Chapter 1
A Brief History of Time Series Econometrics

Peter Fuleky

Abstract This chapter surveys the development of time series econometrics from its statistical foundations to modern, data-rich practice. It traces key methodological turning points, including univariate ARIMA modeling, the rise of vector autoregressions and debates over identification, and the resolution of nonstationarity through unit roots, cointegration, and error-correction. It also reviews advances in volatility and nonlinear modeling, filtering and signal extraction via state-space methods, Bayesian approaches to estimation and model uncertainty, and the emergence of high-dimensional and mixed-frequency forecasting frameworks. The chapter then examines how machine learning and deep learning methods have been integrated into time series workflows, and provides a comparative assessment of major approaches along dimensions such as forecasting performance, structural interpretability, and uncertainty quantification. It concludes with a discussion of testing and forecast evaluation as organizing principles, an agenda for future research, and enduring lessons for remaining relevant at the intersection of econometrics, data science, and real-time decision making.

1.1 Introduction

Time series econometrics develops and applies quantitative methods for analyzing data observed over time. It has become indispensable in modern economics because many central objects of interest—output, inflation, employment, asset prices, exchange rates, and interest rates—are inherently dynamic and often display persistence, cycles, volatility clustering, and structural change. Distinct from cross-sectional work, time series analysis must confront temporal dependence and evolving distributions. If those

Peter Fuleky ✉
University of Hawaii, Honolulu, USA, e-mail: fuleky@hawaii.edu

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 evolved from early descriptive studies of business cycles and stochastic foundations to an integrated toolkit that spans likelihood-based modeling, reduced-form multivariate systems, long-run equilibrium analysis, nonlinear and volatility models, Bayesian and state-space methods, and, more recently, data-rich and machine-learning approaches. This evolution has been driven by a recurring interaction between practical needs—improved forecasting performance, risk measurement, and credible policy analysis—and theoretical innovation in probability foundations, identification, and asymptotic theory. A unifying theme, consistent with the volume's focus on remaining relevant, is that progress in time series econometrics has typically been motivated by concrete empirical needs or new data environments rather than by methodological novelty alone.

The chapter follows a broadly chronological narrative while highlighting major methodological turning points. The first group of sections traces the core arc of the discipline: statistical foundations and the emergence of stochastic interpretations of cyclical behavior (Section 1.2), the rise of formal econometric modeling under the Cowles Commission program (Section 1.3), the Box–Jenkins approach to univariate ARIMA modeling (Section 1.4), the multivariate turn associated with vector autoregressions and the ensuing debate over structure and identification (Section 1.5), and the resolution of nonstationarity through unit roots, cointegration, and error-correction models (Sections 1.6 and 1.7).

Subsequent sections cover extensions and enrichments of this core framework: volatility and nonlinear dynamics (Section 1.8), the applied tradition of filtering, seasonal adjustment, and signal extraction via state-space representations (Section 1.9), Bayesian methods and computation (Section 1.10), the data-rich era of factor models, shrinkage, mixed-frequency methods, and real-time nowcasting (Section 1.11), and the recent integration of machine learning into time series workflows (Section 1.12).

The chapter concludes with assessment and synthesis: developments in testing and forecast evaluation (Section 1.13), a comparative assessment of major approaches (Section 1.14), a forward-looking agenda for future research (Section 1.15), and a closing discussion of enduring lessons and the challenge of remaining relevant (Section 1.16).

1.2 Statistical Foundations (Pre-1940)

The quantitative study of economic time series began in the early twentieth century and established many of the conceptual building blocks of modern econometrics. A central early question concerned the nature of economic fluctuations: are observed cycles driven by deterministic periodic forces, or can they arise from the propagation of random shocks? Researchers approached these issues by analyzing time plots, proposing early dynamic models, and confronting the pitfalls of naive correlation analysis.

An important foundation was the recognition of serial dependence. Yule (1927) introduced autoregressive (AR) models as a parsimonious way to represent persistence, illustrating with a low-order fit to sunspot data that current values can be statistically related to their own past. Closely related was an influential 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. The broader lesson—that trending time series can suggest spurious relationships if treated with static tools—would recur repeatedly in later econometric debates.

At roughly the same time, the stochastic interpretation of cycles gained prominence. Slutsky (1937) 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; the modeling of business cycles in this tradition is taken up in Chapter 22 of this volume. Together, these contributions shifted attention away from purely deterministic cycle narratives toward models in which randomness and dynamics jointly shape observed behavior.

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 clarified why autoregressive and moving average terms are natural ingredients in time series models: under broad conditions, a stationary process can be represented (or approximated) as a weighted sum of current and past shocks, providing the theoretical backbone for later ARMA modeling.

In parallel, foundational contributions to the mathematical theory of stochastic processes were being developed. Kolmogorov's work on the axiomatic foundations of probability and on the theory of stationary processes provided a rigorous framework that would underpin much of subsequent time series analysis (Kolmogoroff, 1933). His results on optimal linear prediction for stationary processes, developed alongside those of Wiener, established the theoretical basis for filtering and forecasting that later found wide application in economics. Complementing the time-domain perspective, Cramér (1942) proved the spectral representation theorem, showing that any stationary process can be decomposed into contributions at different frequencies. This frequency-domain perspective provided a mathematically elegant way to discuss cyclical behavior and would later become central to business-cycle measurement and band-pass filtering, even though its systematic application to economics came somewhat 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—a research program that began well before 1940 and culminated in *Measuring Business Cycles* (Burns & Mitchell, 1946). Although less formal statistically, this empirical taxonomy provided a benchmark chronology that subsequent econometric models sought both to explain and to forecast. The Burns–Mitchell tradition also shaped an empirical disposition toward studying comovements across many series, foreshadowing the factor-analytic approach that would be formalized decades later.

By the late 1930s, the stage was set: economists had 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. What was still missing was a unified econometric framework tying probability-based inference, identification, and dynamic specification together—a synthesis that began in earnest in the 1940s.

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

The 1940s marked a reorientation of econometric methodology, often described as 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 view observed macroeconomic histories not as unique deterministic trajectories, but as realizations of underlying stochastic processes. In this framing, the disturbance term became a substantive component of the model—capturing shocks and measurement error—rather than a residual nuisance, and the credibility of inference depended on explicitly stated probabilistic assumptions.

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—i.e., which theoretical restrictions and exogeneity assumptions are needed to pin down structural parameters from time series data (Koopmans, 1950). The general analysis of identification in simultaneous systems is taken up in Chapter 5 of this volume. 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 then used the available estimation machinery to fit these systems to macroeconomic time series. At the same time, the Cowles experience highlighted a recurring issue for time series work: identification and inference are inseparable from the stochastic properties of the disturbances and the adequacy of the dynamic specification.

In the 1950s, the Cowles approach was operationalized 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; the large-scale macroeconometric models in this tradition are the subject of Chapter 23 in this volume. One related response to dynamics in this period was the distributed-lag model, which offered parsimonious representation of gradual adjustment over time (Koyck, 1954). These models were intended for both explanation and forecasting, yet their treatment of stochastic dynamics was often limited by later time series standards. Disturbances were frequently assumed to be serially independent (or treated as such for estimation), and when residual autocorrelation was detected it was common to apply corrective procedures rather than to revisit the dynamic specification. The best-known example is the transformation proposed by Cochrane and Orcutt (1949), alongside diagnostic tools such as the Durbin–Watson tests for serial correlation in regression residuals (Durbin & Watson, 1950, 1951). In hindsight, 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.

At the same time, a partly separate stream of work—rooted in statistics, signal processing, and electrical engineering—was advancing frequency-domain approaches to stochastic processes. 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 natural way to discuss business-cycle periodicities and noise reduction, but they spread into mainstream econometric practice more slowly than structural equation modeling. Nonetheless, they became part of the toolkit that later researchers would draw on; Granger and Hatanaka (1964), for example, provided an early systematic treatment of spectral methods for economic time series, bridging the engineering and economics literatures and illustrating how the frequency domain could illuminate questions about cyclical comovements and lead–lag relationships.

By the end of the 1950s, the demand for reliable forecasts was growing in policy institutions and business, while digital computing was making iterative estimation increasingly feasible. Together, these forces created fertile ground for a more data-oriented approach to time series modeling and forecasting—an approach that would be popularized in the following decade by Box and Jenkins (1970). The postwar era thus delivered the probabilistic foundations and the first dynamic econometric models, while also exposing the tension between theory-driven structural specification and empirically driven time series modeling that would shape later developments.

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

A landmark development in time series econometrics came with the publication of Box and Jenkins (1970) *Time Series Analysis: Forecasting and Control*. 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. The timing mattered: postwar data production had expanded, and digital computing made iterative estimation and diagnostic checking feasible in routine applications. 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. At a conceptual level, the approach can be viewed as a practical implementation of the Wold representation: if a covariance-stationary process admits a moving-average representation in innovations, then parsimonious autoregressive–moving-average specifications provide workable parametric approximations for prediction and inference (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). 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 systematic 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 was typically carried out by (conditional) least squares or maximum likelihood, after which residuals were scrutinized 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. A complementary selection principle was parsimony, later formalized in information-criterion language—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 not merely technical but methodological. The approach offered a coherent data-oriented alternative to the Cowles Commission tradition of theory-driven simultaneous equations. Rather than starting from a structural system and treating serial correlation as an inconvenience to be corrected after estimation, Box–Jenkins began with the time-series properties of the data and treated dynamic dependence as the object to be modeled. This shift mattered particularly for forecasting, where the criterion of success is predictive performance rather than structural interpretation. 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. In this sense, Box–Jenkins broadened the use of time series modeling—providing a procedure that could be applied without a full structural theory—while raising the standard for empirical credibility by making stochastic assumptions and diagnostics explicit.

The diffusion of these ideas into economics was reinforced by econometric texts that translated the statistical toolkit into an economic forecasting framework. For example, Granger and Newbold (1977) presented univariate ARIMA methods alongside forecasting theory and forecast evaluation, helping to normalize Box–Jenkins procedures for economic data. 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—policy shifts, strikes, regulatory changes—when the background evolution of the series follows an ARIMA process. These extensions mattered for economics because they allowed policy-relevant questions to be studied using time series data without abandoning the diagnostics that made ARIMA models attractive in the first place.

A later development that extended the Box–Jenkins legacy was the automation of ARIMA model selection. Rather than relying on manual inspection of ACF and PACF plots, automated algorithms use information criteria and unit-root pre-tests to select the order of differencing and the ARMA lag structure from the data. The algorithm of Hyndman and Khandakar (2008) formalized this process and made large-scale univariate forecasting practical, a feature that proved especially valuable for applications involving hundreds or thousands of series.

In hindsight, the Box–Jenkins approach also clarified 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 increasingly salient as researchers moved toward multivariate descriptions of macroeconomic dynamics. Finally, the Box–Jenkins practice of differencing to achieve stationarity highlighted—without fully resolving—the conceptual issue of whether trends in macroeconomic data are deterministic or stochastic. Nonetheless, by the late 1970s the Box–Jenkins toolkit had become part of the econometric baseline: it supplied a common language of ARMA dynamics, residual diagnostics, and forecasting benchmarks that shaped subsequent developments in multivariate modeling and structural interpretation.

Box–Jenkins methods also shaped applied practice in official statistics, where seasonal ARIMA modeling became a core ingredient of seasonal adjustment and trend-cycle decomposition workflows. These developments institutionalized the view that careful filtering and decomposition are often prerequisites for modeling and communication—a theme discussed in detail in Section 1.9.

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

By the late 1970s, dissatisfaction with the prevailing large-scale macroeconometric models had grown. Their heavy theoretical structure invited criticism of the identifying restrictions needed for estimation, and their practical forecasting record was mixed in an era marked by stagflation and policy regime change. At the same time, univariate ARIMA models offered rigorous 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.

Against this backdrop, Sims (1980) proposed vector autoregressions (VARs) as a deliberately modest statistical description of the joint dynamics of macroeconomic time series. Sims argued that many structural systems achieved identification through what he memorably called ‘incredible identifying assumptions’. 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 naturally delivers impulse response functions (IRFs) that trace the dynamic effect of a one-time innovation through the entire system, along with forecast error variance decompositions (FEVDs) that summarize the contribution of each innovation to forecast uncertainty at different horizons. Together, IRFs and FEVDs provided a common language for describing ‘how shocks propagate’ in macro data. VARs also offered a convenient operationalization of 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, even though the concept captures predictive precedence rather than structural causation—a distinction that Granger himself emphasized. In monetary debates, these tools helped re-examine hypotheses associated with the monetary history tradition, including the view that money and policy shocks have systematic effects on real activity (M. Friedman & Schwartz, 1963).

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 can be read as an attempt to restore structured 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. Short-run restrictions imposed through contemporaneous exclusion patterns, often motivated by institutional arguments about timing (e.g., that monetary policy does not respond to output within the same month), provided an alternative route to identification that preserves the VAR’s data-driven dynamics while adding just enough theory to isolate specific shocks.

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 (Uhlig, 2005). Externalinstrument ('proxy') approaches use high-frequency movements around policy announcements to isolate shocks (Gertler & Karadi, 2015), while narrative identification uses historical records to construct measures of exogenous policy shifts that can be employed as instruments in VAR analysis (Romer & Romer, 2004). 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, accurate quantification of uncertainty around impulse responses became an important practical concern. Because IRFs are nonlinear functions of estimated coefficients, delta-method approximations can perform poorly in small samples, and bootstrap methods—particularly the bias-corrected bootstrap developed by Kilian (1998)—became the standard tool for constructing confidence bands in applied VAR work. In Bayesian settings, posterior credible sets for IRFs are obtained directly from the MCMC output, providing a natural measure of estimation uncertainty that is now routinely reported.

An important alternative to VAR-based impulse response analysis emerged with Jordà (2005)'s method of local projections (LPs). Rather than estimating a full VAR system, LPs recover impulse responses by running direct regressions at each forecast horizon. This approach is robust to misspecification of the VAR lag structure and extends naturally to nonlinearities and state-dependent effects—for example, by interacting with business-cycle indicators to estimate how policy effects differ across recessions and expansions (Jordà, 2005). 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. In practice, many applied researchers now 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 the trajectory of VARs. 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). Methodologically, 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 questions of regime change emphasized by the Lucas critique (Lucas, 1976). 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 carefully tested simplifications and long-run restrictions (Hendry, 1995).

By the end of the 1980s, VARs had become central tools in empirical macroeconomics and policy institutions, valued for their forecasting performance—often competitive with or superior to large structural systems in direct comparisons (Fair &

Shiller, 1990)—and for the interpretive power of IRFs and related decompositions (for a comprehensive survey, see Stock & Watson, 2001). At the same time, the decade made clear that multivariate dynamics could not be separated from the nonstationary features of many macroeconomic series, which motivated the unit-root and cointegration developments discussed next.

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

A major methodological challenge that crystallized in the late 1970s and 1980s was 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). The common thread is that two independent series, each carrying a trend—whether deterministic or stochastic, as with a unit root—tend to appear statistically related, so it is the presence of unrelated trends, not the unit root specifically, that drives the spurious association. For the integrated case that is the focus of this section, subsequent theory clarified the mechanism: when integrated processes are regressed on each other, standard asymptotic approximations break down and conventional inference can be severely distorted (Phillips, 1986).

Econometrically, the presence of unit roots required rethinking standard tools. If two series are trending (have unit roots), a levels regression can be spurious unless the variables are tied together by a genuine long-run relation—the case later formalized as cointegration. More generally, if a series has a unit root, its variance grows without bound over time, and regressions involving integrated variables typically produce residuals that are themselves nonstationary (often $I(1)$) unless the regressors share a common stochastic trend. 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. More broadly, both estimation and inference can have nonstandard limiting behavior under integration, so classical hypothesis tests can over-reject dramatically.

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. If a series is integrated of order one, $I(1)$, then its first difference is (under regularity conditions) approximately stationary, $I(0)$, whereas 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 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). The deeper theoretical basis for the power problem was provided by Phillips (1987), who developed 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. This framework shows that conventional tests have limited power to distinguish between exact unit roots and highly persistent processes, a finding that informed the development of improved tests with better power under local-to-unity alternatives (Elliott, 1999). Nonrejection of a unit root is therefore not definitive evidence of a random walk, and complementary evidence—economic theory, long-term behavior, stationarity tests—should inform modeling decisions.

The empirical salience of these issues was reinforced by C. R. Nelson and Plosser (1982), who examined long U.S. macroeconomic time series and found 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 stochastic trend components and reduces the risk of spurious findings when variables are driven by unrelated stochastic trends. Reflecting how routine this precaution became, Sims, Stock, and Watson (1990) observed that applied work often defaulted to differencing even when long-run relations were plausible and many inferential targets did not require such transformations. 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. This trade-off—statistical tractability through differencing versus the desire to model long-run relations in levels—directly motivated the development of cointegration methods discussed in the next section.

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 is economically consequential. 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'.

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). By the end of the 1980s, then, empirical macroeconomics had largely internalized the lesson that trends, persistence, and breaks are central features of time series, and that credible long-run inference requires tools explicitly designed for nonstationary data.

The unit-root framework also prompted work on long memory and fractional integration, which offers a middle ground between stationary $I(0)$ dynamics and exact unit roots. In this approach, persistence is captured by fractional differencing, so that shocks decay slowly and autocorrelations exhibit hyperbolic rather than geometric decay. Granger and Joyeux (1980) and Hosking (1981) developed 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) provided influential semiparametric inference results that clarified when long-memory behavior can be distinguished from near-unit-root dynamics in finite samples. Long-memory ideas also found application in volatility modeling, particularly through FIGARCH specifications that combine fractional integration with the GARCH framework 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.

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 with a cross-sectionally augmented Dickey–Fuller (CIPS) approach that controls for common factors through cross-sectional averages, 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 is empirically important. These panel methods are developed in full in Chapter 2 of this volume; here they enter only through their time-series dimension.

These tools became especially important as researchers worked with large international macro panels, firm-level panels, and cross-country comparisons, many of which display strong common components that resemble the factor structures estimated in macro indicator datasets. The constructive insight that followed was that some nonstationary series may nonetheless move together in the long run, so that certain linear combinations are stationary—the idea formalized as cointegration (Engle & Granger, 1987).

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

Cointegration emerged in the 1980s as a powerful and constructive response to the challenges posed by nonstationary time series. Unit roots made classical levels regressions vulnerable to spurious inference, while differencing, though statistically convenient, often removed precisely the low-frequency information economists care about. Cointegration provides a principled middle ground: variables may be individually $I(1)$, but if some linear combination is $I(0)$, then the variables share a common stochastic trend and the combination represents a stable long-run restriction. In this case, long-run relationships can 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 notion operational: 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 and mean-reverting 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 that is often regarded as the first application of what became the LSE methodology in dynamic econometric modeling, and it was subsequently developed into a general modeling strategy in that tradition (Hendry, 1995). 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 the influential 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, 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. Together, these methods helped standardize long-run estimation in cointegrated settings where feedback and serial dependence are empirically important.

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 impact 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 $\Pi = \alpha\beta'$ and the rank of Π equals the number of linearly independent cointegrating vectors. 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 restrictions linking short and long interest rates, or assessing fiscal sustainability by relating revenues and expenditures over time. As system methods matured, researchers increasingly used multivariate VECMs that allow several common trends and multiple long-run restrictions within a single framework, for example open-economy systems combining exchange rates, domestic and foreign prices, and interest differentials, or 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. This interpretive layer helped integrate long-run theoretical restrictions with a flexible representation of short-run dynamics. A natural complement was the concept of common cycles: just as cointegration implies that variables share common stochastic trends, Vahid and Engle (1993) showed that cointegrated variables may also share common cyclical features, so that certain linear combinations eliminate not only the long-run trend but also the transitory dynamics.

An important extension to the cointegration framework came from applied work that did not require prior knowledge of the integration order of 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 such that the inference is valid regardless of the integration order of the variables (within specified ranges). This flexibility proved especially valuable in applied settings where integration status was ambiguous or where sample sizes made pretesting unreliable. The ARDL bounds approach became widely adopted in empirical studies across macroeconomics, finance, and applied microeconomics.

Cointegration methods were also extended to panel settings. Pedroni (1999, 2004) developed panel cointegration tests that allow researchers to test for long-run relationships across multiple cross-sectional units such as countries or regions. These tests accommodate heterogeneity across units and pooled estimation of cointegrating vectors, enabling international macroeconomic studies and cross-sectional panel analysis where common long-run relations are plausible. Panel cointegration became particularly important in growth empirics and in studying international macroeconomic linkages.

Beyond the linear framework, researchers recognized that error-correction mechanisms might depend on the magnitude of disequilibrium. Balke and Fomby (1997) and others developed threshold cointegration models in which adjustment speeds or cointegrating vectors vary with the size of the equilibrium error. Such nonlinear cointegration structures are economically motivated by transaction costs or band-of-inaction behavior: when disequilibrium is small, agents may tolerate deviations, but large deviations trigger strong correction. These extensions broadened the toolkit for capturing realistic long-run adjustment dynamics while maintaining the cointegration framework.

Finally, cointegration methods were adopted rapidly, but 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—its residual remains $I(1)$—even when the included variables are genuinely related, so such a regression is misspecified yet not spurious (Banerjee, Dolado, Galbraith, & Hendry, 1993; Martin, Hurn, & Harris, 2013). By the end of the 1990s, cointegration and error-correction modeling had become part of the standard toolkit for nonstationary data, providing a coherent way to combine short-run dynamics with long-run equilibrium restrictions.

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 naturally than general-purpose statistical systems, and they made visual inspection of the data and routine diagnostic testing a normal part of dynamic modeling in central banks and other institutions. Built around the classical toolkit of their era—ARIMA, VAR, and unit-root and cointegration analysis—these programs translated academic advances into accessible procedures.

1.8 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 developed models for time-varying conditional variance, while another brought nonlinear dynamics and regime dependence into routine empirical modeling.

A widely used starting point for modeling volatility is Engle (1982)'s autoregressive conditional heteroskedasticity (ARCH) framework. 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. It also made clear why constant-variance assumptions can be problematic when the object of interest is uncertainty itself. When conditional variance changes over time, treating innovations as homoskedastic can understate uncertainty in turbulent periods and overstate it in calm periods, even if the conditional mean is specified sensibly.

Bollerslev (1986) generalized ARCH to generalized ARCH (GARCH) 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 clarified 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). More generally, financial and macroeconomic time series frequently exhibit heavier tails than the Gaussian distribution implies. Allowing for Student-t or other heavy-tailed innovations in GARCH and related models improved density forecasts and risk measures, while robust estimation methods offered further protection against the influence of outliers and extreme observations (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, initially under constant conditional correlation (Bollerslev, 1990) and subsequently through dynamic conditional correlation specifications that let comovements evolve through time (Engle, 2002; Engle & Kroner, 1995). Although these models were most influential in finance, they also informed work in macroeconomics where time-varying uncertainty and comovement are empirically important.

Extending the GARCH framework to capture long-memory persistence, Baillie et al. (1996) introduced FIGARCH (Fractionally Integrated GARCH), which uses fractional integration to produce the hyperbolic decay in squared-return autocorrelations observed in practice. 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, Shephard, & Chib, 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 essential for capturing the changing volatility of the postwar U.S. economy. More broadly, the observation-driven versus parameter-driven distinction remains a useful organizing principle: GARCH is computationally convenient and dominates financial risk applications, whereas SV integrates naturally into Bayesian state-space systems and is now the default volatility specification 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. 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. High-frequency data and the broader apparatus of volatility and financial-risk modeling are developed in Chapter 3 of this volume.

Beyond the ARCH/GARCH family, a unifying framework for time-varying parameter models emerged from score-driven (or Generalized Autoregressive Score, GAS) dynamics. Creal, Koopman, and Lucas (2013) developed a general framework in which model parameters evolve through time based on the scaled score of the likelihood function. This approach encompasses GARCH, stochastic volatility, and many other time-varying specifications as special cases, while providing a unified set of inference tools and simulation-based methods. Score-driven models have become increasingly useful for capturing complex patterns of parameter evolution in a flexible yet parsimonious way.

In parallel, econometricians 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 symmetry and time-invariant propagation. Under linearity, the response to a shock does not depend on whether the economy is near a constraint, in a recession, or in a high-volatility period, and adjustment takes the same form throughout the sample. 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 influential 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. More broadly, 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 idea 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 as complementary approaches, depending on whether the data suggested a one-time shift, a small number of recurring regimes, or state-dependent adjustment.

A further departure from classical conditional-mean modeling was the extension of quantile regression to time series settings. 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, this logic led to the conditional autoregressive Value-at-Risk (CAViaR) framework, which models tail quantiles as functions of past returns and past quantile estimates, offering a semiparametric alternative to GARCH-based Value-at-Risk (Jorion, 2007; Engle & Manganelli, 2004). More broadly, 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).

Taken together, volatility models and nonlinear dynamics broadened the scope of time series econometrics. ARCH/GARCH-type specifications made conditional second moments and uncertainty forecasts routine inputs to empirical work, particularly in finance, while asymmetric and multivariate extensions addressed empirically important features of risk. Threshold, smooth-transition, and regime-switching models provided structured departures from linear propagation and made it easier to discuss nonlinear adjustment and regime dependence in macroeconomic data.

These developments 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 became natural 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 it became more systematically connected to econometric modeling through state-space representations and the Kalman filter. The next section reviews this parallel tradition and its role in modern applied work.

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

By the 1960s, time series econometrics was not only developing stochastic models for forecasting and inference, but also shaping a large body of applied practice concerned with separating 'signal' from 'noise'. In macroeconomics and official statistics, this took the form of seasonal adjustment, trend–cycle decomposition, and related filtering procedures that aim to extract interpretable components from noisy and persistent data. Such methods matter because 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, the current standard seasonal adjustment tool used by statistical agencies worldwide. X-13ARIMA-SEATS combines the signal-extraction principles of both the X-11 tradition and the model-based SEATS (Seasonal and Trend decomposition using Time Series models) approach (Gómez & Maravall, 1996), allowing flexible decomposition of series into seasonal, trend, and irregular components through either automatic ARIMA selection or user-specified specifications. Related decomposition methods, including STL, provided flexible ways to represent seasonality and trend components (Cleveland, Cleveland, McRae, & Terpenning, 1990). These tools reinforced the idea that careful preprocessing and component extraction are often prerequisites for meaningful modeling and communication.

A closely related applied tradition is exponential smoothing, which can be understood as a family of adaptive filters designed for routine forecasting in settings 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, later work provided a coherent statistical foundation 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 highlighting 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 (Beveridge & Nelson, 1981). The Hodrick–Prescott filter (Hodrick & Prescott, 1997) became widely used as a pragmatic trend extractor. Hamilton (2018) provides a prominent critique along these lines, emphasizing concerns about 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. The broader lesson is that filtering choices are consequential and should be treated as part of the modeling strategy rather than as innocuous preprocessing.

An important question concerns the relationship between different decomposition approaches. The Beveridge–Nelson trend represents the permanent component as the long-run forecast of the level of the series, 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) showed that the apparent tension between these methods can be resolved by recognizing that when the correlation between trend and cycle shocks is unrestricted (allowed to vary freely), 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 clarified that model specification choices (particularly restrictions on shock correlations) are the key determinant of trends and cycles extracted from the data.

A useful perspective is that several widely used filters can be interpreted as special cases of unobserved-components state-space models. For example, the Hodrick–Prescott trend 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.

A natural complement to ad hoc filters is model-based signal extraction using state-space representations. Once a series is written as the sum of latent components with explicit stochastic laws of motion, filtering and smoothing deliver estimates of unobserved states together with measures of uncertainty. In this framework, decomposition is not imposed externally, but implied by an econometric model, and uncertainty about extracted components can be carried through to subsequent inference. Durbin and Koopman (2001) provide a systematic treatment of this model-based approach to filtering and smoothing.

State-space modeling and Kalman filtering provided the core computational apparatus for such model-based signal extraction. The Kalman filter provides a recursive algorithm for optimal updating in linear Gaussian state-space systems (Kalman, 1960). Its usefulness in economics is closely tied to how naturally many empirical problems can be expressed in a 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 written in this form, filtering and smoothing deliver both estimates of unobserved components and forecast distributions in a unified framework.

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.

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 the Bayesian estimation 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 has become widely influential in 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.

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.

Taken together, 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 would later be reused in mixed-frequency forecasting and nowcasting. The next section turns to Bayesian methods and computation, which provided a complementary set of tools for constraining high-dimensional parameterizations, integrating over latent states, and reporting uncertainty in complex dynamic models.

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

By the 1990s, applied time series econometrics had accumulated models whose appeal came with substantial estimation and inference burdens. VARs were flexible but parameter-heavy, and many nonlinear specifications were not amenable to standard likelihood calculations or conventional asymptotics. At the same time, improvements in computing and numerical methods made intensive optimization and simulation feasible in routine empirical work. These developments fostered a practical Bayesian reorientation in econometrics, one that mattered along two closely related margins. Shrinkage priors made large dynamic systems estimable by regularizing high-dimensional parameterizations, while simulation-based methods made latent-state and nonlinear models empirically tractable, providing coherent ways to integrate over unobserved states and to characterize uncertainty in complex dynamic settings. This section emphasizes Bayesian methods as they bear on time series; Chapter 9 of this volume sets out the general framework.

A widely used 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 is often described in terms of simple 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 matters for practice because 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. First, the prior can be tuned through a small set of hyperparameters that govern overall tightness and lag decay (Giannone, Lenza, & Primiceri, 2015). 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 (Carriero, Clark, & Marcellino, 2015). Second, the prior is transparent enough that it can 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, Giannone, and Reichlin (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 clarified a key point for practice: 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 objective is not parameter interpretation per se, but reliable real-time prediction and scenario analysis.

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 through influential statistical work (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 widely used 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 expanded the range of models amenable to Bayesian estimation. 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). These computational developments have been especially important for scaling Bayesian methods to larger and more realistic dynamic models.

Stochastic volatility models, introduced in Section 1.8, 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—reflecting 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, developed by Uhlig (2005), impose constraints on the sign (positive or negative) of impulse responses in the periods immediately following shocks, allowing researchers to identify structural shocks through economic theory without requiring fully parametrized structural models. This approach proved particularly useful for policy applications where the direction of transmission is known but the magnitude is uncertain. 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. These Bayesian identification strategies made it feasible to perform credible structural inference in large systems where classical parametric identification would be intractable, and they became standard tools for policy analysis and historical shock decomposition.

Bayesian computation also 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 influential 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 a coherent accounting of uncertainty. DSGE models themselves are the subject of Chapter 21 in this volume.

For models where the likelihood function is intractable but simulation is feasible, approximate Bayesian computation (ABC) offers an alternative approach. ABC methods, reviewed by 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 and large-scale DSGE specifications where likelihood evaluation is prohibitively expensive (Grazzini, Richiardi, & Tsionas, 2017). ABC trades computational burden for some loss of precision in posterior approximation but remains a valuable tool for disciplines where likelihood-free inference is the only feasible approach.

Bayesian ideas entered forecasting through explicit treatment of model uncertainty as well. Bayesian model averaging (BMA) replaces model selection by weighting models according to posterior probabilities, which can improve predictive performance when the true data-generating process is uncertain and when many specifications fit comparably well. A general discussion of BMA, including practical guidance and its relation to predictive performance, is provided by Hoeting, Madigan, Raftery, and Volinsky (1999). In time series settings, BMA also fits naturally with the idea that different models may be locally useful at different horizons or under different volatility regimes. A complementary perspective emphasizes evaluation of forecasts themselves rather than relying on model probabilities. Geweke and Amisano (2010) developed predictive Bayesian inference methods centered on log predictive scores, which measure the density forecast performance. Log predictive scores and related metrics have become standard tools for evaluating and comparing dynamic models, connecting to forecast-evaluation traditions in time series econometrics.

Bayesian weighting ideas also reinforced the broader case for forecast combination. Even when no single specification dominates, combining predictive distributions can stabilize forecasts and reduce the risk of large errors. Early forecast-combination arguments, developed outside a Bayesian framework, emphasized that averaging can outperform selection when models are misspecified or when relative performance shifts over time (Bates & Granger, 1969). BMA can be viewed as a probabilistic implementation of the same insight, with weights determined by posterior model probabilities rather than by ad hoc rules.

More broadly, Bayesian methods affected 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. At the same time, Bayesian analysis 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.

Bayesian methods also opened the door to nonparametric dynamic models in which the number of latent regimes or states is learned from the data rather than fixed in advance. Drawing on Dirichlet process priors and related constructions from machine learning, 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, demonstrating the practical value of letting the data determine the number and character of regimes.

By the early 2000s, Bayesian VARs and posterior simulation had become standard components of the time series econometric toolkit. They enabled empirical work with large parameterizations, latent states, time variation, and structural models whose likelihoods are difficult to handle analytically. They also supplied computational infrastructure that would later be reused in high-dimensional forecasting, mixed-frequency modeling, and machine-learning-inspired regularization, while keeping the econometric focus on identification, uncertainty quantification, and forecast evaluation.

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

Entering the 2000s, empirical macroeconomics and finance were increasingly shaped by larger and more heterogeneous information sets. 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 methodological response of the decade was to exploit information in large datasets while imposing enough structure to keep estimation feasible and forecasts reliable. This agenda overlaps with the Bayesian shrinkage logic discussed earlier, but here the focus is on extracting signal from many series, often observed at mixed frequencies and released at different times. The present section concentrates on the time-series response to large datasets; Chapter 4 of this volume treats the wider data environment.

Two complementary strategies became standard. The first is dimensionality reduction, especially factor models that replace a large panel of observables with a small number of latent components that summarize pervasive comovement. The second is shrinkage or variable selection, 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. A third theme, 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.

A widely used 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 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. While the idea of extracting common movements has deep roots in business-cycle measurement (Burns & Mitchell, 1946), it was developed into a practical forecasting device through 'diffusion index' methods, which use estimated factors as predictors for macroeconomic targets (Stock & Watson, 2002). The practical appeal is that one can replace a large, collinear predictor set with a small number of estimated factors that capture the dominant covariance structure of the data. 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 widely used 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. These contributions clarified when high-dimensional macro datasets are informative for forecasting. When comovement is strong and pervasive, a few factors can summarize much of the signal, while remaining idiosyncratic components capture measurement error, sector-specific shocks, and short-lived noise.

A second line of work emphasized the dynamic structure of the common component and its connection to spectral representations. The generalized dynamic factor model of Forni, Hallin, Lippi, and Reichlin (2000) treats the common component as driven by a small number of dynamic shocks and shows how it can be recovered using frequency-domain methods. A key distinction is between the static dimension of the factor space and the number of underlying dynamic shocks that generate the common component. This matters for practice because a panel may require several static factors to approximate comovement well, even if the underlying common dynamics are driven by only a few shocks. Later work refined estimation and forecasting procedures within this framework and reinforced the view that factor methods are coherent time-series models for high-dimensional systems (Forni, Hallin, Lippi, & Reichlin, 2005).

DFMs also connect naturally to 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, crucially, can handle ragged-edge datasets in which some entries of x_t are missing at the end of the sample. This state-space perspective became central for institutions that needed continuous real-time monitoring of the economy. 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. In practice, 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.

Factor methods were also integrated into multivariate time-series analysis aimed at policy and structural interpretation. The factor-augmented VAR (FAVAR) framework of Bernanke, Boivin, and Eliasz (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.

The real-time dimension of large datasets motivated formal nowcasting models. When key aggregates such as GDP are released with delay, while many indicators are observed monthly, weekly, or even daily, the information set at any point in time is incomplete and uneven. Some series are available promptly while others arrive with long lags, and the most recent observations are often missing for parts of the dataset. Giannone, Reichlin, and Small (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. Their framework also clarifies an operational feature that mattered for institutions. As each data release arrives, the nowcast can be updated mechanically through the Kalman filter. The update can be decomposed into a 'news' component that reflects the surprise in the new release relative to what the model predicted, and a weight that reflects 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, which is valuable in policy communication and real-time monitoring.

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 dictionary 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 resulting 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). More broadly, the rise of penalization reinforced a shift in practice toward tuning procedures that are explicitly linked to out-of-sample performance, such as cross-validation and rolling-window forecast evaluation.

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. From an applied perspective, 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. In practice, complete subset regression has proven competitive with regularization methods in macroeconomic forecasting while offering a more transparent framework that does not require tuning a penalty parameter. The method illustrates a broader principle that when model uncertainty is pervasive, combining diverse specifications can outperform the selection of any single model, a theme developed further in the evaluation literature discussed below.

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). 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. In practice, MIDAS provided a flexible middle ground between simple aggregation rules and full state-space mixed-frequency systems, particularly in cases where high-frequency predictors are plentiful and timeliness is critical.

Large datasets also reinvigorated panel time-series econometrics. The panel unit-root and cointegration tools discussed earlier (Section 1.6 and Section 1.7) became increasingly important as researchers worked with cross-country and cross-sectional panels containing dozens or hundreds of units. A distinctive concern in this setting 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. The strong common components in many large panels resemble the factor structures estimated in macro indicator datasets, further blurring the boundary between panel and factor-based approaches.

High dimensionality also 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 and grouped time series reconciliation. Many real-world datasets are organized in hierarchies—for example, sales forecasts at product, store, and regional levels must be coherent across levels. Hyndman, Ahmed, Athanasopoulos, and Shang (2011) developed methods for reconciling forecasts across hierarchical aggregation structures, ensuring that forecasts at different levels satisfy accounting constraints. This problem became increasingly important as forecasters worked with larger and more complex datasets, and efficient reconciliation procedures reduced the effective dimensionality of large forecasting problems.

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). They made it feasible to exploit richer information sets for forecasting and monitoring while clarifying the limits of naively scaling up traditional models. Computational advances were also important, as faster linear algebra and improved optimization routines made large-scale estimation feasible in routine applied work. They also reinforced a principle that would remain central in later developments, namely that the value of big data depends on suitable restrictions that control estimation noise and on evaluation protocols that prioritize out-of-sample performance. These advances set the stage for the subsequent integration of machine-learning-style algorithms, which would extend the same logic of high-dimensional prediction while further shifting attention toward computational scalability, model selection, and robust assessment of forecast accuracy.

1.12 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 influence time-series econometrics in a systematic way. On the practical side, 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. On the methodological side, the high-dimensional toolkit discussed in the previous section—factor models, shrinkage, and rigorous out-of-sample evaluation—made it natural to ask how far one could move beyond linear and Gaussian benchmarks without sacrificing stability (Varian, 2014; Mullainathan & Spiess, 2017). Our focus is the integration of these tools into time-series workflows; Chapter 12 of this volume treats the methods themselves.

The ML turn can be understood as a shift in 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. The econometric contribution was not to abandon time-series structure, but to embed it in procedures that respect temporal dependence and real-time forecasting constraints (Bergmeir, Hyndman, & Koo, 2018). A concrete illustration is Prophet, which decomposes a series into trend, seasonal, and holiday components using a modular additive model and applies automated parameter tuning (Taylor & Letham, 2018). By encoding time-series structure directly into the model specification, such tools made structured forecasting accessible to practitioners outside the traditional econometrics community and underscored the principle that flexible estimation is most productive when paired with domain-specific temporal structure.

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—penalizing model complexity to stabilize estimation—provided a natural bridge because 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 increasingly explored nonlinear supervised learning algorithms that can adapt to interactions, thresholds, and other forms of nonlinearity without requiring the analyst to pre-specify a parametric form. Tree-based methods—classification and regression trees, random forests, and boosting—proved especially attractive because they 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 can be understood as ensemble methods: they combine many weak learners to reduce variance (in forests) or sequentially correct errors (in boosting). Their appeal in time-series contexts often lies in forecasting with large sets of potential predictors, where nonlinear relationships or regime-specific effects might matter. 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.

This flexibility comes with familiar concerns. Without careful validation, ensemble methods can overfit, particularly when the analyst searches over many feature sets and tuning parameters or when the sample includes structural breaks that limit the relevance of older data. Even when a model forecasts well, the mapping from predictors to outcomes can be difficult to interpret economically. As a result, interpretability tools became part of the applied workflow. Partial dependence diagnostics and permutation-based importance measures provide summaries of average predictive effects, while Shapley-value decompositions offer local explanations of individual forecast movements (J. H. Friedman, 2001; Ribeiro, Singh, & Guestrin, 2016; Lundberg & Lee, 2017). In econometric practice, these diagnostics are best viewed as complements to structural analysis. They can be useful for monitoring and communication, but they do not, by themselves, establish causal mechanisms or policy-invariant relationships.

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 and the sequential nature of forecasting, and hyperparameter tuning must be nested within the evaluation scheme to avoid look-ahead bias (Bergmeir et al., 2018). This concern has long been present in econometrics under the headings of data mining and specification search, but ML made the issue operational 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, whose stabilizing benefits under model misspecification have been recognized since at least Bates and Granger (1969) (see also Timmermann, 2006; Elliott et al., 2013).

Deep learning represented a more substantial departure from traditional econometric modeling. While neural networks had been used in forecasting earlier, 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, Spiliotis, & Assimakopoulos, 2020). Yet M4 also highlighted a durable 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.

In parallel, the forecasting literature 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). For broader surveys of deep-learning forecasting architectures and design choices, see Benidis et al. (2022).

A second recurring 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, and one practical 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, & Assimakopoulos, 2022). A close econometric parallel is panel forecasting. Pooling across related series can raise accuracy, but evaluation should keep highly comoving units separated across training and test samples so that shared shocks do not mechanically inflate measured out-of-sample performance. The competitive evaluation tradition continued to evolve with the M6 competition, which shifted focus to financial return prediction and investment performance, reinforcing the value of structured comparisons as the set of candidate methods expanded (Makridakis et al., 2025).

A recent and significant development is the emergence of *foundation models* for time series—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, Challu, & Mergenthaler-Canseco, 2023) and Chronos (Ansari et al., 2024), which treat time series forecasting as a sequence modeling problem and leverage transformer architectures trained on large collections of publicly available time series. The promise of foundation models lies in their ability to capture general temporal patterns—trends, seasonality, level shifts, and nonlinear dynamics—from pre-training, potentially reducing the need for domain-specific feature engineering and model selection. However, their economic relevance remains an active area of investigation. Macroeconomic and financial time series have distinctive properties—structural breaks, policy feedback, nonstationarity, and relatively short histories—that differ substantially from the retail, energy, and web-traffic data that dominate many pre-training corpora. 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.

These developments underscore that, in many applications, the most consequential modeling choice is not the learning algorithm but the representation of the time-series problem. ML methods typically operate on supervised-learning inputs, so the analyst must decide which lags, seasonal components, and transformations enter the feature set. More broadly, combining econometric preprocessing with flexible learners is often essential for stability and for maintaining a clear link between modeling choices and the underlying data-generating features.

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 (or weaker conditions in the sequential setting) rather than parametric distributional assumptions (Vovk, Gammerman, & Shafer, 2005). Adaptations of conformal methods to time series, which must contend with temporal dependence and potential nonstationarity, have been developed and show promise for providing calibrated uncertainty bands around ML forecasts without relying on model-specific distributional assumptions (Zaffran, Feron, Goude, Josse, & Dieuleveut, 2022; Barber, Candès, Ramdas, & Tibshirani, 2023). This line of work helps bridge the gap between the flexibility of ML prediction and the uncertainty quantification that is central to econometric practice and policy communication.

Nowcasting and real-time monitoring provided another natural 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 can complement these frameworks by extracting predictive signals from large and heterogeneous feature sets, including alternative data such as internet search series and text. For example, internet search indices have been used to improve short-horizon monitoring of macro indicators, and text-as-data methods have been used to construct sentiment and uncertainty measures that enter forecasting exercises (Choi & Varian, 2012; Gentzkow, Kelly, & Taddy, 2019; Ardia, Bluteau, & Boudt, 2019).

Finally, the ML turn clarified the 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 aligns naturally with shrinkage, ensembles, and automated feature construction. It is less aligned with the goals of structural modeling, where identification, interpretation, and counterfactual analysis are central. This distinction became more explicit as researchers adapted ML methods to settings with formal inferential targets. Work on orthogonalization and debiasing provides one pathway for valid inference on low-dimensional targets in the presence of high-dimensional nuisance components (Chernozhukov et al., 2018). Related developments in causal ML—including tree-based estimators of heterogeneous treatment effects and synthetic control methods for policy evaluation—show how algorithmic tools can be combined with explicit inferential targets, although time-series dependence and feedback remain active complications in economic applications (Athey & Imbens, 2016; Wager & Athey, 2018; Abadie, Diamond, & Hainmueller, 2010).

In sum, the 2010s–2020s did not overturn the foundations of time-series econometrics. Rather, they expanded the admissible set of forecasting strategies and sharpened attention to rigorous out-of-sample validation. The result is a more pluralistic toolkit in which classical time-series structure and ML flexibility are combined, tested, and compared under transparent evaluation protocols. At the same time, the vastly larger model spaces made possible by tuning, feature construction, and alternative data heightened the risk of overfitting and spurious improvements, reinforcing the importance of credible evaluation. The next section takes up testing, forecast evaluation, and model comparison as organizing principles.

1.13 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. The discipline therefore evolved not only through new stochastic specifications, but also through a growing statistical arsenal for diagnostics, forecast evaluation, and principled model choice.

Early diagnostic testing was exemplified in Box–Jenkins analysis, where residual checks and portmanteau statistics (such as the Ljung–Box test) were used to assess whether an ARIMA specification had adequately captured serial dependence (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 suites of specification tests—discussed in Sections 1.6 and 1.7—that became part of routine diagnostic practice.

Forecast evaluation developed in parallel, motivated by the recognition that comparing forecast accuracy requires formal statistical inference rather than informal comparisons of mean squared error. A key milestone was the Diebold–Mariano test, which provides a general procedure for testing equal predictive accuracy under an explicit loss function (Diebold & Mariano, 1995). The test is based on the time series of loss differentials and uses long-run variance estimation to accommodate serial dependence induced by multi-step-ahead and overlapping forecasts. This made it possible to attach sampling uncertainty to claims of predictive improvement and helped standardize forecast comparisons in macroeconomics and finance. Subsequent work highlighted that the appropriate null and asymptotic distribution depend on the forecasting environment—for example, whether one compares fixed methods or recursively re-estimated models. West (1996) developed the asymptotic theory for out-of-sample forecast comparisons when model parameters are estimated, clarifying when estimation uncertainty affects the limiting distribution of predictive-accuracy statistics. Further refinements included tests of conditional predictive ability (Giacomini & White, 2006).

A foundational diagnostic in forecast evaluation 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 from these restrictions signal systematic bias or inefficiency—the forecast either consistently over- or under-predicts, or fails to exploit available information optimally. Although simple, rationality tests remain widely used as a first check on forecast quality and as a diagnostic for comparing the efficiency properties of competing forecasting models.

A recurring 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 covariance estimation, and Andrews (1991) developed influential results on bandwidth choice and long-run variance estimation under dependence. 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). A related concern is the use of real-time versus revised data in evaluation. Croushore (2011) demonstrated that using final revised figures rather than the data vintage actually available to the forecaster can produce misleading assessments of forecast quality, a point that is especially consequential for nowcasting and short-horizon prediction where the goal is to exploit the information set 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), clarified that forecast evaluation requires aligning the statistical framework with the specific model comparison at hand.

Beyond pairwise tests, time series econometrics developed concepts of forecast encompassing and forecast combination. Forecast encompassing asks whether one forecast subsumes the information in another. If forecast combinations offer no improvement, one model's forecast may be said to encompass the other. Formal tests of encompassing and related regression-based diagnostics became widely used (D. I. Harvey, Leybourne, & Newbold, 1998). At the same time, the empirical success of forecast combinations reinforced 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 (also known as the Bayesian information criterion) 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). The resulting contrast—AIC as asymptotically efficient for prediction in many settings, BIC as consistent for model identification when a true model exists in the candidate set—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 workhorses for 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 sufficiently rich dynamic model and simplifies through sequential testing, diagnostic checking, and encompassing considerations, with the aim of arriving at a parsimonious yet congruent representation (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—from the Chow test through the unknown-break-date procedures discussed in Section 1.8 on volatility and nonlinear dynamics—extended this logic to parameter stability, formalizing the principle that in dynamic environments it is rarely sufficient to estimate a model; one must also test whether its key invariances plausibly hold.

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, because the best-performing model in a large search may look impressive even when all models have equal population predictive ability. This critique was influential in finance and macro-forecasting and helped shift attention toward multiple-testing adjustments and joint inference. 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 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 clarified that 'accuracy' encompasses more than point prediction, including the calibration and sharpness of predictive uncertainty. Institutionally, this shift was reinforced by the Bank of England's adoption of 'fan charts' for inflation and GDP growth forecasts beginning in 1996, which made density forecasting a routine element of policy communication and encouraged other central banks to report forecast uncertainty systematically.

Finally, the practice of model assessment reinforced the importance of out-of-sample validation. 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. This emphasis on predictive validation, long ingrained in time series econometrics, also anticipated and influenced broader data-science practices such as cross-validation and benchmark competitions.

A similar eclectic approach 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). This comparison between methods has become a standard expectation in leading research, reflecting that different methods serve different purposes, and using multiple approaches often yields a more robust understanding.

In sum, the evolution of testing, forecast evaluation, and model comparison underscores a commitment to scientific discipline. New models must demonstrate value either through well-posed hypothesis tests, through predictive improvements under transparent loss functions, or through robustness across plausible alternatives. The resulting toolkit—ranging from residual diagnostics and unit-root tests to predictive-accuracy tests, multiple-comparison adjustments, and probabilistic scoring rules—has not merely accompanied methodological innovation; it has shaped which innovations endure by rewarding those that survive rigorous empirical scrutiny.

The next section compares the major classes of methods along common dimensions—forecasting ability, structural insight, scalability, and practical usability.

1.14 Comparative Assessment of Time Series Methods

Throughout this chapter, we encountered a variety of models and approaches, each with its own philosophy, mathematical structure, and range of applicability. It is useful to step back and compare these tools along key dimensions—such as their ability to forecast, to provide structural insight, to handle certain data features, and their practicality in implementation. This section provides a comparative evaluation of the major categories of time series econometric methods discussed, highlighting their strengths, limitations, and typical use cases. We also consider how they can complement each other. For clarity, we organize the comparison as a nested bulleted list and then elaborate on specific points.

- **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 or support causal analysis.
    - Assumes linearity and fixed parameters; struggles with structural breaks and regime changes.
    - Purely data-driven and provides no structural interpretation.

- **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.
    - Can handle regime changes in policy by design, addressing concerns raised by the Lucas critique (Lucas, 1976).
  - *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).
    - Often competitive in short-run forecasting exercises, especially relative to large structural systems (Fair & Shiller, 1990).
    - 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.
    - ARCH/GARCH is used most heavily in finance, though stochastic volatility has become central in macro through 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).
    · Less purely reduced-form in spirit, because priors encode external information.
    · Classical p-values are replaced by Bayesian credible intervals, which some practitioners may find less familiar.

- **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; J. Bai & Ng, 2002).
  – *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 remain interpretable as in standard regression.
  - 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.

- **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.

The list above organizes approaches by model class. The points below highlight cross-cutting dimensions along which these classes tend to differ 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).
- **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 in the short samples typical of macroeconomics unless combined with penalization or sparsity constraints (Tibshirani, 1996; Banbura et al., 2010).
- **Uncertainty and density forecasts.** Quantifying forecast uncertainty is increasingly seen as no less important than producing point predictions. State-space models, BVARs, and stochastic-volatility specifications (both GARCH-type and latent SV models) naturally produce full predictive distributions, which are essential for 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 systems (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, which matters for accountability (Smets & Wouters, 2007). As more complex methods enter applied practice, maintaining a clear link between modeling choices and their rationale becomes increasingly important—a theme discussed further in Sections 1.12 and 1.15.

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 toolkit and matching the right tool—or combination of tools—to the task. With this comparative landscape in mind, the next section considers the challenges and opportunities that are likely to reshape practice going forward.

1.15 Emerging Trends and Future Challenges

Having reviewed the methodological arc of time series econometrics, it is natural to ask what challenges and opportunities lie ahead. A central theme is the discipline’s continuous adaptation: time series econometrics has stayed relevant by responding to new data environments, new decision problems, and the empirical shortcomings of earlier models. The issues discussed below concern developments most likely to reshape practice over the coming decade.

Perhaps the most consequential trend is the continued proliferation of data sources. Econometric analysis increasingly extends beyond traditional aggregates to encompass high-frequency transaction data, text from news and social media, satellite imagery, mobility indicators, and other sources available in near real time. Incorporating these heterogeneous inputs into time series models in a principled way will require further development of data-rich approaches, including dynamic factor and mixed-frequency methods, as well as 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 increased granularity—hundreds of micro indicators rather than a single GDP series—which 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; Im et al., 2003).

A closely related challenge is real-time adaptability. The COVID-19 recession demonstrated vividly 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. In response, researchers are developing methods for rapid structural-break detection and monitoring, as well as time-varying parameter models that place greater weight on recent observations (Andrews, 1993; J. Bai & Perron, 1998; Cogley & Sargent, 2005; Primiceri, 2005). Beyond sudden shocks, slow-moving structural trends—productivity slowdowns, demographic shifts, 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 with substantial room for progress. The broader econometric emphasis on credible research designs—natural experiments, instrumental variables, difference-in-differences—has only partly penetrated aggregate time series analysis, in part because many macroeconomic questions rely on observational data with strong temporal dependence; these credible-design methods are developed for the cross-sectional setting in Chapter 6 of this volume. Promising directions include the use of high-frequency or external instruments for identifying structural shocks in VARs, as in the proxy-SVAR approach (Gertler & Karadi, 2015), 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 technically challenging but would be highly valuable for policy analysis.

Explainability and transparency will grow in importance as complex models become more common 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 articulated clearly (Taylor & Letham, 2018). Complementary tools such as Shapley-value decompositions and permutation-based importance measures provide post-hoc explanations of individual predictions (Lundberg & Lee, 2017; J. H. Friedman, 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.

The historical lesson is instructive: Box–Jenkins ARIMA modeling succeeded in part because it came with a clear, teachable diagnostic workflow (Box & Jenkins, 1970), and 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 have natural applications in these domains, while insights from domain sciences—epidemiological compartment models, 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 increasing use of granular and individual-level data introduces privacy and ethical considerations. Differential privacy and federated learning provide frameworks for combining information across institutions without centralizing sensitive microdata (Dwork & Roth, 2014; McMahan, Moore, Ramage, Hampson, & Agüera y Arcas, 2017) (see Chapter 15 in this volume), and these constraints will increasingly shape which data can be used and how models are estimated. These emerging challenges underscore the broader question of how the field can continue to adapt while preserving the core principles that have sustained it.

1.16 Enduring Lessons to Remain Relevant

The historical arc of time series econometrics—from early empirical studies and probability foundations to contemporary data-rich practice—reveals a field that has repeatedly adapted to new challenges. With each era, researchers developed tools to address limitations of prevailing approaches, whether by formalizing diagnostics for dependence and misspecification, resolving the puzzle of spurious regression and nonstationarity, modeling time-varying risk through conditional heteroskedasticity, or exploiting expanding information sets using shrinkage and factor structures. This dynamism reflects a dual commitment to scientific rigor and practical relevance, and the volume's theme of 'remaining relevant' resonates strongly with this record.

Reflecting on a century of developments, 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—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.** The field has 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), structural VARs with theory-motivated restrictions, and hybrid approaches that embed domain knowledge within predictive tools. 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 the field has increasingly formalized this recognition. 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 reinforces that pooling across diverse models can be more robust than selection when model uncertainty is pervasive (Bates & Granger, 1969; Timmermann, 2006).
- **Adapt continuously.** A meta-lesson from this history is the field's capacity for self-renewal. 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).

These lessons, distilled from successive waves of innovation, point to broader patterns. A defining feature of this history is that no single paradigm retains dominance indefinitely. The field has progressed through productive tension—between theory-driven and data-driven modeling, between structural interpretation and predictive performance, and between complexity and parsimony. VARs challenged large-scale structural models on empirical grounds, yet also stimulated syntheses that linked reduced-form dynamics to structural restrictions. Machine learning has expanded the forecasting toolkit, but its successful integration depends on rigorous validation, careful treatment of dependence, and clarity about inferential targets. The historical record suggests that progress is most durable when new methods are absorbed into, rather than positioned against, established econometric principles.

Two dimensions of this interplay deserve particular emphasis. First, economic theory and time series analysis are complementary. Cointegration illustrates how long-run economic reasoning can be embedded within empirically grounded short-run dynamics, while reduced-form evidence has repeatedly tempered theoretical constructions by revealing misspecification and instability. Maintaining this dialogue remains essential for policy analysis, where counterfactual questions require structural interpretation while real-time forecasting demands models that perform well under evolving conditions. The model best suited for forecasting is not always best suited for structural explanation, and practitioners have learned to navigate this trade-off by using multiple models for distinct purposes and cross-checking conclusions.

Second, structural change remains a persistent and underappreciated complication. No model can anticipate all discontinuities, and the practical implication is to combine modeling with robust monitoring, scenario analysis, and explicit quantification of uncertainty rather than to treat any fitted specification as definitive. Forecast evaluation and comparison tools—from predictive-accuracy tests through model confidence sets and proper scoring rules—help determine model choice and guard against overinterpretation of incremental gains, while maintaining diverse approaches reduces the risk of collective blind spots.

Looking ahead, the challenge of remaining relevant is ongoing. Time series econometrics will continue to be shaped by expanding data availability, computational scale, and demands for credible causal interpretation. The field's comparative advantage lies in combining a deep understanding of temporal dependence, equilibrium concepts, and economic context with rigorous evaluation practices and honest quantification of uncertainty. If it continues to absorb new tools while retaining this core discipline, time series econometrics is well positioned to remain a vital and evolving methodology at the intersection of statistics and economics.

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.

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, N., & 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., & 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

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

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

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. *Journal of Computational and Graphical Statistics*, *27*(2), 392–401. doi: 10.1080/10618600.2017.1316722

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

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.
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.
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.
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., & 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

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.

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

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.

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., 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.

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: 10.1198/016214505000000511

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., 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., 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.
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.
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.

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.

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

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.

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)

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.2015.02.003

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

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

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.

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
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
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., & Assimakopoulos, V. (2022). The m5 competition: Conclusions. *International Journal of Forecasting*, *38*(4), 1313–1337.
Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The m4 competition: Results, findings, and conclusions. *International Journal of Forecasting*, *36*(1), 54–74.
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
Martin, V. L., Hurn, A. S., & Harris, D. (2013). *Econometric modelling with time series: Specification, estimation and testing*. New York: Cambridge University Press.
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.
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
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
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

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., & 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*.

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.

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

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.

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

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)

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. (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., & 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.

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.

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.

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).

Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world*. New York: Springer. doi: 10.1007/0-387-25061-1

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

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