the 'illusion of sparsity'.

Regularization was extended to dynamic multivariate systems in which the number of coefficients is especially large. High-dimensional VARs can be estimated under sparsity or shrinkage assumptions on the coefficient matrices, which provide a frequentist analog to Bayesian priors in large BVARs. Kock and Callot (2015) developed theoretical results for lasso-type estimators in high-dimensional VARs, formalizing conditions under which prediction error can be controlled despite the large parameter count. In applied work, these approaches treat the large VAR as a forecasting device. They can also be used for measures such as forecast error variance decompositions, but structural interpretation requires caution when dynamics are estimated under strong regularization, because tuning choices can affect impulse responses even when forecasting performance is stable.

An alternative to penalty-based selection is complete subset regression, which averages forecasts across all possible predictor subsets of a given size rather than selecting a single model (Elliott, Gargano, & Timmermann, 2013). This approach sidesteps the discrete instability of variable selection and pools information across many plausible specifications. Empirically, complete subset regression has proven competitive with regularization methods in macroeconomic forecasting, while being more transparent and requiring no penalty tuning.

Large datasets also reinvigorated panel time-series econometrics. The panel unit-root and cointegration tools discussed earlier (Section 6 and Section 7) became increasingly important as researchers worked with cross-country and cross-sectional panels containing dozens or hundreds of units. In such panels the common shocks that generate cross-sectional dependence are often well described by factor structures, further blurring the boundary between panel and factor-based approaches. The global VAR of Pesaran, Schuermann, and Weiner (2004) offered a complementary route to high-dimensional multivariate modeling, linking country-specific models through cross-unit averages to trace international transmission in systems spanning dozens of economies (Dées, di Mauro, Pesaran, & Smith, 2007).

High dimensionality placed renewed emphasis on covariance estimation and on network-style summaries of dependence. In finance, estimating covariance matrices for many assets is central for risk management and portfolio construction, yet sample covariance matrices can be poorly conditioned when the number of assets is large relative to the available time span. Ledoit and Wolf (2004) proposed shrinkage estimators that stabilize covariance estimation by blending the sample covariance with a structured target, delivering improved conditioning and better out-of-sample portfolio performance in many settings. In macro-finance, related questions arise when measuring systemic risk, spillovers, and connectedness across markets or sectors. Methods that summarize dependence networks, including

variance-decomposition based connectedness measures, gained popularity as a way to describe high-dimensional interaction patterns without estimating unrestricted large systems. Diebold and Yilmaz (2009, 2012) provided an influential variance-decomposition approach to measuring spillovers across variables, illustrating how the language of networks can be combined with familiar time-series objects.

Another practical consideration in big-data forecasting is hierarchical time series reconciliation. Sales forecasts at product, store, and regional levels, for example, must add up across the hierarchy, and Hyndman, Ahmed, Athanasopoulos, and Shang (2011) developed methods that adjust forecasts at different levels so that they satisfy these aggregation constraints.

Empirical comparisons in macroeconomic forecasting often find that factor methods and Bayesian shrinkage approaches perform well relative to simple autoregressive benchmarks when information sets are large (Stock & Watson, 2002; Banbura et al., 2010). By the end of the 2000s, factor methods, regularization, mixed-frequency modeling, and large-panel tools had become standard components of applied time-series econometrics (for a comprehensive treatment, see Fuleky, 2020). The value of big data depends on restrictions that control estimation noise and on evaluation protocols that prioritize out-of-sample performance.

12 Machine Learning and AI in Time Series Econometrics (2010s–2020s)

By the 2010s, rapid progress in machine learning (ML) and artificial intelligence (AI) began to reshape time-series econometrics. Computation and software ecosystems made large-scale estimation routine, and new predictors became available in real time, including text, internet search activity, and high-frequency financial indicators. The high-dimensional methods of the previous section (factor models, shrinkage, out-of-sample evaluation) raised an obvious question: how far could one move beyond linear and Gaussian benchmarks without sacrificing stability (Varian, 2014; Mullainathan & Spiess, 2017)? This section covers the integration of these tools into time-series workflows.

The ML turn shifted emphasis from specifying a single parametric dynamic model to constructing a *prediction pipeline*. The pipeline typically combines feature construction (lags, moving averages, interactions, and indicators of seasonality and calendar effects), regularization or early stopping to control overfitting, and systematic validation through rolling or expanding windows. Econometrics did not abandon time-series structure; it embedded that structure in procedures that respect temporal dependence and real-time forecasting constraints (Bergmeir, Hyndman, & Koo, 2018). A concrete illustration is Prophet, a modular additive model that decomposes a series into trend, seasonal, and holiday components (Taylor & Letham, 2018). It automates fitting and break-date detection, leaving other parameters for the analyst to set. By encoding time-series structure directly into the model specification, such tools made structured forecasting accessible to practitioners outside econometrics.

One important point of contact between econometrics and ML was the shared problem of prediction with many predictors. As discussed in earlier sections, high-dimensional macroeconomic and financial datasets overwhelm classical estimators unless strong restrictions are imposed. Regularization, which penalizes model complexity to stabilize estimation, provided the obvious bridge: it retained regression-based structure while formalizing the bias–variance trade-off that underpins statistical learning (Tibshirani, 1996; Zou & Hastie, 2005). In time-series applications, regularization is often paired with econometric transformations (detrending, seasonal adjustment, variance stabilization) so that flexible estimators do not waste capacity rediscovering trivial persistence or deterministic seasonal patterns.

Beyond regularized linear models, econometricians explored nonlinear supervised learners that capture interactions and thresholds without a pre-specified parametric form. Tree-based methods proved especially attractive: classification and regression trees, random forests, and boosting accommodate many predictors, capture nonlinearities and interactions, and often perform well out of sample with modest tuning (Breiman, Friedman, Olshen, & Stone, 1984; Breiman, 2001; J. H. Friedman, 2001). Random forests and boosting are ensemble methods that use decorrelated trees to reduce variance (in forests) or combine many weak learners to sequentially correct errors (in boosting). In macro and finance forecasting problems, trees can automatically capture asymmetric predictor effects (e.g., a variable matters in recessions but not expansions) and interaction patterns (e.g., the effect of a credit spread depends on monetary conditions) that would be cumbersome to pre-specify in linear models. Systematic evaluations found that these methods can deliver material gains in macroeconomic forecasting, especially at longer horizons and in turbulent periods, with random forests proving notably robust (Medeiros et al., 2021; Goulet Coulombe, Leroux, Stevanovic, & Surprenant, 2022); Masini, Medeiros, and Mendes (2023) survey this rapidly growing literature.

This flexibility comes with familiar concerns. Without careful validation, ensemble methods can overfit, especially when the analyst searches over many feature sets and tuning parameters, and even an accurate model can be hard to

interpret economically. As a result, interpretability tools became part of the applied workflow. Partial dependence plots trace the average effect of a predictor on the forecast (J. H. Friedman, 2001), while permutation importance measures how much accuracy the model loses when a predictor is shuffled (Breiman, 2001). Local methods such as LIME and Shapley-value decompositions instead attribute a single forecast movement to individual predictors (Ribeiro, Singh, & Guestrin, 2016; Lundberg & Lee, 2017). In econometric practice, such diagnostics are best treated as aids to monitoring and communication rather than evidence of causal mechanisms, a distinction taken up below.

The emphasis on honest forecast evaluation became more explicit with the spread of ML workflows. In time-series data, evaluation must account for temporal dependence, though ordinary cross-validation can be valid for stationary autoregressive models with uncorrelated errors (Bergmeir et al., 2018). Tuning must use only data available before the forecast date; otherwise look-ahead bias inflates measured performance. This concern has long been present in econometrics under the headings of data mining and specification search, but ML made the issue far more acute because the feasible search space expanded. Practical remedies include strict separation between model development and assessment windows, comparisons against strong benchmarks (including simple autoregressive models), and the routine use of forecast combinations, which hedge against over-committing to any single specification drawn from a large search space (Timmermann, 2006; Elliott et al., 2013).

Deep learning was a more substantial departure from traditional econometric modeling. While neural networks had attracted econometric attention much earlier (Kuan & White, 1994; Swanson & White, 1997), the 2010s brought architectures and training methods that made large models feasible. Recurrent neural networks, and long short-term memory (LSTM) networks in particular, were designed to represent sequential dependence and longer-range dynamics (Hochreiter & Schmidhuber, 1997; Hewamalage, Bergmeir, & Bandara, 2021). However, early evidence in economic forecasting was mixed, largely because many macroeconomic time series are short relative to the number of parameters in deep networks.

This tension between flexibility and sample size became visible in large-scale evaluations. In the Makridakis M4 forecasting competition, purely ML methods did not dominate traditional statistical baselines on average, and combinations of classical methods remained difficult to beat (Makridakis et al., 2020). Yet M4 also pointed to a lasting theme for subsequent work. The winning approach combined a recurrent network with exponential smoothing components, pairing a flexible learner with strong time-series structure (Smyl, 2020). The implication was that deep learning can add value when embedded within, or constrained by, time-series structure rather than deployed as an unconstrained black box. More generally, hybrid workflows can use econometric components to handle detrending, seasonal adjustment, or baseline linear dynamics, while ML components capture residual nonlinearities and interactions.

The forecasting literature also adapted generic deep-learning templates to time series. Temporal convolutional networks, which use causal dilated convolutions to capture long-range dependencies without recurrence, offered a computationally efficient alternative to LSTMs (S. Bai, Kolter, & Koltun, 2018). Attention-based architectures, inspired by the transformer model of Vaswani et al. (2017), were adapted for multi-horizon forecasting. The Temporal Fusion Transformer of Lim, Arik, Loeff, and Pfister (2021) incorporated explicit temporal structure, including variable selection networks and interpretable attention mechanisms, and subsequent work produced further variants optimized for long-horizon prediction (Nie, Nguyen, Sinthong, & Kalagnanam, 2023; Zhou et al., 2021; Wu, Xu, Wang, & Long, 2021). That said, simple linear baselines have proved surprisingly hard to beat in several of these benchmarks, a reminder that architectural sophistication does not guarantee forecasting gains (Zeng, Chen, Zhang, & Xu, 2023). For broader surveys of deep-learning forecasting architectures and design choices, see Benidis et al. (2022).

A second theme was pooling across many related series. Unlike i.i.d. cross-sectional settings, time-series applications can remain data-poor in the time dimension even as the cross-sectional or predictor dimension grows. One response is to expand the effective training set by learning jointly across many comparable series. In such settings, so-called *global* models can learn shared representations and outperform separate univariate fits, as illustrated by methods such as DeepAR and N-BEATS (Salinas, Flunkert, Gasthaus, & Januschowski, 2020; Oreshkin, Carpov, Chapados, & Bengio, 2019). This perspective also clarifies why ML methods performed particularly well in the M5 competition, which involved thousands of related retail series and rich covariates (Makridakis, Petropoulos, & Spiliotis, 2022). A close econometric parallel is panel forecasting, which poses the same evaluation question: whether the target is future values for the series already in hand, or generalization to series the model has never seen. The competition tradition itself continued with M6, which turned to financial return prediction and investment performance (Makridakis et al., 2025).

More recently, *foundation models* for time series have emerged. These are large pre-trained models designed to forecast or represent temporal patterns across diverse domains without task-specific retraining. Analogous to large language models in natural language processing, these models are trained on massive corpora of time series from heterogeneous sources and can be applied zero-shot or with minimal fine-tuning to new series. Early examples include

TimeGPT (Garza et al., 2023) and Chronos (Ansari et al., 2024), which treat time series forecasting as a sequence modeling problem and build on transformer architectures trained on large collections of publicly available time series. Foundation models promise to capture general temporal patterns from pre-training, including trends, seasonality, level shifts, and nonlinear dynamics, which could reduce the need for domain-specific feature engineering and model selection. Their economic relevance is still being tested. Macroeconomic and financial series differ from the retail, energy, and web-traffic data that dominate many pre-training corpora: they have structural breaks, policy feedback, nonstationarity, and relatively short histories. Whether pre-trained representations transfer effectively to economic forecasting tasks, and whether they can be combined productively with domain-specific econometric structure, are important open questions. Across these developments, the most consequential modeling choice is often the representation of the time-series problem rather than the learning algorithm.

Uncertainty quantification in ML-based time series forecasting also received growing attention. Traditional econometric models produce prediction intervals from distributional assumptions or posterior simulation, but many ML methods yield only point forecasts. Conformal prediction provides a distribution-free framework for constructing prediction intervals with finite-sample coverage guarantees, requiring only exchangeability rather than parametric distributional assumptions (Vovk, Gammerman, & Shafer, 2005). Adaptations to time series, which must contend with temporal dependence and potential nonstationarity, show promise for providing calibrated uncertainty bands around ML forecasts (Zaffran, Feron, Goude, Josse, & Dieuleveut, 2022; Barber, Candès, Ramdas, & Tibshirani, 2023). This line of work helps connect the flexibility of ML prediction with the uncertainty quantification that is central to econometric practice and policy communication.

Nowcasting and real-time monitoring provided another obvious entry point for ML. Many institutions routinely work with mixed-frequency datasets, ragged edges, and rapidly updating indicators. Classical state-space and dynamic factor model frameworks remain strong baselines because they accommodate missing observations and provide coherent updating through the Kalman filter (Giannone et al., 2008). ML methods complement these frameworks mainly by widening the set of usable inputs, converting alternative data that arrive well ahead of official statistics into predictors for the monitoring system. Internet search indices were an early example, improving short-horizon prediction of series such as unemployment claims and retail sales (Choi & Varian, 2012). Text-as-data methods extended the idea, distilling news and other documents into quantitative indicators of sentiment and economic policy uncertainty (Gentzkow, Kelly, & Taddy, 2019; Ardia, Bluteau, & Boudt, 2019; Baker, Bloom, & Davis, 2016). Because these signals are timely and observed at high frequency, they slot into nowcasting models alongside conventional data releases.

Finally, the ML turn drew a sharper boundary between forecasting and structural inference. ML excels at high-dimensional prediction and pattern extraction, but improved forecast accuracy does not automatically deliver interpretable causal mechanisms or policy counterfactuals. In most time-series applications, ML methods are used for reduced-form prediction, where the objective is to minimize forecast loss subject to out-of-sample validation. That objective sits comfortably with shrinkage, ensembles, and automated feature construction, but not with the goals of structural modeling, where identification, interpretation, and counterfactual analysis are central. The distinction grew starker as researchers adapted ML to problems where the goal is inference on a specific parameter rather than prediction alone. Double, or 'debiased,' machine learning, for instance, delivers valid inference on a target parameter such as a treatment effect, while using flexible ML to absorb many nuisance controls (Chernozhukov et al., 2018). Related work in causal ML shows how algorithmic tools can serve explicit inferential targets, including tree-based estimators of heterogeneous treatment effects and synthetic control methods for policy evaluation, though time-series dependence and feedback remain active complications (Athey & Imbens, 2016; Wager & Athey, 2018; Abadie, Diamond, & Hainmueller, 2010).

Classical time-series structure and ML flexibility are now combined, tested, and compared under transparent evaluation protocols. But the vastly larger model spaces opened up by tuning, feature construction, and alternative data raise the risk of overfitting and spurious improvements, which makes credible evaluation more important than ever.

13 Testing, Forecast Evaluation, and Model Comparison

Throughout its development, time series econometrics has maintained a strong tradition of model assessment and comparison. New models and methods matter only to the extent that they can be validated and shown to improve either inference or prediction. New stochastic specifications drove the discipline forward, but so did a growing set of methods for diagnostics, forecast evaluation, and principled model choice.

Box–Jenkins analysis exemplified early diagnostic testing. Residual checks and portmanteau statistics such as the Ljung–Box test asked whether an ARIMA specification had captured serial dependence adequately (Box & Jenkins,

1970; Ljung & Box, 1978). The Breusch–Godfrey Lagrange multiplier test generalized these checks to regression models with lagged dependent variables and other regressors, providing a more reliable test for residual serial correlation than the earlier Durbin–Watson statistic in dynamic settings (Breusch, 1978; Godfrey, 1978). As richer models emerged, analogous diagnostics followed. For volatility models, Engle's Lagrange multiplier test for ARCH effects became a standard tool for detecting conditional heteroskedasticity (Engle, 1982), and subsequent GARCH practice inherited a parallel suite of diagnostics for standardized residuals. For VARs, practitioners routinely examine residual autocorrelation (using multivariate portmanteau tests), stability (all eigenvalues inside the unit circle), and the plausibility of impulse responses under alternative identification schemes. The unit-root and cointegration revolutions brought their own specification tests (Sections 6 and 7), which became part of routine diagnostic practice.

Forecast evaluation developed alongside these diagnostics, once it became clear that comparing forecast accuracy requires formal statistical inference rather than informal comparisons of mean squared error. Theil (1961, 1966) laid early foundations, with an inequality coefficient and a decomposition of forecast errors into bias, variance, and covariance components that provided some of the first systematic accuracy diagnostics. Another early diagnostic is the rationality regression, in which realized values are regressed on point forecasts to check whether the intercept is zero and the slope is unity (Mincer & Zarnowitz, 1969). Deviations signal systematic bias or a failure to exploit available information, and these simple tests remain a standard first check on forecast quality.

A key milestone was the Diebold–Mariano test, a general procedure for testing whether two sets of forecasts are equally accurate under an explicit loss function (Diebold & Mariano, 1995). It operates on the time series of loss differentials, using long-run variance estimation to handle the serial dependence that multi-step and overlapping forecasts induce. By attaching sampling uncertainty to claims of predictive improvement, the test helped standardize forecast comparisons in macroeconomics and finance. An important subtlety concerns what is actually being compared. The test takes the two forecasts as given and compares only their realized errors, so it applies however they were produced—by models, surveys, or judgment. Judging the underlying models is a different task, because their parameters were estimated and the test must allow for that. How much it matters depends on the estimation scheme: whether parameters are fixed once on an initial sample, or refit as the window rolls forward or expands. West (1996) supplied the asymptotic theory for comparing forecasts from estimated models, showing when estimation error and the scheme used to update parameters alter the test's distribution, and Giacomini and White (2006) extended it to tests of conditional predictive ability. Diebold (2015) later reflected on the widespread use, and frequent misuse, of such predictive-accuracy tests in applied work.

A long-standing technical issue in forecast evaluation and testing is that serial dependence invalidates naive variance formulas. Long-run variance estimation therefore became routine in time series inference, both in predictive-accuracy testing and in regression-based diagnostics. Newey and West (1987) popularized heteroskedasticity and autocorrelation consistent (HAC) covariance estimation, and Andrews (1991) developed key results on bandwidth choice and long-run variance estimation under dependence; Lazarus, Lewis, Stock, and Watson (2018) later distilled decades of subsequent refinements into concrete recommendations for practice.

Resampling methods provided a complementary route to inference in settings where asymptotic approximations are unreliable or statistics are complex. The block bootstrap of Künsch (1989) and the stationary bootstrap of Politis and Romano (1994) are widely used examples that preserve dependence structure while enabling approximate sampling distributions for test statistics and forecast comparisons.

An important consideration for real-time forecasting is whether model performance remains stable over time. Fluctuation tests for forecast breakdown can detect periods when a model's predictive ability declines, providing early warning that recalibration may be needed (Giacomini & Rossi, 2009); Rossi (2021) surveys the broader problem of forecasting under instabilities and the tools available for detecting and adapting to them. A related concern is the use of real-time versus revised data in evaluation. Croushore (2011) showed that using final revised figures rather than the data vintage actually available to the forecaster can produce misleading assessments of forecast quality. The point is especially relevant in nowcasting and short-horizon prediction, where the goal is to exploit the information available today.

A further complication arises in nested forecast comparisons, where one model is a restriction of another (for example, adding predictors to a benchmark autoregression). In such settings, naive comparisons of out-of-sample loss can be misleading because the larger model may overfit in-sample noise while offering similar population-level predictive ability. Tests that adjust for estimation error under the nesting null, such as the procedure proposed by Clark and West (2007), showed that forecast evaluation requires aligning the statistical framework with the specific model comparison at hand.

Beyond pairwise accuracy tests, time series econometrics developed the concept of forecast encompassing. Forecast

encompassing asks whether one forecast subsumes the information in another. If combining them yields no improvement, the first is said to encompass the second. Formal tests of encompassing and related regression-based diagnostics became widely used (D. I. Harvey, Leybourne, & Newbold, 1998).

When neither forecast encompasses the other, combining them can improve accuracy. The empirical success of forecast combinations confirmed an older insight that, under model uncertainty, pooling information can be more robust than selecting a single 'best' model. Classic work on combining forecasts (Bates & Granger, 1969) and later regression-based combination schemes (Granger & Ramanathan, 1984) helped turn forecast combination into a practical methodology rather than an ad hoc practice. A persistent empirical finding is that simple equal-weighted averages often perform comparably to or better than estimated optimal weights, suggesting that the gains from diversification across models dominate the costs of estimation error in constructing combination weights (Timmermann, 2006). In modern applications, this logic extends naturally to Bayesian model averaging, ensemble methods, and other forms of aggregation that trade off bias and variance under pervasive model misspecification.

For in-sample model selection, information criteria became central, particularly in ARIMA and VAR specification. Akaike's AIC provided a parsimonious rule, $\text{AIC} = -2\ln(L) + 2k$, where L is the maximized likelihood and k the number of estimated parameters; it approximates out-of-sample predictive performance by penalizing complexity (Akaike, 1974). Schwarz's BIC introduced a stronger penalty, of order $\ln(n)$ with n the sample size, and is consistent for selecting the true finite-dimensional model under standard regularity conditions (Schwarz, 1978). AIC is asymptotically efficient for prediction in many settings, while BIC is consistent for model identification when a true model exists in the candidate set. The contrast helped formalize a distinction that runs throughout the history of time series econometrics: prediction-oriented choice versus structure-oriented selection. Intermediate criteria such as Hannan–Quinn further illustrated this spectrum by using a penalty between AIC and BIC (Hannan & Quinn, 1979). In practice, these criteria became the standard means of selecting lag orders, choosing ARMA and GARCH specifications, and guiding parsimonious parametrizations when likelihood-based estimation is feasible.

Alongside information criteria, model selection in applied time series econometrics also developed as a set of specification-search protocols. A prominent approach, associated with the LSE tradition, is the general-to-specific method, which begins from a deliberately general dynamic model and simplifies it through sequential testing and diagnostic checking, aiming at a parsimonious model that still fits the data (Hendry, 1995). The rise of large candidate model sets sharpened concerns about search-induced overfitting, which helped motivate explicit discussions of data mining and the role of encompassing in specification search (Hoover & Perez, 1999). Subsequent work on automated general-to-specific procedures, such as PcGets, further formalized these protocols and studied their performance under extensive search (Krolzig & Hendry, 2001).

Classical hypothesis testing nevertheless remained indispensable. Researchers routinely test individual restrictions (t-tests), joint restrictions (Wald and likelihood ratio tests), and Granger non-causality restrictions in VARs. Structural-break methods extended this logic to parameter stability, from the Chow test through the unknown-break-date procedures discussed in Section 8 on volatility and nonlinear dynamics. In dynamic environments it is rarely enough to estimate a model; one must also test whether its estimated relationships hold across sample periods and policy regimes.

By the 1990s, forecast evaluation faced a new challenge as empirical researchers increasingly compared large collections of forecasting models, decision rules, or candidate predictors. This raised 'data snooping' concerns, since the best-performing model in a large search may look impressive even when all models have equal population predictive ability. That helped shift attention toward multiple-testing adjustments and joint inference, particularly in finance and macro-forecasting. White's 'reality check' provided an early framework for testing whether the best model in a set genuinely improves expected loss relative to a benchmark, accounting for the search over alternatives (White, 2000). Hansen's superior predictive ability test refined this approach to improve power against relevant alternatives (Hansen, 2005). Related procedures, such as model confidence sets, aim to identify a subset of models that are statistically indistinguishable from the best performer under a chosen loss function, reinforcing the practical message that model uncertainty often persists even after extensive evaluation (Hansen, Lunde, & Nason, 2011).

Another important extension concerns probabilistic forecasts. As Bayesian methods, state-space models, and ensemble approaches encouraged reporting full predictive distributions rather than point forecasts, evaluation increasingly relied on scoring rules and calibration diagnostics. Probability integral transform (PIT) techniques assess whether realized outcomes are consistent with a model's predictive distribution, and they became widely used for diagnosing miscalibration (Dawid, 1984; Diebold, Gunther, & Tay, 1998). Proper scoring rules such as the log predictive score or the continuous ranked probability score (CRPS) provide coherent loss functions for density forecasts, allowing direct comparisons of predictive distributions (Gneiting & Raftery, 2007). Geweke and Amisano (2010) developed methods for using cumulative log predictive scores to compare Bayesian models in terms of sequential out-of-sample predictive

performance, formalizing density forecast comparison as a natural extension of point-forecast evaluation. These tools helped connect classical forecast evaluation to modern probabilistic forecasting and showed that 'accuracy' means more than point precision, extending to the calibration and sharpness of predictive uncertainty. The Bank of England reinforced the shift by adopting 'fan charts' for inflation and GDP growth forecasts in 1996 (Britton, Fisher, & Whitley, 1998), which made density forecasting a routine element of policy communication and encouraged other central banks to report forecast uncertainty systematically.

Finally, model assessment established out-of-sample validation as essential. Rolling or expanding-window exercises simulate real-time forecasting and provide a structured way to evaluate how models perform under evolving data and potential structural change. In empirical work, reserving the final portion of the sample for genuine out-of-sample assessment became commonplace, both to guard against overfitting and to ensure that proposed innovations deliver predictive gains under realistic information sets.

Cross-checking across methods is increasingly common in applied work. Researchers often validate structural models by checking forecasting performance or consistency with reduced-form evidence. For example, if a DSGE model's forecasts are much worse than a BVAR's, one might suspect misspecification in the theoretical structure. Similarly, structural shock identifications in VARs are now frequently cross-checked against external evidence, including narrative measures or instrumental-variable approaches (Romer & Romer, 2004; Gertler & Karadi, 2015).

Testing, forecast evaluation, and model comparison impose a discipline on the field. New models must demonstrate value either through well-posed hypothesis tests, through predictive improvements under transparent loss functions, or through robustness across plausible alternatives. These methods range from residual diagnostics and unit-root tests to predictive-accuracy tests, multiple-comparison adjustments, and probabilistic scoring rules. By rewarding what survived empirical scrutiny, they shaped which innovations endured.

14 Comparative Assessment of Time Series Methods

The chapter has covered a variety of models and approaches, each with its own philosophy, mathematical structure, and range of applicability. They differ along four dimensions: forecasting ability, structural insight, the data features they handle, and ease of implementation.

- **Univariate Methods (ARIMA and Exponential Smoothing)**
  - *Key features.* ARIMA models represent a single series through its own lags, differences, and moving-average terms, following the Box–Jenkins identification–estimation–diagnostic cycle (Box & Jenkins, 1970). Exponential smoothing methods decompose a series into level, trend, and seasonal components with adaptive smoothing weights (Holt, 1957; Winters, 1960); the ETS (error, trend, seasonality) framework places these methods on statistical foundations with explicit likelihoods and prediction intervals (Hyndman et al., 2002).
  - *Strengths.*
    - Strong short-term forecasting performance; both families remain competitive benchmarks in forecasting competitions, including as components of hybrid models (Smyl, 2020).
    - Well-understood diagnostics (ACF/PACF for ARIMA; information criteria and residual checks for both).
    - Captures autocorrelation and seasonality (SARIMA, Holt–Winters).
    - Robust, easy to automate, and widely used in business, government, and official statistics.
  - *Limitations.*
    - Ignores relationships with other variables; cannot capture co-movement, support causal analysis, or provide structural interpretation.
    - Fixed coefficients and smoothing weights; struggles with structural breaks and regime changes.
- **Structural Econometric Models (Cowles/DSGE)**
  - *Key features.* Systems of equations grounded in economic theory (e.g., simultaneity, rational expectations) with parameters tied to economic interpretation. This class includes dynamic stochastic general equilibrium (DSGE) models and earlier Cowles Commission simultaneous-equation systems, both of which impose theoretical structure to achieve identification.

  - *Strengths.*
    * Interpretability. Parameters are tied to economic concepts (elasticities, etc.), which is useful for counterfactual policy analysis and scenario exercises.
    * Micro-founded variants that model expectations explicitly (notably DSGE models) are designed to withstand policy regime changes, addressing the Lucas critique (Lucas, 1976)—a critique originally aimed at the Cowles-style systems themselves.
  - *Limitations.*
    * Can be misspecified if the underlying theory is incomplete or wrong (e.g., relying on fragile identifying assumptions).
    * Large systems can be unwieldy or over-parameterized; many early structural models were not fully estimated.
    * Have often been less competitive than reduced-form alternatives in forecast comparisons, though Bayesian estimation has narrowed this gap for medium-scale DSGE models (Sims, 1980; Smets & Wouters, 2007).

- **VAR (Vector Autoregression)**
  - *Key features.* Treats all variables as endogenous; each equation includes lags of all variables; minimal theory imposed aside from variable selection (Sims, 1980; Stock & Watson, 2001).
  - *Strengths.*
    * Captures joint dynamics of multiple series (accounts for feedback loops).
    * Competitive with large structural systems in short-run forecasting exercises.
    * Impulse response analysis allows tracing out effects of shocks (with identifying assumptions), facilitating economic interpretation. Local projections provide a complementary, robust alternative for estimating impulse responses without committing to the full VAR specification (Jordà, 2005).
    * Flexible: can include trends, seasonal dummies, etc.
  - *Limitations.*
    * Parametric 'curse of dimensionality.' The number of parameters grows quickly with more variables and lags, creating overfitting risk unless sample size is large.
    * Requires identification (additional assumptions) for causal interpretation of impulses; identifying restrictions can be debated.
    * Assumes linear structure; cannot easily handle regime switches or nonlinear dynamics without extensions.
    * Not designed for long-run equilibrium constraints (unless extended to a VECM for cointegration).

- **Cointegrated VAR / Error-Correction (VECM)**
  - *Key features.* Extension of VAR that incorporates cointegration relationships (long-run equilibrium constraints) among variables; error-correction terms push the system back toward equilibrium after shocks (Engle & Granger, 1987; Johansen, 1991).
  - *Strengths.*
    * Long-run insight: distinguishes permanent vs. transitory movements; estimated cointegration vectors often correspond to meaningful equilibria (e.g., a money demand relationship).
    * Avoids spurious regression when dealing with nonstationary $I(1)$ series.
    * Can improve long-horizon forecasts by anchoring to long-run equilibria.
  - *Limitations.*
    * Requires sufficiently long samples to estimate cointegration reliably (tests can have low power).
    * Sensitive to specification (e.g., whether to include trends, how many lags).
    * Interpretation of multiple cointegrating relations can be difficult.

    - If structural breaks alter long-run relationships, cointegration analysis can be invalid unless such breaks are modeled.

- **Volatility Models (ARCH/GARCH/SV)**
  - *Key features.* Models time-varying conditional variance using past errors (ARCH terms) and past variances (GARCH terms); many variants (EGARCH, GJR for asymmetry, multivariate GARCH for covariances), and can be combined with other models (e.g., a VAR with GARCH errors) (Engle, 1982; Bollerslev, 1986). Stochastic volatility (SV) models offer a parameter-driven alternative in which the log-variance follows its own latent process, estimated via MCMC or particle methods (Jacquier et al., 1994).
  - *Strengths.*
    - Captures volatility clustering in financial and economic data.
    - Provides improved forecast intervals and risk measures (e.g., Value-at-Risk) by forecasting time-varying volatility. Quantile regression methods such as CAViaR offer a semiparametric alternative for modeling tail risk directly (Engle & Manganelli, 2004).
    - Flexible extensions allow modeling asymmetry (leverage effects) and multivariate volatility (dynamic correlations).
    - Benchmark tool in finance for volatility forecasting.
  - *Limitations.*
    - Focuses on second moments (variance); mean dynamics are often modeled simply (e.g., i.i.d. or a low-order AR).
    - Sensitive to distributional assumptions; non-normal errors complicate inference.
    - Numerous variants require careful specification choice; risk of overfitting in small samples.
    - Adoption outside finance was long limited; in macroeconomics, time-varying volatility entered mainly through SV within TVP-VAR frameworks (Primiceri, 2005).

- **Nonlinear / Regime-Switching (TAR, Markov-Switching)**
  - *Key features.* Allows different dynamics in different regimes. Examples include threshold autoregressions (TAR), where regimes are defined by an observed variable crossing a threshold; smooth-transition autoregressions (STAR), where parameters change gradually (Granger & Teräsvirta, 1993); and Markov-switching models, where a latent state follows a stochastic transition (Tong, 1983; Hamilton, 1989).
  - *Strengths.*
    - Captures asymmetries or regime-specific behavior that linear models miss (e.g., recession vs. expansion dynamics; high- vs. low-volatility periods).
    - Markov-switching models can detect turning points probabilistically.
    - Can represent discrete structural change via shifts to new regimes.
    - Often improves fit when nonlinearities are empirically important.
  - *Limitations.*
    - Estimation is more complex (nonlinear likelihoods, potential for local maxima; often need good initial values).
    - Interpretation can be difficult if inferred regimes do not map cleanly into observable states.
    - Forecasting is complicated because one must forecast both values and regime probabilities.
    - Threshold choice and search can be ad hoc and may raise data-mining concerns.
    - Typically applied to small systems because extending to high dimensions is difficult.

- **Bayesian VAR (BVAR)**
  - *Key features.* VAR estimated with Bayesian priors (e.g., the Minnesota prior, which shrinks coefficients toward random-walk or zero values) (Litterman, 1986). Typically solved via MCMC or analytical posterior formulas under Gaussian assumptions.

- *Strengths.*
  - Handles higher dimensions through shrinkage toward reasonable defaults, allowing many variables without severe overfitting.
  - Often improves forecast accuracy relative to unrestricted VAR by addressing the bias–variance tradeoff.
  - Delivers predictive distributions, useful for uncertainty quantification.
  - Allows incorporating expert judgment or external information via priors.
- *Limitations.*
  - Results can be sensitive to prior choices (though default priors often work well for macro data).
  - Computationally heavier than standard VARs (though modern computing makes large BVARs feasible).
  - Shrinkage makes large systems estimable but does not identify structural shocks.
  - Priors encode external information, and classical p-values give way to posterior credible intervals, which some practitioners find unfamiliar.

- **State-Space / Unobserved-Components**
  - *Key features.* Represents observed series through measurement equations linked to latent states, with filtering and smoothing used to infer trends, cycles, seasonal components, or time-varying parameters (Kalman, 1960; A. C. Harvey, 1989; Durbin & Koopman, 2001). Time-varying parameter (TVP) models, in which coefficients drift as latent processes, are a prominent application (Cogley & Sargent, 2005; Primiceri, 2005).
  - *Strengths.*
    - Naturally handles missing data, mixed frequencies, and real-time updating.
    - Delivers interpretable decompositions and full predictive distributions.
    - Provides a flexible platform for time-varying parameters and unobserved components.
  - *Limitations.*
    - Results can be sensitive to how latent components are specified.
    - Nonlinear or non-Gaussian versions can be computationally demanding.
    - Structural interpretation depends on whether the latent states have a convincing economic meaning.

- **Factor Models (Dynamic Factor, FAVAR)**
  - *Key features.* Reduce dimensionality by extracting latent factors from many series (via principal components or state-space methods); factors are then used in regressions or VARs. FAVAR (factor-augmented VAR) combines factor analysis with VAR structure to incorporate broad information (Stock & Watson, 2002; Bernanke et al., 2005).
  - *Strengths.*
    - Dimension reduction: can use information from dozens or hundreds of series without exhausting degrees of freedom.
    - Factors often have intuitive interpretations (e.g., a general activity factor or a financial conditions factor).
    - Strong performance for nowcasting and forecasting when broad signals matter.
    - FAVARs allow impulse-response analysis that reflects information from many variables without explicitly modeling each one.
  - *Limitations.*
    - Estimated factors are statistical objects and not always easy to interpret causally.
    - Require large cross-sections and common co-movement to obtain stable factors.
    - Choosing the number of factors can be subjective (though criteria exist) (J. Bai & Ng, 2002).
    - Important predictors that do not co-move with others may be poorly represented by factors.

- **Penalized Regression (Lasso, Ridge, Elastic Net)**

- *Key features.* Linear models with data-driven shrinkage or variable selection applied to high-dimensional regressions and VARs. The lasso imposes an ℓ_1 penalty that can set coefficients exactly to zero; ridge regression uses an ℓ_2 penalty; the elastic net combines both (Tibshirani, 1996; Zou & Hastie, 2005). These methods bridge classical econometrics and machine learning by retaining the linear regression framework while automating model selection.
- *Strengths.*
  - Handles many predictors even in short samples by controlling overfitting through penalization.
  - Coefficients stay interpretable, though shrinkage biases them and selection complicates inference.
  - Debiasing and orthogonalization techniques extend penalized methods to formal inference on low-dimensional targets (Chernozhukov et al., 2018).
  - Computationally efficient and easy to integrate into existing econometric workflows.
- *Limitations.*
  - Assumes linearity; cannot capture nonlinear relationships or interactions without manual feature construction.
  - Performance depends on tuning the penalty parameter, typically via cross-validation, which must respect temporal dependence.
  - Sparsity assumptions may not hold when many predictors contribute small effects (Giannone et al., 2021).

- **Nonparametric Machine Learning (Trees, Neural Nets, Deep Learning)**
  - *Key features.* Flexible nonparametric or semi-parametric models that can accommodate nonlinear relationships and a large number of predictors. Examples include tree-based ensembles (random forests, gradient boosting), support vector machines, recurrent and convolutional neural networks, and transformer architectures. Many incorporate regularization or ensemble techniques to mitigate overfitting (Breiman, 2001).
  - *Strengths.*
    - Can capture nonlinear relationships, interactions, and complex temporal patterns.
    - High-dimensional predictive performance when sufficient data are available.
    - Tree ensembles are often robust to overfitting via averaging; neural networks can approximate complex functions given sufficient data.
    - Some methods handle missing data and outliers well.
  - *Limitations.*
    - Data-intensive; can underperform in the short samples typical of macroeconomics unless combined with strong temporal structure.
    - Limited interpretability; post-hoc explanation tools provide partial remedies but do not establish causal mechanisms.
    - Risk of overfitting without careful tuning and validation, particularly when model spaces are large.
    - Primarily designed for prediction; structural interpretation and policy counterfactual analysis require additional identification assumptions.

- **Forecast Combination and Ensembling**
  - *Key features.* Combines forecasts from multiple models (simple averages, weighted averages, or more complex weighting schemes) (Bates & Granger, 1969).
  - *Strengths.*
    - Robustness to model uncertainty; averaging reduces the risk of large forecast errors.
    - Often improves accuracy empirically.
    - Easy to implement; even equal-weight combinations perform well in many settings.
    - Can exploit complementary strengths across distinct modeling approaches.

- *Limitations.*
  - Does not itself provide structural interpretation.
  - If component models are too similar or share the same biases, combining yields limited gains.
  - Complex weighting schemes can overfit past performance.

Several dimensions cut across these model classes and separate them in practice.

- **Forecasting vs. inference.** Methods range from primarily predictive (ARIMA, factor models, machine learning) to primarily structural (Cowles/DSGE models, identified SVARs). VARs span both roles: they provide competitive short-run forecasts in reduced form and, with identifying assumptions, support impulse-response analysis. In practice, major institutions maintain portfolios that pair theory-driven frameworks for scenario and counterfactual analysis with reduced-form or data-driven tools for near-term prediction, cross-checking the two for consistency (Sims, 1980; Smets & Wouters, 2007). Fair and Shiller (1990) give the empirical case for keeping both: structural and VAR forecasts each carried information absent from the other.

- **Model stability and structural change.** A perennial concern in applied work is whether a model's parameters remain stable over the sample. Most specifications assume fixed parameters and can deteriorate when relationships shift. Regime-switching models address this by allowing dynamics to change across states (Hamilton, 1989), and time-varying parameter models accommodate gradual evolution (Cogley & Sargent, 2005; Primiceri, 2005). In routine practice, simpler strategies—rolling-window estimation, break dummies, or periodic re-specification—are common complements to formal structural-change methods.

- **Dimensionality.** With expanding data, models that handle many series are increasingly important. BVARs and factor models manage high-dimensional information through shrinkage or dimension reduction; machine learning methods are designed for large predictor sets but can overfit when time series are short unless combined with penalization or sparsity constraints (Tibshirani, 1996; Banbura et al., 2010).

- **Uncertainty and density forecasts.** Quantifying forecast uncertainty now matters as much as producing point predictions. State-space models, BVARs, and time-varying volatility models, including GARCH-type and latent SV specifications, naturally produce full predictive distributions, which are central to risk management and policy communication. Evaluation then shifts from mean-squared error to calibration diagnostics and proper scoring rules such as the log score or CRPS (Diebold et al., 1998; Gneiting & Raftery, 2007). Many machine learning algorithms yield only point forecasts and require supplementary tools to quantify uncertainty.

- **Real-time information and mixed frequencies.** Methods differ in how well they accommodate real-time forecasting constraints. Key targets are released with delay and are revised, while predictors arrive at different frequencies and dates (Croushore, 2011). State-space and dynamic factor frameworks are designed for this environment. They handle missing observations at the end of the sample and update forecasts mechanically as new releases arrive, which is central in practical nowcasting (Giannone et al., 2008). MIDAS regressions offer a complementary approach by relating low-frequency targets directly to high-frequency predictors through parsimonious distributed-lag weight functions (Ghysels et al., 2004).

- **Ease of use and communication.** The practical adoption of a method depends not only on its statistical properties but also on how readily it can be implemented, automated, and explained to decision makers. Simpler tools such as ARIMA and exponential smoothing remain popular in industry because they are robust, easy to automate, and easy to explain. Structural models retain influence in policy institutions partly because their outputs can be communicated in terms of familiar economic concepts (Smets & Wouters, 2007). As more complex methods enter applied practice, maintaining a clear link between modeling choices and their rationale becomes ever more pressing, a theme discussed further in Sections 12 and 15.

No single method dominates across all dimensions. The 'best' approach depends on the purpose (forecasting, explanation, or scenario analysis), the data at hand (sample size, number of variables, presence of structural change), and practical considerations (interpretability, computational resources, audience). A key skill for the applied econometrician is knowing the available methods and matching the right one, or the right combination, to the task.

15 Emerging Trends and Future Challenges

What challenges and opportunities lie ahead? The discipline has adapted continuously to new data environments, new decision problems, and the empirical shortcomings of earlier models. Several developments now seem most likely to reshape practice over the coming decade.

The most consequential trend is probably the continued proliferation of data sources. Econometric analysis now reaches beyond traditional aggregates to high-frequency transaction data, text from news and social media, satellite imagery, mobility indicators, and other sources available in near real time. Bringing these varied inputs into time series models in a principled way will require further work on data-rich methods, including dynamic factor and mixed-frequency models, and more systematic use of regularization for high-dimensional feature sets (Stock & Watson, 2002; Banbura et al., 2010; Ghysels et al., 2004). The 'text as data' agenda, which maps unstructured language into time-varying indices suitable for combination with conventional predictors, is a prominent example (Gentzkow et al., 2019). Data proliferation also means more granularity, with hundreds of micro indicators rather than a single GDP series. That blurs the boundary between time series and panel analysis and encourages methods that borrow strength across related series, including hierarchical reconciliation and panel time series tools (Hyndman et al., 2011).

A closely related challenge is real-time adaptability. The COVID-19 recession made clear that models trained on historical regularities can fail when confronted with shocks far outside the range of past experience. Standard models struggled not because their statistical apparatus was wrong, but because previously stable relationships changed abruptly; pandemic-era observations were extreme enough to distort estimated dynamics, prompting practical fixes that downweight or explicitly model the outliers (Lenza & Primiceri, 2022; Foroni, Marcellino, & Stevanovic, 2022). The pandemic renewed interest in adapting established structural-break and time-varying-parameter methods for real-time monitoring (Andrews, 1993; J. Bai & Perron, 1998; Cogley & Sargent, 2005; Primiceri, 2005). Beyond sudden shocks, slow-moving structural trends such as productivity slowdowns, demographic shifts, and changes in the inflation process can gradually erode a model's forecasting performance. Distinguishing whether such low-frequency change is permanent or transitory remains difficult, and robust inference about long-run objects under persistent dynamics is likely to remain an active research area (Müller & Watson, 2015).

Causal inference in time series settings is another frontier. The broader econometric emphasis on credible research designs, such as natural experiments, instrumental variables, and difference-in-differences, has only partly penetrated aggregate time series analysis. One reason is that many macroeconomic questions rely on observational data with strong temporal dependence. Promising directions include the use of high-frequency or external instruments for identifying structural shocks in VARs, as in the proxy-SVAR approach (Mertens & Ravn, 2013; Gertler & Karadi, 2015; Stock & Watson, 2018), and local projection methods that accommodate nonlinearities and state dependence more naturally than standard VAR impulse responses (Jordà, 2005). Machine learning tools are also being adapted for causal targets, including double machine learning for valid inference on low-dimensional parameters in the presence of high-dimensional nuisance components (Chernozhukov et al., 2018), and synthetic control methods for policy evaluation (Abadie et al., 2010). Extending these methods to settings with temporal dependence and feedback remains challenging and is a priority for policy analysis.

Explainability and transparency matter more as complex models spread in forecasting and policy analysis. Time series econometricians can contribute by designing predictive methods that explicitly represent trend, cycle, seasonality, and calendar effects, so that the contribution of each component to a forecast can be explained clearly (Taylor & Letham, 2018). Shapley-value decompositions explain individual predictions, while permutation importance assesses the model's overall predictive reliance on each feature (Lundberg & Lee, 2017; Breiman, 2001). More broadly, improved visualization and communication tools will be needed to convey results of complex analyses, particularly when forecast objects are high-dimensional or when scenario analyses involve many moving parts. History offers a precedent: Box–Jenkins ARIMA modeling succeeded in part because it came with a clear, teachable diagnostic workflow (Box & Jenkins, 1970). Methods that are difficult to explain risk being ignored regardless of their technical sophistication.

Computational advances will continue to expand the feasible model space. Simulation-based inference methods, including approximate Bayesian computation, may become more common for models too complex to yield tractable likelihoods (Beaumont et al., 2002; Grazzini et al., 2017). Automation is also expanding through time-series variants of AutoML, in which suites of models are fit, tuned, and compared under standardized out-of-sample protocols. In this environment, the econometrician's role shifts toward ensuring that automation is statistically sound: using appropriate time-series validation schemes (Bergmeir et al., 2018), adjusting for multiple comparisons when many models are searched (White, 2000; Hansen, 2005), and maintaining economic plausibility alongside predictive accuracy. Forecasting competitions continue to evolve toward decision-focused evaluation that links predictive accuracy to downstream utility

(Makridakis et al., 2025).

Interdisciplinary engagement will also deepen. Time series econometrics shares methodological concerns with climate science, epidemiology, and network science, all of which confront problems of trend versus cycle, structural change, and high-dimensional dependence. Econometric tools such as cointegration, state-space filtering, and formal forecast evaluation apply directly in these domains, while insights from domain sciences, such as epidemiological compartment models and climate physics constraints, can serve as structural priors within econometric frameworks. At the intersection of economics and finance, high-frequency data from algorithmic trading raise new modeling challenges for irregularly spaced observations, motivating further development of duration models and point-process approaches (Engle & Russell, 1998; Hawkes, 1971).

Finally, the growing use of granular and individual-level data raises privacy and ethical questions. Federated learning trains models across institutions without pooling raw microdata (McMahan, Moore, Ramage, Hampson, & Agüera y Arcas, 2017), while differential privacy limits what any released estimate or update reveals about a single record, in centralized or decentralized settings (Dwork & Roth, 2014). The two are complementary rather than interchangeable, and are often combined. Such safeguards will increasingly shape which data can be used and how models are estimated.

16 Conclusion: Enduring Lessons to Remain Relevant

The history of time series econometrics, from early empirical studies and probability foundations to contemporary data-rich practice, is one of repeated adaptation. Four broad lessons stand out.

- **Respect the data's properties.** A foundational lesson is 'know thy data.' Yule (1927) and Slutsky (1937) showed that apparent regularities can arise from underlying stochastic processes; Granger and Newbold (1974) demonstrated that ignoring nonstationarity leads to spurious regression. The field responded with unit root tests (Dickey & Fuller, 1979), cointegration analysis (Engle & Granger, 1987), and ARCH models for time-varying volatility (Engle, 1982). The consistent lesson is that methods must be tailored to the characteristics of the data-generating process, including trends, seasonality, breaks, and heteroskedasticity, and that routine diagnostic checks remain essential before settling on a specification (Andrews, 1993; J. Bai & Perron, 1998).

- **Balance parsimony with structural insight.** Econometricians have learned to navigate the tension between data-driven flexibility and theory-driven discipline. Simpler models often generalize better, a principle articulated in the Box–Jenkins tradition (Box & Jenkins, 1970) and revived in modern regularization (Tibshirani, 1996) and Bayesian shrinkage (Litterman, 1986). At the same time, economic structure has not been abandoned. It re-emerged through cointegration (Engle & Granger, 1987) and even in VARs. Sims (1980) introduced them as a data-driven alternative to large structural models, but structural VARs later brought theory back through identifying restrictions. Purely mechanical fitting risks chasing noise; purely theoretical modeling can miss empirically salient features. The most productive work blends both.

- **Quantify uncertainty and resist overclaiming.** No specification is definitive, and econometricians have built that caution into their tools. Forecast intervals, fan charts, density forecast evaluation, and formal out-of-sample comparison provide safeguards against overstating what models can deliver (Diebold & Mariano, 1995; Diebold et al., 1998; Gneiting & Raftery, 2007). The empirical success of forecast combinations shows that pooling across diverse models can be more robust than selection when model uncertainty is pervasive (Bates & Granger, 1969; Timmermann, 2006).

- **Adapt continuously.** Time series econometrics has repeatedly renewed itself. When traditional macro models faltered in the 1970s, ARIMA and VAR methods gained prominence (Box & Jenkins, 1970; Sims, 1980). In the 1980s, unit roots and cointegration forced economists to rethink strategies for trending data (Dickey & Fuller, 1979; Engle & Granger, 1987). Financial risk management needs spurred volatility models in the 1990s (Engle, 1982; Bollerslev, 1986). High-dimensional data in the 2000s motivated factor models and shrinkage (Stock & Watson, 2002; Banbura et al., 2010), and the rise of machine learning has prompted integration of new tools alongside established evaluation principles (Makridakis et al., 2020; Chernozhukov et al., 2018).

Expanding data availability, computational scale, and demands for credible causal interpretation will continue to shape time series econometrics. New methods have lasted when they were absorbed into the field's established principles, which combine a deep understanding of temporal dependence, equilibrium concepts, and economic context

with rigorous evaluation of uncertainty. That combination, rather than any particular model class, has kept the discipline relevant.

Acknowledgments

I am grateful for valuable suggestions provided by David Harris and Farshid Vahid.

Declaration on the Use of Generative AI in the Writing Process

During the preparation of this chapter, the author used Claude to improve the readability of the text. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the chapter.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, *105*(490), 493–505. doi: 10.1198/jasa.2009.ap08746

Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *American Economic Review*, *109*(4), 1263–1289.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*(2), 579–625.

Anderson, T. W., & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, *20*(1), 46–63.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*(3), 817–858.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, *61*(4), 821–856.

Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., ... Wang, Y. (2024). Chronos: Learning the language of time series. *Transactions on Machine Learning Research*. Retrieved from https://arxiv.org/abs/2403.07815 (Also available as arXiv:2403.07815)

Antolín-Díaz, J., & Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for SVARs. *American Economic Review*, *108*(10), 2802–2829. doi: 10.1257/aer.20161852

Ardia, D., Bluteau, K., & Boudt, K. (2019). Questioning the news about economic growth: Sparse forecasting using thousands of news-based sentiment values. *International Journal of Forecasting*, *35*(4), 1370–1386.

Arellano, M., & Bond, S. (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.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, *68*(1), 29–51.

Athey, S., & Imbens, G. W. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360. doi: 10.1073/pnas.1510489113

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, *71*(1), 135–171. doi: 10.1111/1468-0262.00392

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221. doi: 10.1111/1468-0262.00273

Bai, J., & Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, *25*(1), 52–60.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, *66*(1), 47–78.

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, *73*(1), 5–59. doi: 10.1016/0304-4076(95)01732-1

Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*(1), 3–30. doi: 10.1016/S0304-4076(95)01749-6

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636.

Balke, N. S., & Fomby, T. B. (1997). Threshold cointegration. *International Economic Review*, *38*(3), 627–645. doi: 10.2307/2527284

Bañbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). *Now-casting and the real-time data flow* (Working Paper Series No. 1564). Frankfurt am Main, Germany: European Central Bank. Retrieved from https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1564.pdf

Banbura, M., Giannone, D., & Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71–92.

Banerjee, A., Dolado, J. J., Galbraith, J. W., & Hendry, D. F. (1993). *Co-integration, error correction, and the econometric analysis of non-stationary data*. Oxford: Oxford University Press.

Barber, R. F., Candès, E. J., Ramdas, A., & Tibshirani, R. J. (2023). Conformal prediction beyond exchangeability. *The Annals of Statistics*, *51*(2), 816–845. doi: 10.1214/23-AOS2276

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(2), 253–280. doi: 10.1111/1467-9868.00336

Basmann, R. L. (1957). A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica*, *25*(1), 77–83.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, *20*(4), 451–468.

Baumeister, C., & Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica*, *83*(5), 1963–1999.

Baxter, M., & King, R. G. (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics*, *81*(4), 575–593. doi: 10.1162/003465399558454

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate bayesian computation in population genetics. *Genetics*, *162*(4), 2025–2035. doi: 10.1093/genetics/162.4.2025

Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, Y., Maddix, D. C., Türkmen, C., ... Januschowski, T. (2022). Deep learning for time series forecasting: Tutorial and literature survey. *ACM Computing Surveys*, *55*(6), 1–36. doi: 10.1145/3533382

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, *120*, 70–83. doi: 10.1016/j.csda.2017.11.003

Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, *120*(1), 387–422. doi: 10.1162/0033553053327452

Beveridge, S., & Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics*, *7*(2), 151–174. doi: 10.1016/0304-3932(81)90040-4

Beveridge, W. H. (1922). Wheat prices and rainfall in western Europe. *Journal of the Royal Statistical Society*, *85*(3), 412–475.

Blanchard, O. J., & Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, *79*(4), 655–673.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. doi: 10.1080/01621459.2017.1285773

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

Boivin, J., & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, *132*(1), 169–194.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, *72*(3), 498–505.

Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, *11*(2), 143–172.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Box, G. E. P., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, *65*(332), 1509–1526. doi: 10.1080/01621459.1970.10481180

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, *70*(349), 70–79. doi: 10.1080/01621459.1975.10480264

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth.

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, *17*(31), 334–355.

Britton, E., Fisher, P., & Whitley, J. (1998). The Inflation Report projections: Understanding the fan chart. *Bank of England Quarterly Bulletin*, *38*(1), 30–37.

Brown, R. G. (1959). *Statistical forecasting for inventory control*. New York: McGraw-Hill.

Burns, A. F., & Mitchell, W. C. (1946). *Measuring business cycles*. New York: National Bureau of Economic Research.

Campbell, J. Y., & Mankiw, N. G. (1987). Are output fluctuations transitory? *The Quarterly Journal of Economics*, *102*(4), 857–880.

Campbell, J. Y., & Shiller, R. J. (1987). Cointegration and tests of present value models. *Journal of Political Economy*, *95*(5), 1062–1088.

Canova, F., & De Nicoló, G. (2002). Monetary disturbances matter for business fluctuations in the G-7. *Journal of Monetary Economics*, *49*(6), 1131–1159.

Carriero, A., Clark, T. E., & Marcellino, M. (2015). Bayesian VARs: Specification choices and forecast accuracy. *Journal of Applied Econometrics*, *30*(1), 46–73.

Carter, C. K., & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, *81*(3), 541–553.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21*(1), C1–C68.

Choi, H., & Varian, H. R. (2012). Predicting the present with Google trends. *Economic Record*, *88*(s1), 2–9. doi: 10.1111/j.1475-4932.2012.00809.x

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, *28*(3), 591–605.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? In J. B. Taylor & M. Woodford (Eds.), *Handbook of macroeconomics* (Vol. 1A, pp. 65–148). Amsterdam: Elsevier.

Christiano, L. J., & Fitzgerald, T. J. (2003). The band pass filter. *International Economic Review*, *44*(2), 435–465.

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, *138*(1), 291–311.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3–73.

Cochrane, D., & Orcutt, G. H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association*, *44*(245), 32–61. doi: 10.1080/01621459.1949.10483290

Cochrane, J. H. (1988). How big is the random walk in GNP? *Journal of Political Economy*, *96*(5), 893–920.

Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, *8*(2), 262–302.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Cramér, H. (1942). On harmonic analysis in certain functional spaces. *Arkiv för Matematik, Astronomi och Fysik*, *28B*(12). (Article 17)

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795. doi: 10.1002/jae.1279

Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of Economic Literature*, *49*(1), 72–100. doi: 10.1257/jel.49.1.72

Davidson, J. E. H., Hendry, D. F., Srba, F., & Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *The Economic Journal*, *88*(352), 661–692.

Dawid, A. P. (1984). Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *Journal of the Royal Statistical Society: Series A (General)*, *147*(2), 278–292.

Dées, S., di Mauro, F., Pesaran, M. H., & Smith, L. V. (2007). Exploring the international linkages of the euro area: A global VAR analysis. *Journal of Applied Econometrics*, *22*(1), 1–38.

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.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business & Economic Statistics*, *33*(1), 1–9.

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863–883.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

Diebold, F. X., & Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, *119*(534), 158–171. doi: 10.1111/j.1468-0297.2008.02208.x

Diebold, F. X., & Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, *28*(1), 57–66. doi: 10.1016/j.ijforecast.2011.02.006

Doucet, A., de Freitas, N., & Gordon, N. J. (Eds.). (2001). *Sequential Monte Carlo methods in practice*. New York: Springer.

Doz, C., & Fuleky, P. (2020). Dynamic factor models. In P. Fuleky (Ed.), *Macroeconomic forecasting in the era of big data: Theory and practice* (Vol. 52, pp. 27–64). Cham: Springer.

Doz, C., Giannone, D., & Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, *164*(1), 188–205. doi: 10.1016/j.jeconom.2011.02.012

Doz, C., Giannone, D., & Reichlin, L. (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models. *The Review of Economics and Statistics*, *94*(4), 1014–1024. doi: 10.1162/REST_a_00225

Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*. Oxford: Oxford University Press.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika*, *37*(3-4), 409–428. doi: 10.1093/biomet/37.3-4.409

Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression: II. *Biometrika*, *38*(1-2), 159–178. doi: 10.1093/biomet/38.1-2.159

Dwork, C., & Roth, A. (2014). *The algorithmic foundations of differential privacy*. now Publishers Inc. doi: 10.1561/0400000042

Elliott, G. (1999). Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. *International Economic Review*, *40*(3), 767–783.

Elliott, G., Gargano, A., & Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, *177*(2), 357–373.

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, *64*(4), 813–836. doi: 10.2307/2171846

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica*, *50*(4), 987–1008.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, *55*(2), 251–276.

Engle, R. F., & Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, *11*(1), 122–150.

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367–381.

Engle, R. F., & Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, *66*(5), 1127–1162. doi: 10.2307/2999632

Fair, R. C., & Shiller, R. J. (1990). Comparing information in forecasts from econometric models. *American Economic Review*, *80*(3), 375–389.

Faust, J. (1998). The robustness of identified VAR conclusions about money. *Carnegie-Rochester Conference Series on Public Policy*, *49*, 207–244.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, *1*(2), 209–230.

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., & Chen, B.-C. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business & Economic Statistics*, *16*(2), 127–152. doi: 10.1080/07350015.1998.10524743

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation. *The Review of Economics and Statistics*, *82*(4), 540–554. doi: 10.1162/003465300559037

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, *100*(471), 830–840. doi: 10.1198/016214504000002050

Foroni, C., Marcellino, M., & Stevanovic, D. (2022). Forecasting the Covid-19 recession and recovery: Lessons from the financial crisis. *International Journal of Forecasting*, *38*(2), 596–612.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. doi: 10.1214/aos/1013203451

Friedman, M., & Schwartz, A. J. (1963). *A monetary history of the united states, 1867–1960*. Princeton, NJ: Princeton University Press. (National Bureau of Economic Research, Studies in Business Cycles No. 12)

Frisch, R. (1933). Propagation problems and impulse problems in dynamic economics. In *Economic essays in honour of gustav cassel* (pp. 171–205). London: George Allen & Unwin.

Fuleky, P. (Ed.). (2020). *Macroeconomic forecasting in the era of big data: Theory and practice* (Vol. 52). Cham: Springer.

Garza, A., Challu, C., & Mergenthaler-Canseco, M. (2023). Timegpt-1. *arXiv preprint*. Retrieved from https://arxiv.org/abs/2310.03589 doi: 10.48550/arXiv.2310.03589

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*(410), 398–409.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574. doi: 10.1257/jel.20181020

Gertler, M., & Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, *7*(1), 44–76.

Geweke, J., & Amisano, G. (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*(2), 216–230. doi: 10.1016/j.ijforecast.2009.10.007

Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, *4*(4), 221–238.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). *The MIDAS touch: Mixed data sampling regression models* (Scientific Series No. 2004s-20). Montreal, QC, Canada: CIRANO. Retrieved from https://www.cirano.qc.ca/files/publications/2004s-20.pdf

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: Getting the most out of return data by mixing frequencies. *Journal of Econometrics*, *131*(1–2), 59–95.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). Midas regressions: Further results and new directions. *Econometric Reviews*, *26*(1), 53–90. doi: 10.1080/07474930600972467

Giacomini, R., & Rossi, B. (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies*, *76*(2), 669–705. doi: 10.1111/j.1467-937X.2009.00545.x

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.

Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, *97*(2), 436–451. doi: 10.1162/REST_a_00483

Giannone, D., Lenza, M., & Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, *89*(5), 2409–2437.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data releases. *Journal of Monetary Economics*, *55*(4), 665–676.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, *46*(6), 1293–1301.

Gómez, V., & Maravall, A. (1996). *Programs TRAMO and SEATS: Instructions for the user (beta version: September 1996)* (Working Paper No. 9628). Madrid: Banco de España.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, *37*(5), 920–964.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438.

Granger, C. W. J., & Hatanaka, M. (1964). *Spectral analysis of economic time series*. Princeton, NJ: Princeton University Press.

Granger, C. W. J., & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, *1*(1), 15–29.

Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, *2*(2), 111–120.

Granger, C. W. J., & Newbold, P. (1977). *Forecasting economic time series*. New York: Academic Press.

Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, *3*(2), 197–204.

Granger, C. W. J., & Teräsvirta, T. (1993). *Modelling nonlinear economic relationships*. Oxford: Oxford University Press.

Grazzini, J., Richiardi, M. G., & Tsionas, M. (2017). Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control*, *77*, 26–47. doi: 10.1016/j.jedc.2017.01.014

Gregory, A. W., & Hansen, B. E. (1996). Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics*, *70*(1), 99–126.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, *12*(Supplement), iii–115.

Hallin, M., & Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, *102*(478), 603–617.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*(2), 357–384.

Hamilton, J. D. (2018). Why you should never use the hodrick–prescott filter. *The Review of Economics and Statistics*, *100*(5), 831–843.

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, *41*(2), 190–195.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, *23*(4), 365–380.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Harvey, A. C. (1989). *Forecasting, structural time series and the kalman filter*. Cambridge: Cambridge University Press.

Harvey, A. C., & Jaeger, A. (1993). Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics*, *8*(3), 231–247. doi: 10.1002/jae.3950080302

Harvey, D. I., Leybourne, S. J., & Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business & Economic Statistics*, *16*(2), 254–259.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83–90. doi: 10.1093/biomet/58.1.83

Hendry, D. F. (1995). *Dynamic econometrics*. Oxford: Oxford University Press.

Hendry, D. F., & Morgan, M. S. (Eds.). (1995). *The foundations of econometric analysis*. Cambridge: Cambridge University Press.

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, *37*(1), 388–427. doi: 10.1016/j.ijforecast.2020.06.008

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Hodrick, R. J., & Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, *29*(1), 1–16. doi: 10.2307/2953682

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*(4), 382–417.

Holt, C. C. (1957). *Forecasting seasonals and trends by exponentially weighted moving averages* (ONR Research Memorandum No. 52). Pittsburgh: Carnegie Institute of Technology.

Hoover, K. D., & Perez, S. J. (1999). Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *The Econometrics Journal*, *2*(2), 167–191. doi: 10.1111/1368-423X.00025

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, *68*(1), 165–176.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*(1), 73–101.

Hylleberg, S., Engle, R. F., Granger, C. W. J., & Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, *44*(1–2), 215–238. doi: 10.1016/0304-4076(90)90080-D

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, *55*(9), 2579–2589. doi: 10.1016/j.csda.2011.03.006

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *27*(3). Retrieved from https://www.jstatsoft.org/article/view/v027i03 doi: 10.18637/jss.v027.i03

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. D. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*(3), 439–454. doi: 10.1016/S0169-2070(01)00110-8

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, *115*(1), 53–74. doi: 10.1016/S0304-4076(03)00092-7

Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, *12*(4), 371–417.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2–3), 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, *59*(6), 1551–1580.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.

Jordà, O. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, *95*(1), 161–182.

Jordà, Ò., & Taylor, A. M. (2025). Local projections. *Journal of Economic Literature*, *63*(1), 59–110.

Jorion, P. (2007). *Value at risk: The new benchmark for managing financial risk* (3rd ed.). New York: McGraw-Hill.

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, *105*(3), 1177–1216.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, *82*(Series D), 35–45.

Keynes, J. M. (1939). Professor Tinbergen's method. *The Economic Journal*, *49*(195), 558–577.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics*, *80*(2), 218–230.

Kilian, L., & Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge: Cambridge University Press. doi: 10.1017/9781108164818

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, *60*(1-2), 1–22.

Kim, C.-J., & Nelson, C. R. (1999a). Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *The Review of Economics and Statistics*, *81*(4), 608–616.

Kim, C.-J., & Nelson, C. R. (1999b). *State-space models with regime switching: Classical and gibbs-sampling approaches with applications*. Cambridge, MA: MIT Press.

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, *65*(3), 361–393.

Klein, L. R. (1950). *Economic fluctuations in the united states, 1921–1941*. New York: John Wiley & Sons. (Cowles Commission for Research in Economics, Monograph No. 11)

Klein, L. R., & Goldberger, A. S. (1955). *An econometric model of the united states, 1929–1952*. Amsterdam: North-Holland.

Kock, A. B., & Callot, L. (2015). Oracle inequalities for high-dimensional vector autoregressions. *Journal of Econometrics*, *186*(2), 325–344. doi: 10.1016/j.jeconom.2015.02.013

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.

Koenker, R., & Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, *101*(475), 980–990.

Kolmogoroff, A. (1933). *Grundbegriffe der wahrscheinlichkeitsrechnung* (Vol. 2). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-49888-6

Kolmogorov, A. N. (1941). Interpolation und Extrapolation von stationären zufälligen Folgen. *Izvestiya Akademii Nauk SSSR, Seriya Matematicheskaya*, *5*(1), 3–14.

Koopmans, T. C. (1947). Measurement without theory. *The Review of Economic Statistics*, *29*(3), 161–172.

Koopmans, T. C. (Ed.). (1950). *Statistical inference in dynamic economic models*. New York: John Wiley & Sons. (Cowles Commission for Research in Economics, Monograph No. 10)

Koyck, L. M. (1954). *Distributed lags and investment analysis*. Amsterdam: North-Holland.

Krolzig, H.-M., & Hendry, D. F. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*, *25*(6–7), 831–866. doi: 10.1016/S0165-1889(00)00058-0

Kuan, C.-M., & White, H. (1994). Artificial neural networks: An econometric perspective. *Econometric Reviews*, *13*(1), 1–91.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, *17*(3), 1217–1241. doi: 10.1214/aos/1176347265

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.

Lazarus, E., Lewis, D. J., Stock, J. H., & Watson, M. W. (2018). HAR inference: Recommendations for practice. *Journal of Business & Economic Statistics*, *36*(4), 541–559.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. doi: 10.1016/S0047-259X(03)00096-4

Lenza, M., & Primiceri, G. E. (2022). How to estimate a vector autoregression after March 2020. *Journal of Applied Econometrics*, *37*(4), 688–699.

Levin, A., Lin, C.-F., & Chu, C.-S. J. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics*, *108*(1), 1–24. doi: 10.1016/S0304-4076(01)00098-7

Li, D., Plagborg-Møller, M., & Wolf, C. K. (2024). Local projections vs. VARs: Lessons from thousands of DGPs. *Journal of Econometrics*, *244*(2), 105722.

Lim, B., Arik, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748–1764. doi: 10.1016/j.ijforecast.2021.03.012

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions—five years of experience. *Journal of Business & Economic Statistics*, *4*(1), 25–38.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303. doi: 10.1093/biomet/65.2.297

Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, *1*, 19–46.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems* (Vol. 30, pp. 4765–4774).

Makridakis, S., Petropoulos, F., & Spiliotis, E. (2022). The M5 competition: Conclusions. *International Journal of Forecasting*, *38*(4), 1576–1582. doi: 10.1016/j.ijforecast.2022.04.006

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74. doi: 10.1016/j.ijforecast.2019.04.014

Makridakis, S., Spiliotis, E., Hollyman, R., Petropoulos, F., Swanson, N. R., & Gaba, A. (2025). The M6 forecasting competition: Bridging the gap between forecasting and investment decisions. *International Journal of Forecasting*, *41*(4), 1315–1354. doi: 10.1016/j.ijforecast.2024.11.002

Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, *18*(4), 427–443.

Martin, V. L., Hurn, A. S., & Harris, D. (2013). *Econometric modelling with time series: Specification, estimation and testing*. New York: Cambridge University Press.

Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of Economic Surveys*, *37*(1), 76–111.

McConnell, M. M., & Pérez-Quirós, G. (2000). Output fluctuations in the united states: What has changed since the early 1980s? *American Economic Review*, *90*(5), 1464–1476.

McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.

McMahan, B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th international conference on artificial intelligence and statistics* (Vol. 54, pp. 1273–1282). PMLR. Retrieved from https://proceedings.mlr.press/v54/mcmahan17a.html

Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, *39*(1), 98–119.

Mertens, K., & Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States. *American Economic Review*, *103*(4), 1212–1247.

Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In J. A. Mincer (Ed.), *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3–46). Cambridge, MA: National Bureau of Economic Research, Inc. Retrieved from https://www.nber.org/system/files/chapters/c1214/c1214.pdf

Morgan, M. S. (1990). *The history of econometric ideas*. Cambridge: Cambridge University Press.

Morley, J. C., Nelson, C. R., & Zivot, E. (2003). Why are the beveridge–nelson and unobserved-components decompositions of GDP so different? *The Review of Economics and Statistics*, *85*(2), 235–243. doi: 10.1162/003465303765299765

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106. doi: 10.1257/jep.31.2.87

Müller, U. K., & Watson, M. W. (2015). *Low-frequency econometrics* (NBER Working Paper No. 21564). National Bureau of Economic Research. (Revised November 2016) doi: 10.3386/w21564

Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, *55*(290), 299–306.

Neal, R. M. (2011). Mcmc using hamiltonian dynamics. In S. Brooks, A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *Handbook of markov chain monte carlo* (pp. 113–162). Boca Raton, FL: Chapman and Hall/CRC. doi: 10.1201/b10905

Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the U.S. economy. *American Economic Review*, *62*(5), 902–917.

Nelson, C. R., & Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, *10*(2), 139–162.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347–370.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Ng, S., & Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, *69*(6), 1519–1554. doi: 10.1111/1468-0262.00256

Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. In *International conference on learning representations.*

Ord, J. K., Koehler, A. B., & Snyder, R. D. (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association*, *92*(440), 1621–1629. doi: 10.1080/01621459.1997.10473684

Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*.

Orphanides, A., & van Norden, S. (2002). The unreliability of output-gap estimates in real time. *The Review of Economics and Statistics*, *84*(4), 569–583.

Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, *61*(S1), 653–670. doi: 10.1111/1468-0084.61.s1.14

Pedroni, P. (2004). Panel cointegration: Asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econometric Theory*, *20*(3), 597–625. doi: 10.1017/S0266466604203073

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, *57*(6), 1361–1401.

Persons, W. M. (1919). Indices of business conditions. *The Review of Economic Statistics*, *1*(1), 5–107.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, *22*(2), 265–312. doi: 10.1002/jae.951

Pesaran, M. H., Schuermann, T., & Weiner, S. M. (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics*, *22*(2), 129–162.

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, *16*(3), 289–326. doi: 10.1002/jae.616

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, *33*(3), 311–340.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica*, *55*(2), 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika*, *74*(3), 535–547. doi: 10.1093/biomet/74.3.535

Phillips, P. C. B., & Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies*, *57*(1), 99–125.

Phillips, P. C. B., & Ouliaris, S. (1990). Asymptotic properties of residual-based tests for cointegration. *Econometrica*, *58*(1), 165–193.

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*(2), 335–346.

Phillips, P. C. B., Shi, S., & Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, *56*(4), 1043–1078.

Phillips, P. C. B., Wu, Y., & Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review*, *52*(1), 201–226.

Plagborg-Møller, M., & Wolf, C. K. (2021). Local projections and VARs estimate the same impulse responses. *Econometrica*, *89*(2), 955–980.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*(428), 1303–1313. doi: 10.1080/01621459.1994.10476870

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, *72*(3), 821–852. doi: 10.1111/j.1467-937X.2005.00353.x

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, *55*(290), 324–330.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135–1144). ACM. doi: 10.1145/2939672.2939778

Rigobon, R. (2003). Identification through heteroskedasticity. *The Review of Economics and Statistics*, *85*(4), 777–792. doi: 10.1162/003465303772815727

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, *23*(5), 1630–1661.

Romer, C. D., & Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, *94*(4), 1055–1084. doi: 10.1257/0002828042002651

Rossi, B. (2021). Forecasting in the presence of instabilities: How we know whether models predict well and how to improve them. *Journal of Economic Literature*, *59*(4), 1135–1190.

Rubio-Ramírez, J. F., Waggoner, D. F., & Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies*, *77*(2), 665–696.

Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, *71*(3), 599–607.

Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181–1191.

Sargan, J. D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology. In P. E. Hart, G. Mills, & J. K. Whitaker (Eds.), *Econometric analysis for national economic planning* (Vol. 16, pp. 25–54). London: Butterworth. (Reprinted as Ch. 10 in D. F. Hendry and K. F. Wallis (Eds.), *Econometrics and Quantitative Economics*, Oxford: Basil Blackwell, 1984)

Schorfheide, F., & Song, D. (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics*, *33*(3), 366–380.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464.

Schwert, G. W. (1989). Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics*, *7*(2), 147–159.

Shiskin, J., Young, A. H., & Musgrave, J. C. (1967). *The X-11 variant of the census method II seasonal adjustment program* (Technical Paper No. 15). Washington, DC: U.S. Bureau of the Census.

Sims, C. A. (1972). Money, income, and causality. *American Economic Review*, *62*(4), 540–552.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, *48*(1), 1–48.

Sims, C. A., Stock, J. H., & Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica*, *58*(1), 113–144. doi: 10.2307/2938337

Slutsky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica*, *5*(2), 105–146.

Smets, F., & Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, *1*(5), 1123–1175.

Smets, F., & Wouters, R. (2007). Shocks and frictions in US business cycles: A bayesian DSGE approach. *American Economic Review*, *97*(3), 586–606.

Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, *36*(1), 75–85.

Song, Y. (2014). Modelling regime switching and structural breaks with an infinite hidden Markov model. *Journal of Applied Econometrics*, *29*(5), 825–842.

Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, *55*(5), 1035–1056.

Stock, J. H., & Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association*, *83*(404), 1097–1107.

Stock, J. H., & Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, *61*(4), 783–820.

Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, *15*(4), 101–115. doi: 10.1257/jep.15.4.101

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162.

Stock, J. H., & Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, *39*(s1), 3–33.

Stock, J. H., & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, *128*(610), 917–948.

Swanson, N. R., & White, H. (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *The Review of Economics and Statistics*, *79*(4), 540–550.

Taylor, S. J. (1986). *Modelling financial time series*. Chichester: Wiley.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37–45. doi: 10.1080/00031305.2017.1380080

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, *89*(425), 208–218.

Theil, H. (1953). *Repeated least squares applied to complete equation systems* (Memorandum). The Hague: Central Planning Bureau.

Theil, H. (1961). *Economic forecasts and policy* (2nd ed.). Amsterdam: North-Holland.

Theil, H. (1966). *Applied economic forecasting*. Amsterdam: North-Holland.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B*, *58*(1), 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 1, pp. 135–196). Amsterdam: Elsevier.

Tinbergen, J. (1939). *Statistical testing of business-cycle theories, vol. I: A method and its application to investment activity; vol. II: Business cycles in the united states of america, 1919–1932*. Geneva: League of Nations.

Tong, H. (1983). *Threshold models in non-linear time series analysis*. New York: Springer.

Uhlig, H. (2005). What are the effects of monetary policy? results from an agnostic identification procedure. *Journal of Monetary Economics*, *52*(2), 381–419.

Vahid, F., & Engle, R. F. (1993). Common trends and common cycles. *Journal of Applied Econometrics*, *8*(4), 341–360.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28. doi: 10.1257/jep.28.2.3

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (Vol. 30, pp. 5998–6008).

Vining, R. (1949). Koopmans on the choice of variables to be studied and of methods of measurement. *The Review of Economics and Statistics*, *31*(2), 77–86.

Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world*. New York: Springer. doi: 10.1007/b106715

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. doi: 10.1080/01621459.2017.1319839

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, *64*(5), 1067–1084.

White, H. (2000). A reality check for data snooping. *Econometrica*, *68*(5), 1097–1126.

Whittle, P. (1953). The analysis of multiple stationary time series. *Journal of the Royal Statistical Society. Series B (Methodological)*, *15*(1), 125–139. doi: 10.1111/j.2517-6161.1953.tb00131.x

Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series: With engineering applications*. Cambridge, MA: Technology Press of the Massachusetts Institute of Technology and John Wiley & Sons.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, *6*(3), 324–342. doi: 10.1287/mnsc.6.3.324

Wold, H. (1938). *A study in the analysis of stationary time series*. Uppsala: Almqvist & Wiksells Boktryckeri-A.-B.

Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in neural information processing systems.*

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? a study in sampling and the nature of time series. *Journal of the Royal Statistical Society*, *89*(1), 1–63.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A*, *226*, 267–298.

Zaffran, M., Feron, O., Goude, Y., Josse, J., & Dieuleveut, A. (2022). Adaptive conformal predictions for time series. In *Proceedings of the 39th international conference on machine learning* (Vol. 162, pp. 25834–25866). PMLR. Retrieved from https://proceedings.mlr.press/v162/zaffran22a.html

Zellner, A. (1971). *An introduction to bayesian inference in econometrics*. New York: Wiley.

Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 37, pp. 11121–11128).

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence.*

Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics*, *10*(3), 251–270.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429. doi: 10.1198/016214506000000735

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, *67*(2), 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

Index

adaptive lasso, 19
Akaike information criterion (AIC), 5
approximate Bayesian computation (ABC), 16
augmented Dickey–Fuller (ADF) test, 8
autocorrelation function (ACF), 5
autoregressive (AR) model, 3
autoregressive conditional heteroskedasticity (ARCH), 11
Autoregressive Distributed Lag (ARDL), 10
autoregressive integrated moving-average (ARIMA), 5

band-pass filter, 14
Bayesian information criterion (BIC), 5
Bayesian model averaging (BMA), 16
Bayesian VAR (BVAR), 15
Beveridge–Nelson decomposition, 14
block bootstrap, 23
boosting, 20
bubble detection, 9

causal inference, 31
Cholesky identification, 6
Chow test, 13
cointegration, 9
common features, 10
complete subset regression, 19
conformal prediction, 22
connectedness measure, 20
continuous ranked probability score (CRPS), 24
Cowles Commission, 4
cross-sectional dependence, 9
cross-validation, 19

deep learning, 21
Dickey–Fuller test, 8
Diebold–Mariano test, 23
differential privacy, 32
diffusion index, 17
distributed-lag model, 4
double machine learning, 31
duration model, 32
Durbin–Watson test, 4
dynamic conditional correlation, 12
dynamic factor model (DFM), 17
dynamic OLS (DOLS), 10
dynamic stochastic general equilibrium (DSGE) model, 16

elastic net, 19
error-correction model (ECM), 9
exponential smoothing, 13

factor-augmented VAR (FAVAR), 18
fan chart, 25
federated learning, 32
forecast combination, 17
forecast encompassing, 23
forecast error variance decomposition (FEVD), 6
foundation model, 21
fractional integration, 8
fractionally integrated ARIMA (ARFIMA), 8
fractionally integrated GARCH (FIGARCH), 12
fully modified OLS (FM-OLS), 10

general-to-specific modeling, 7
generalized autoregressive conditional heteroskedasticity (GARCH), 11
Generalized Autoregressive Score (GAS) model, 12
generalized dynamic factor model, 18
Gibbs sampler, 16
Granger causality, 6
Granger representation theorem, 9
Great Moderation, 13
growth-at-risk, 13

Hamiltonian Monte Carlo (HMC), 16
heteroskedasticity and autocorrelation consistent (HAC) covariance, 23
Hodrick–Prescott filter, 14

illusion of sparsity, 19
impulse response function (IRF), 6
impulse-propagation paradigm, 3
intervention analysis, 5

Kalman filter, 13
KPSS test, 8

lasso, 19
local projections, 7
local-to-unity asymptotics, 8
long short-term memory (LSTM) network, 21
Lucas critique, 6

machine learning, 20
macroeconometric model, 4
Markov chain Monte Carlo (MCMC), 16
Markov-switching model, 12
Mincer–Zarnowitz regression, 23
Minnesota prior, 7
Mixed data sampling (MIDAS), 19
model confidence set, 24
multivariate GARCH, 27

narrative identification, 7
National Bureau of Economic Research (NBER), 3

nowcasting, 17

panel cointegration, 10
panel unit-root test, 9
partial autocorrelation function (PACF), 5
particle filter, 15
Phillips–Perron test, 8
portmanteau test, 5
probability approach, 4
probability integral transform (PIT), 24
proxy-SVAR, 31

quantile regression, 13

random forest, 20
realized volatility, 12

seasonal adjustment, 13
shrinkage estimation, 7
sign restrictions, 6
simulation smoothing, 16
simultaneous-equations model, 4
smooth-transition autoregressive model, 12
spectral representation theorem, 3
spurious regression, 7
state-space model, 13
stationary bootstrap, 23
stochastic volatility (SV), 12
structural change, 8
structural time series model, 14
structural VAR (SVAR), 6
synthetic control method, 22

temporal convolutional network, 21
Temporal Fusion Transformer, 21
text as data, 31
threshold autoregressive (TAR) model, 12
threshold cointegration, 11
time-varying parameter (TVP) model, 14

unit root, 8
unobserved-components model, 14

Value-at-Risk, 13
variational inference, 16
vector autoregression (VAR), 6
vector error-correction model (VECM), 10

Wold decomposition, 3

X-11 procedure, 13
X-13ARIMA-SEATS, 13