Abstract

Essays in Nonparametric Econometrics

Timothy Christensen

2014

This dissertation consists of four chapters which study nonparametric identification and estimation of models which have been the subject of much recent research in economic and econometric theory. The first three chapters study nonparametric identification and nonparametric sieve estimation of positive eigenfunction problems in economics, with particular application to dynamic asset pricing models. The fourth chapter, which is jointly authored with Xiaohong Chen, studies optimal uniform convergence rates for nonparametric instrumental variables models.

The first chapter introduces new econometric techniques for studying the long-run implications of dynamic asset pricing models. The long-run implications of a model are jointly determined by the functional form of the stochastic discount factor (SDF) and the dynamic behavior of the variables in the model. The estimators introduced in this chapter treat the dynamics as an unknown nuisance parameter. Nonparametric sieve estimators of the positive eigenfunction and eigenvalue used to decompose the SDF into its permanent and transitory components are proposed, together with estimators of the long-term yield and entropy of the permanent component of the SDF. The estimators are particularly simple to implement, and may be used to numerically compute the long-run implications of fully specified models for which analytical solutions are unavailable. Nonparametric identification conditions are presented. Consistency and convergence rates of the estimators are established. An approach for conducting asymptotic inference on the eigenvalue, long-term yield, and entropy of the permanent component of the SDF is provided. The semiparametric efficiency bounds for these parameters are derived and their estimators are shown to be efficient. The long-run implications of the consumption CAPM are investigated using these

methods. This investigation reveals a long-run version of the equity premium puzzle which is robust to certain augmentations of the representative agent's utility function. The estimators, identification conditions, and large sample theory presented in this chapter have broader application in economics including, for example, the nonparametric estimation of marginal utilities of consumption in representative agent models.

In the second chapter, the nonparametric identification conditions presented in Chapter 1 are both weakened and extended to more general function spaces. High-level conditions for consistency and convergence rates for nonparametric sieve estimators of the positive eigenfunctions of a collection of nonselfadjoint operators and their adjoints are also presented, along with useful results on the convergence of random matrices.

The third chapter further investigates the conditions under which the positive eigenfunction of an operator is nonparametrically identified. Identification is achieved if the operator satisfies two mild positivity conditions and a power compactness condition. Both existence and identification are achieved under an additional non-degeneracy condition. The identification conditions are presented for the general case of positive operators on Banach lattices. The general identification conditions presented in this chapter are applied to obtain new identification conditions for the positive eigenfunctions which are used to extract the long-run implications of dynamic asset pricing models. The identification conditions have other applications in economics. For example, the conditions may be applied to obtain new primitive nonparametric identification conditions for marginal utilities in heterogeneous-agent and representative-agent consumption-based asset pricing models, and to facilitate future nonparametric estimation of these models.

The fourth chapter is joint work with Xiaohong Chen. This chapter establishes optimality properties of sieve estimators of nonparametric regression models with endogeneity. Nonparametric regression with an endogenous regressor is an important nonparametric instrumental variables (NPIV) regression in econometrics and a difficult ill-posed inverse problem with unknown operator in statistics. This chapter establishes new uniform (sup-norm) convergence rates for sieve NPIV estimators, which are a nonparametric version of two-stage least squares estimators. The literature on NPIV estimation has so far only studied mean-square convergence rates. This chapter derives the best possible (i.e. minimax optimal) uniform convergence rates for estimators of NPIV models, and provides conditions under which sieve NPIV estimators attain their best possible rates. As an indication of the sharpness of the convergence rates obtained, it is shown that spline or wavelet nonparametric series regression estimators attain their well-known best possible rates even with weakly dependent data and heavy-tailed error terms.

Essays in Nonparametric Econometrics

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Timothy Christensen

Dissertation Directors: Xiaohong Chen and Peter C.B. Phillips

May 2014

Contents

List of Figures

List of Tables

Chapter 1

Estimating the long-run implications of dynamic asset pricing models

Dynamic asset pricing models link the prices of future state-contingent payoffs with sources of risk, the payoff horizon, and the preferences of economic agents. The stochastic discount factor (SDF, or pricing kernel) within a dynamic asset pricing model assigns values to future state-contingent payoffs. Recent work in macroeconomics and asset pricing has shown how to extract information about the long-run pricing implications of a model by analyzing the permanent component of the SDF (see, for example, Alvarez and Jermann (2005); Hansen and Scheinkman (2009); Hansen (2012); Backus, Chernov, and Zin (2012)). As this work has highlighted, long-run implications provide a powerful and robust means with which to analyze dynamic asset pricing models. Different assumptions about the preferences of economic agents may, in some cases, result in different short-run implications but the same long-run implications. Consequently, the long-run implications of classes of asset pricing models may, in some cases, be inferred by studying just one model.

This chapter introduces an econometric framework for extracting information about the permanent component of the SDF, and the pricing of long-horizon assets, from a dynamic asset pricing model. The permanent component of the SDF and the long-run implications of the model are jointly determined by both the functional form of the SDF and the short-run dynamics, or law of motion, of the variables in the model. The framework introduced in this chapter treats the dynamics as an unknown nuisance parameter. Economic theory is often vague regarding the precise form that the dynamics should take. In practice, dynamics are usually specified parametrically in a way that makes analytical solution of the model feasible. Changing the dynamics can change the long-run implications of a model. One might, therefore, be concerned that the long-run implications of a model may be sensitive to the specification of the dynamics. Generalized method of moments (GMM) is a popular technique for estimating asset pricing models because it uses moment restrictions, typically based on an Euler equation or asset-pricing equation, which are derived from economic theory and places only weak assumptions on the dynamics of the data. The estimators proposed in this chapter are based on the same Euler equation or asset-pricing equation one would use to estimate a model with GMM. Rather than placing parametric restrictions on the dynamics, the estimators proposed in this chapter nonparametrically infer, from a time series of data, attributes of the dynamics from which the long-run implications of the model are obtained.

As shown by Hansen and Scheinkman (2009) and Hansen (2012), information about the permanent component of the SDF and the pricing of long-horizon assets can be extracted by studying a positive eigenfunction problem related to an appropriately chosen operator. Their analysis applies to economies in which there exists a Markov state process whose value at each point in time contains all the relevant information for valuation. The operator is determined jointly by the SDF and the dynamics of the state process. The positive eigenfunction characterizes the state dependence of the prices of long-horizon assets, and its eigenvalue is related to the yield on long-term zero-coupon bonds and the entropy of the permanent component of the SDF. The entropy of the permanent component of the SDF is a joint measure of the dispersion of the SDF and persistence of the SDF process.

This metric can be used to place an upper bound on average excess returns on risky assets relative to long-term bonds (see Alvarez and Jermann (2005)). Whether or not the bound is satisfied by historical average returns on equity relative to long-term bonds provides a measure with which the predictions of the model may be evaluated.

The central focus of this chapter is the nonparametric estimation of the positive eigenfunction and its eigenvalue, the long-term yield, and the entropy of the permanent component. The operator is unknown when the dynamics are unknown. Extracting the long-run implications of the model given a time series of data on the state process therefore requires estimating a positive eigenfunction of an unknown operator. A feasible nonparametric sieve estimator is proposed, inspired by earlier work of Chen, Hansen, and Scheinkman (2000). Sieve estimation methods are appealing in this context as they reduce an intractable infinite-dimensional eigenfunction problem to a simple matrix eigenvector problem. The matrix eigenvector problem is formed by instrumenting the Euler equation or asset-pricing equation in the model by a growing collection of basis functions. The estimators are particularly easy to implement: no simulation, optimization or numerical integration is required. By contrast, the use of kernel-based methods in this context would involve nonparametric estimation of a conditional density, numerical computation of an integral, and solution of an infinite-dimensional eigenfunction problem. The sieve estimators may also be used to numerically compute the long-run implications of fully specified asset pricing models for which analytical solutions are unavailable.

Large sample properties of the estimators are established. The eigenfunction estimators are consistent and converge at reasonable nonparametric rates under appropriate regularity conditions. The asymptotic distribution and semiparametric efficiency bounds of the eigenvalue, long-term yield and entropy of the permanent component are derived, and the estimators of these quantities are shown to be efficient. An approach to performing asymptotic inference is provided. The derivation of the large sample properties is nonstandard, as the eigenfunction and eigenvalue being estimated are defined implicitly by an unknown nonselfadjoint operator. Favorable small-sample performance of the estimators is illustrated in a Monte Carlo study. The large sample theory is first presented for the case in which the long-run implications of a given SDF are to be investigated. The large sample theory is then extended to the case in which the researcher first estimates a SDF, either parametrically or semi/nonparametrically, from a time series of data on returns and the state process, then estimates the long-run implications of the estimated SDF. Other extensions of the large sample theory are explored, including nonparametric sieve estimation of marginal utilities in representative agent models.

To simplify the econometric analysis, the scope of this chapter is confined to discrete-time economies with finite-dimensional stationary state processes. Primitive nonparametric identification conditions for the positive eigenfunction in stationary discrete-time environments are provided. The conditions are formulated in terms of positivity and integrability conditions on the SDF and the stationary and transition densities of the state process. The existence and identification conditions complement those that Hansen and Scheinkman (2009) provide for general continuous-time environments. Existence of the positive eigenfunction is guaranteed under the identification conditions. A version of the long-run pricing result of Hansen and Scheinkman (2009) also obtains under the identification conditions.

The estimators are used to study the long-run implications of the consumption capital asset pricing model (CAPM). High levels of risk aversion are required to generate an estimated entropy of the permanent component of the SDF that is consistent with historical average returns on equities relative to long-term bonds. Moreover, when risk aversion is set high enough to rationalize this excess return the implied long-term yield is much larger than historical long-term yields. The long-run implications of the consumption CAPM are the same as a wider class of consumption-based representative agent models, including some habit formation models and limiting versions of recursive preference models. These empirical findings therefore have broader import beyond the standard consumption CAPM.

The estimators introduced in this chapter cannot, in their present form, be used to study models with latent state variables. Latent variables are a useful modeling tool for incorporating features such as stochastic growth and stochastic volatility in an analytically tractable way. For example, the popular long-run risks model of Bansal and Yaron (2004) specifies that (log) consumption growth is the sum of a latent predictable component and a stochastic component, in which both the latent predictable component and stochastic volatility evolve as first-order Gaussian processes. The estimators introduced in this chapter may be used to analyze models whose state processes exhibit time-variation in growth, conditional volatility, and other nonlinearities provided these features are state-dependent (instead of latent). Allowing for nonlinear, state-dependent dynamics goes some way to incorporate features that might otherwise be modeled by latent processes. So although the scope of the estimators is confined to models with observable variables, the restrictions this imposes on the dynamics the observable variables is less severe than it may first appear. Moreover, the version of the long-run pricing result presented in this chapter applies equally to models with latent state variables, and the sieve approach may be used to numerically calculate the long-run implications of fully specified models with latent state variables.

The remainder of the chapter is structured as follows. Section 1.1 discusses three other nonparametric eigenfunction problems in economics that the research developed in this chapter could be used to analyze. Section 1.2 reviews the decomposition of the SDF into its permanent and transitory components using the positive eigenfunction and its eigenvalue, and introduces other quantities to be studied. Identification and a version of the long-term pricing result are presented in Section 1.3. The nonparametric sieve estimators are introduced in Section 1.4, and their large sample properties are derived. Section 1.5 discusses extension of the large sample theory to cover estimated SDFs, more general SDFs, and nonparametric sieve estimation of marginal utilities. Section 1.6 examines the performance of the estimators in a Monte Carlo exercise. Section 1.7 studies the consumption CAPM using the estimators introduced in this chapter, and Section 1.8 concludes. Section 1.9 contains all proofs.

1.1 Nonparametric eigenfunction problems in economics

The identification conditions, estimators, and large sample theory developed in this chapter have broader application to nonparametric identification and estimation of economic models. Three other applications, namely nonparametric Euler equations, household consumption models, and transitory misspecification of asset pricing models, are now briefly outlined.

1.1.1 Nonparametric Euler equations

The Euler equation within consumption-based asset pricing models places restrictions on the comovement of asset returns and the marginal utility of consumption of economic agents. Such restrictions have been the basis for a vast literature on estimating consumption-based asset pricing models from time series of asset returns and consumption data. Recent work has shown that marginal utility of consumption may be represented as a positive eigenfunction of an appropriately chosen operator. This eigenfunction representation provides an alternative framework in which to study nonparametric identification and estimation of semi/nonparametric consumption-based asset pricing models.

Let MU_t^h denote the marginal utility of consumption of agent h at time t. Consider an economy in which the gross return on asset i from time t to $t+1$, denoted $R_{i,t+1}$, is determined by the Euler equation

$$MU_t^h = E[\beta MU_{t+1}^h R_{i,t+1} | \mathcal{I}_t^h] \tag{1.1}$$

where $\beta > 0$ is a time-preference parameter, and $\mathcal{I}_t^h$ is the information set of the agent at time t. Assume MU_t^h is a function (known to the agent/s but unknown to the econometrician) of a vector of explanatory variables X_t^h

$$MU_t^h = MU(X_t^h)$$

and that the explanatory variables belong to the agent's information set, i.e. $\sigma(X_t^h) \subseteq \mathcal{I}_t^h$. By iterated expectations, the Euler equation (1.1) can be rewritten as

$$E[MU(X_{t+1}^h)R_{i,t+1}|X_t^h] = \beta^{-1}MU(X_t^h)\,. \tag{1.2}$$

Expression (1.2) defines (MU, β) as the solution to nonparametric eigenfunction problem

$$\mathbb{T}_i MU = \beta^{-1} MU \tag{1.3}$$

where $\mathbb{T}_i f(X_t^h) = E[f(X_{t+1}^h)R_{i,t+1}|X_t^h]$. Marginal utility of consumption is typically assumed to be positive, in which case MU is a positive eigenfunction of $\mathbb{T}_i$.

Linton, Lewbel, and Srisuma (2011) and Escanciano and Hoderlein (2012) use this positive eigenfunction representation of marginal utility to analyze identification in representative agent models.[1] Chen, Chernozhukov, Lee, and Newey (2014a) use a similar eigenfunction representation to provide nonparametric identification conditions in a representative agent model with external habit formation. The eigenfunction representation of marginal utility of consumption does not appear to have been used to study heterogeneous-agent models to date.

The sieve estimators introduced in this chapter extend to the nonparametric estimation of the marginal utility function MU and time-preference parameter β of a representative agent, given a time series of data on $\{(X_t, R_{i,t+1})\}$ (see Section 1.5.3 for further details). This sieve-based approach is an alternative to the kernel-based procedure introduced in Linton, Lewbel, and Srisuma (2011). That the same pair (MU, β) are the solution to (1.3) for each asset i for which the Euler equation holds provides a source of over-identifying restrictions with which to test the model in both the representative- and heterogeneous-agent cases.

[1] This approach also has some similarities with Ross (2013), who nonparametrically recovers the pricing kernel from panels of option prices by solving a positive eigenvector problem.

1.1.2 Household consumption models

Eigenfunction techniques may also be used to study semiparametric Euler equations. Consider a semiparametric variant of the preceding model, in which MU_t^h is of the form

$$MU_t^h = [C_t^h]^{-\gamma} v(Z_t^h) \tag{1.4}$$

where C_t^h is the consumption of household h at time t, Z_t^h is a vector of explanatory variables, and v is a positive function. Attanasio and Weber (1993, 1995) use a model of this form to estimate preference parameters from household-level panel data. In their treatment, the function v is used to correct for the effect that a household's demographic structure may have on the marginal utility of a given level of consumption expenditure. A common approach for estimating these models from household-level panel data is to (i) assume a parametric form for v, (ii) log linearize the Euler equation, and (iii) estimate the log-linearized model by a panel instrumental variables regression. Marginal utility of the form (1.4) could equally be a heterogeneous-agent variant of the semiparametric consumption CAPM with external habit formation studied by Chen and Ludvigson (2009) and Chen, Chernozhukov, Lee, and Newey (2014a). The function v would represent an external habit formation component in this interpretation of (1.4). The identification conditions, estimators and large sample theory developed in this chapter may be extended to provide an alternative means with which to study nonparametric identification and estimation of these models.

The function v may be represented as the positive eigenfunction of an operator related to the Euler equation. Substituting MU_t^h of the form (1.4) into the Euler equation (1.1) yields

$$[C_t^h]^{-\gamma} v(Z_t^h) = E[\beta [C_{t+1}^h]^{-\gamma} v(Z_{t+1}^h) R_{i,t+1} | \mathcal{I}_t^h] \,. \tag{1.5}$$

Let $G_{t+1}^h = C_{t+1}^h / C_t^h$ denote the growth in consumption of household h from time t to time

$t+1$. When $\sigma((C_t^h, Z_t^h)) \subseteq \mathcal{I}_t^h$, the Euler equation (1.5) can be rewritten as

$$E[(G_{t+1}^h)^{-\gamma} R_{i,t+1} v(Z_{t+1}^h) | Z_t^h] = \beta^{-1} v(Z_t^h)\,.$$

Therefore, (v, β^{-1}) are the solution to the eigenfunction problem:

$$\mathbb{T}_{i,h} v = \beta^{-1} v \tag{1.6}$$

where $\mathbb{T}_{i,h} f(Z_t^h) = E[(G_{t+1}^h)^{-\gamma} R_{i,t+1} f(Z_{t+1}^h) | Z_t^h]$. The same (v, β^{-1}) must solve the eigenfunction relation (1.6) for each household h and asset i, providing a source of overidentifying restrictions.

1.1.3 Diagnosing transitory misspecifications in asset pricing models

The recent literature on extracting the long-run implications of asset pricing models has highlighted the fact that classes of asset pricing model may yield the same long-run implications but different short-run implications (see, e.g., Bansal and Lehmann (1997); Hansen (2012); Hansen and Scheinkman (2013); Backus, Chernov, and Zin (2012)). This line of research may also be used to study transitory misspecifications of SDFs in asset pricing models.

Let the economy be characterized by discrete-time Markov state process $\{X_t\}$ and consider SDF misspecification of the form

$$m(X_t, X_{t+1}) = \alpha m_{mis}(X_t, X_{t+1}) \frac{h(X_{t+1})}{h(X_t)} \tag{1.7}$$

where m is the true SDF and m_{mis} is the misspecified SDF used by the econometrician. The constant $\alpha > 0$ in expression (1.7) plays the role of a discount rate distortion, and the function $h > 0$ captures transitory misspecification of the SDF. Hansen (2012) shows that both m and m_{mis} will have the same permanent component, but different transitory components. If $\alpha = 1$ then m and m_{mis} will imply the same long-run rate of return. If

$h = 1$ then both m and m_{mis} will share the same positive eigenfunction.

Hansen and Scheinkman (2013) show that the true SDF m may be recovered from the misspecified SDF m_{mis} by solving a positive eigenfunction problem. Assume assets are priced using the true SDF m, i.e.

$$E\left[\, m(X_t, X_{t+1}) R_{i,t+1} |\, X_t\right] = 1\,. \tag{1.8}$$

Substitution of (1.7) into (1.8) yields

$$E\left[\, m_{mis}(X_t, X_{t+1}) R_{i,t+1} h(X_{t+1}) |\, X_t\right] = \alpha^{-1} h(X_t)\,.$$

The transitory adjustment h and multiplicative constant α are therefore the solution to the positive eigenfunction problem

$$\mathbb{T}_i h = \alpha^{-1} h \tag{1.9}$$

where $\mathbb{T}_i f(X_t) = E[m_{mis}(X_t, X_{t+1}) R_{i,t+1} f(X_{t+1}) | X_t]$. The techniques developed in this chapter may be applied to study nonparametric identification and estimation of (h, α) from a time-series of data on $(X_t, R_{i,t+1})$, thereby providing a means with which to diagnose transitory misspecifications of SDFs. Over-identifying restrictions are again implicit since the same (h, α) must solve the eigenfunction relation (1.9) for each asset i for which (1.8) holds.

1.2 Review of the long-run implications of dynamic asset pricing models

This section briefly reviews the positive eigenfunction problem and its relation to the long-run implications of asset pricing models, as exposited by Hansen and Scheinkman (2009) and Hansen (2012). The connection between these quantities and other metrics developed by Alvarez and Jermann (2005) and Backus, Chernov, and Zin (2012) is also discussed.

1.2.1 Model

Consider a class of economy characterized by a discrete-time (first-order) Markov state process $\{X_t\}$ defined on a complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, where time is indexed by $t \in \mathbb{Z}$ and where $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$ denotes the completion of the σ-algebra generated by $\{X_t, X_{t-1}, \ldots\}$. Let $\{X_t\}$ have support $\mathcal{X} \subseteq \mathbb{R}^d$. Assume further that in this economy the date-t price of a claim to the date-$(t+\tau)$ state-dependent payoff $Z_{t+\tau}$ is given by

$$E\left[\left(\prod_{s=t}^{t+\tau-1} m(X_s, X_{s+1})\right) Z_{t+\tau} \middle| X_t\right] \tag{1.10}$$

for some positive measurable function $m : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$, for all $t \in \mathbb{Z}$ and $\tau \geq 1$. The function m will be referred to generically as the SDF. The sequence

$$\{\ldots, m(X_{t-1}, X_t), m(X_t, X_{t+1}), m(X_{t+1}, X_{t+2}), \ldots\}$$

forms a stochastic process called the SDF process, which is denoted $\{m(X_t, X_{t+1})\}$.

Example: Consumption CAPM

Consider the consumption CAPM with complete, frictionless markets and a representative agent who maximizes, subject to a budget constraint, expected utility given by

$$\sum_{s \geq 0} \beta^s E[u(C_{t+s})|\mathcal{I}_t]$$

where $\mathcal{I}_t$ is the information set at time t, C_{t+s} is consumption of a representative good at date $t+s$, and β is a time preference parameter. The SDF is given by

$$m(X_t, X_{t+1}) = \beta \frac{u'(C_{t+1})}{u'(C_t)} .$$

When $u(c) = (1-\gamma)^{-1}(c^{1-\gamma} - 1)$ the SDF takes the familiar form

$$m(X_t, X_{t+1}) = \beta G_{t+1}^{-\gamma}$$

where $G_{t+1} = C_{t+1}/C_t$ is aggregate consumption growth and γ is the coefficient of relative risk aversion. Let $\{X_t\}$ be a strictly stationary and ergodic Markov state process and let $\mathcal{I}_t = \sigma(X_t)$. The consumption CAPM falls within the scope of the analysis of this chapter when $G_t = g(X_{t-1}, X_t)$ for some known function $g : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. For instance, one might take $X_t = (G_t, Y_t')'$.

Example: External habit formation

Following Abel (1990, 1999) and Galí (1994), consider an environment with complete, frictionless markets and a representative agent who maximizes, subject to a budget constraint, expected utility given by

$$\sum_{s \geq 0} \beta^s E[u(C_{t+s}, v_{t+s})|\mathcal{I}_t]$$

where v_{t+s} is a benchmark level of consumption which the agent takes as exogenous. When

$$u(c_t, v_t) = \frac{(c_t/v_t)^{1-\gamma} - 1}{1-\gamma}, \quad v_t = C_t^{\gamma_0} C_{t-1}^{\gamma_1}$$

the SDF is of the form

$$m(X_t, X_{t+1}) = \beta G_{t+1}^{\lambda} G_t^{\alpha}$$

where λ and α are functions of risk aversion γ and the consumption externality parameters γ_0 and γ_1. This model falls within the scope of the analysis in this chapter when $\{X_t\}$ is a strictly stationary and ergodic Markov state process, $\mathcal{I}_t = \sigma(X_t)$ and $G_t = g(X_t)$ for some known $g : \mathcal{X} \to \mathbb{R}$.

1.2.2 The principal eigenpair

As described in Hansen and Scheinkman (2009), the restriction of equation (1.10) to payoffs of the form $Z_{t+\tau} = \psi(X_{t+\tau})$ for suitable $\psi : \mathcal{X} \to \mathbb{R}$ defines a collection of linear operators $\{\mathbb{M}_\tau : \tau \geq 1\}$.[2] For each $\tau \geq 1$, the operator $\mathbb{M}_\tau$ is defined as

$$\mathbb{M}_\tau \psi(x) = E\left[\left(\prod_{s=t}^{t+\tau-1} m(X_s, X_{s+1}) \right) \psi(X_{t+\tau}) \middle| X_t = x \right] . \tag{1.11}$$

Given a payoff function ψ, the operator $\mathbb{M}_\tau$ assigns a date-t price to a claim to the date-$(t+\tau)$ state-dependent payoff $\psi(X_{t+\tau})$. For example, if $\iota(x) = 1$ for all $x \in \mathcal{X}$ then $\mathbb{M}_\tau \iota(X_t)$ is the date-t price of a τ-period zero-coupon bond.

As $\{X_t\}$ is a Markov process, the pricing operators factorize as $\mathbb{M}_\tau = \mathbb{M}_1^\tau$ for each $\tau \geq 1$. Let $\mathbb{M} := \mathbb{M}_1$ denote the 1-period pricing operator, i.e.

$$\mathbb{M}\psi(X_t) = E[m(X_t, X_{t+1})\psi(X_{t+1})|X_t] .$$

A function ϕ is an *eigenfunction* of the collection $\{\mathbb{M}_\tau : \tau \geq 1\}$ with *eigenvalue* ρ if

$$\mathbb{M}_\tau \phi = \rho^\tau \phi \tag{1.12}$$

for all $\tau \geq 1$. If, in addition, ϕ is positive then ϕ is referred to as the *principal eigenfunction*, ρ is the *principal eigenvalue*, and (ρ, ϕ) are the *principal eigenpair*.

Hansen and Scheinkman (2009) and Hansen (2012) show that principal eigenpairs may be used to decompose the SDF into its permanent and transitory components. That is,

$$m(X_t, X_{t+1}) = M^P_{t,t+1} M^T_{t,t+1}$$

[2] The set of "suitable" functions $\psi : \mathcal{X} \to \mathbb{R}$ will be defined subsequently.

where the permanent component of the SDF is

$$M_{t,t+1}^P = \rho^{-1}\frac{\phi(X_{t+1})}{\phi(X_t)}m(X_t, X_{t+1})$$

and the transitory component is

$$M_{t,t+1}^T = \rho\frac{\phi(X_t)}{\phi(X_{t+1})}$$

(cf. Equation (20) in Hansen (2012)).[3] The notion of permanent and transitory components employed here is different from that used in the study of nonstationary time series. For example, Beveridge and Nelson (1981) additively decompose a nonstationary time series into the sum of a random walk (martingale) permanent component and a stationary transitory component. In contrast, here the SDF process is multiplicatively decomposed into the product of $M_{t,t+1}^P$ and $M_{t,t+1}^T$ where the permanent component $M_{t,t+1}^P$ is a multiplicative martingale: $E[M_{t,t+1}^P|\mathcal{F}_t] = 1$ almost surely. By the definition of $M_{t,t+1}^P$, equation (1.11) may be rewritten as

$$\rho^{-\tau}\mathbb{M}_\tau\psi(x) = E\left[\left(\prod_{s=t}^{t+\tau-1} M_{s,s+1}^P\right)\frac{\psi(X_{t+\tau})}{\phi(X_{t+\tau})}\right| X_t = x\left.\vphantom{\prod_{s=t}^{t+\tau-1}}\right]\phi(x)$$

where $\rho^{-\tau}\mathbb{M}_\tau$ may be interpreted as an horizon-normalized price. Therefore, any differences between the horizon-normalized prices of claims to distinct future payoffs $\psi(X_{t+\tau})$ are due to differences between the covariation of the permanent component of the SDF and the scaled payoff $\psi(X_{t+\tau})/\phi(X_{t+\tau})$.

Under stochastic stability and integrability conditions, Hansen and Scheinkman (2009) and Hansen (2012) obtain a single-factor representation of the prices of long-horizon assets, namely

$$\lim_{\tau\to\infty}\rho^{-\tau}\mathbb{M}_\tau\psi(X_t) = \widetilde{E}[\psi(X)/\phi(X)]\phi(X_t) \tag{1.13}$$

[3]The definitions of the permanent and transitory components used here are the same as the definitions in Backus, Chernov, and Zin (2012), and correspond with what Alvarez and Jermann (2005) define as the growth in the permanent and transitory components of the pricing kernel process.

where $\widetilde{E}[\cdot]$ denotes expectation under a "twisted" probability measure associated with the permanent component. Equation (1.13) shows that when τ is large, the yield implied by the date-t price of a claim to $\psi(X_{t+\tau})$ is approximately $-\log\rho$, the *long-term yield*. Moreover, after discounting by ρ, state-dependence of the price is captured solely through $\phi(X_t)$. A restatement of (1.13) is provided in Theorem 1.3.2 below. This theorem shows how to calculate $\widetilde{E}[\cdot]$ in stationary discrete-time environments, and makes precise the sense in which the limit in (1.13) holds under the identification conditions presented in this chapter.

1.2.3 Entropies

The *entropy of the permanent component of the SDF* is defined as

$$L(M^P_{t,t+1}) = \log E[M^P_{t,t+1}] - E[\log M^P_{t,t+1}]\,.$$

Backus, Chernov, and Zin (2012) refer to $L(M^P_{t,t+1})$ as the "long-horizon entropy". Alvarez and Jermann (2005) show that

$$L(M^P_{t,t+1}) \geq E[\log R_{t+1}] - E[\log R_{\infty,t+1}] \tag{1.14}$$

where R_{t+1} is the gross return on a risky asset from time t to $t+1$ and $R_{\infty,t+1}$ is the gross return on a risk-free bond with infinite maturity from time t to $t+1$. For an asset pricing model to be consistent with observed returns on risky assets relative to long-term bonds, its permanent component must be large enough to satisfy the bound (1.14). In the stationary discrete-time environment considered in this chapter, the entropy of the permanent component takes the convenient form

$$L(M^P_{t,t+1}) = \log\rho - E[\log m(X_t, X_{t+1})] \tag{1.15}$$

whenever $E[\log \phi(X_t)]$ and $E[\log m(X_t, X_{t+1})]$ are finite. Given ρ and m, the premium on risky assets in excess of long-term bonds may be bounded by

$$\log \rho - E[\log m(X_t, X_{t+1})] \geq E[\log R_{t+1}] - E[\log R_{\infty,t+1}]$$

as a consequence of (1.14) and (1.15). By contrast, the entropy of the SDF is defined as

$$L(m(X_t, X_{t+1})) = \log E[m(X_t, X_{t+1})] - E[\log m(X_t, X_{t+1})] \tag{1.16}$$

and may be used to bound returns relative to short-term risk-free bonds:

$$L(m(X_t, X_{t+1})) \geq E[\log R_{t+1}] - E[\log R_{1,t+1}] \tag{1.17}$$

where $R_{1,t+1}$ is the gross return on a one-period risk-free bond from time t to $t+1$ (Cochrane, 1992; Bansal and Lehmann, 1997; Backus, Chernov, and Martin, 2011).

The entropy of the SDF measures the "roughness" or "dispersion" of the SDF, whereas the entropy of the permanent component measures both the roughness of the SDF and the persistence of the SDF process $\{m(X_t, X_{t+1})\}$. This latter point is reflected by the bound (1.15), which shows that the entropy of the permanent component may be used to bound the return on risky assets relative to short-term bonds minus the term premium $E[\log R_{\infty,t+1}] - E[\log R_{1,t+1}]$. The return on risky assets relative to short-term bonds depends on the dispersion of the SDF (cf. expression (1.17)) whereas the term premium depends on the dynamics of the SDF process. If the SDF is i.i.d. (independent and identically distributed) each period, then the entropy of the SDF and the entropy of the permanent component of the SDF are equal and the term premium is zero.

1.2.4 Robustness

An attractive reason for focusing on the long-run implications of an asset pricing model is that different models can have the same long-run implications but different short-run implications. This property was first noted by Bansal and Lehmann (1997), and is explored further by Hansen (2012), Hansen and Scheinkman (2013) and Backus, Chernov, and Zin (2012). This robustness property makes the long-run implications of a model a powerful means with which to analyze dynamic asset pricing models.

Let m and m^* be two SDFs that differ by the ratio of two transitory terms, i.e.

$$m^*(X_t, X_{t+1}) = m(X_t, X_{t+1}) \frac{f(X_{t+1})}{f(X_t)}$$

for some positive function f. For instance, m could be the SDF in the consumption CAPM and f might be an external habit formation term or a term that represents a limiting version of recursive preferences (Hansen, 2012; Hansen and Scheinkman, 2013). Although the short-run implications of m and m^* may differ, the permanent components of m and m^* will be the same (and so the entropy of the permanent components of m and m^* will be the same), and m and m^* will imply the same long-term yield. This robustness property means that the long-run implications of classes of asset pricing models can be inferred from the analysis of one model.

1.3 Nonparametric identification

Nonparametric identification of the positive eigenfunction ϕ is a consequence of the law of motion of the state variables, the form of the SDF, and the space of functions to which the eigenfunction is assumed to belong. Hansen and Scheinkman (2009) study identification of the positive eigenfunction in continuous-time economies. They use Markov process theory to derive sufficient conditions for identification of the positive eigenfunction. This section presents nonparametric identification conditions for the positive eigenfunction in stationary discrete-time economies. The conditions are also sufficient for existence of the positive eigenfunction. A function-analytic approach is used to establish identification and existence: existence follows by application of the Perron-Frobenius theorem for positive integral operators, and identification is established by a version of the Kreĭn-Rutman theorem. A version of the long-term pricing result of Hansen and Scheinkman (2009) holds under the identification conditions.

Function-analytic methods have been used recently to study identification of positive eigenfunctions related to other operators in economics. Chen, Chernozhukov, Lee, and Newey (2014a) study nonparametric identification of a habit formation component in a semiparametric consumption CAPM using these methods. In ongoing work, Linton, Lewbel, and Srisuma (2011) and Escanciano and Hoderlein (2012) use related techniques to analyze nonparametric identification of marginal utilities of consumption in representative agent models. However, in each of these studies the model and operator analyzed is different from the operator studied here.

1.3.1 Identification and existence

The conditions presented below are sufficient for nonparametric identification and existence of the positive eigenfunction ϕ. The conditions are stronger than required for identification, but are convenient for establishing both identification and the large sample properties of the estimators. Weaker nonparametric identification and existence conditions are presented in Section 2.2. Alternative identification conditions for stationary discrete- and continuous-time environments are explored in Chapter 3.

Assumption 1.3.1. $\{X_t\}$ *and* m *satisfy the following conditions:*

(i) $\{X_t\}$ *is a strictly stationary and ergodic (first-order) Markov process with support* $\mathcal{X} \subseteq \mathbb{R}^d$

(ii) the stationary distribution Q *of* $\{X_t\}$ *has density* q *(wrt Lebesgue measure) s.t.* $q(x) > 0$ *almost everywhere*

(iii) (X_0, X_1) *has joint density* f *(wrt Lebesgue measure) s.t.* $f(x_0, x_1) > 0$ *almost everywhere and* $f(x_0, x_1)/(q(x_0)q(x_1))$ *is uniformly bounded away from infinity*

(iv) $m : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *has* $m(x_0, x_1) > 0$ *almost everywhere and* $E[m(X_0, X_1)^2] < \infty$ *.*

Stationarity and ergodicity (Assumption 1.3.1(i)) is a stronger assumption than the irreducibility condition of Hansen and Scheinkman (2009). However, the requirement of stationarity is not necessarily restrictive. For example, consumption-based asset pricing models are typically written in terms of consumption growth to avoid potential nonstationarity in aggregate consumption (Hansen and Singleton, 1982; Gallant and Tauchen, 1989). Stationarity of the state process is also convenient for the derivation of the large sample properties of the estimators. Positivity of the joint density and boundedness of the ratio of joint to marginal densities (Assumptions 1.3.1(ii) and (iii)) is used both for identification and to develop the large sample theory. In particular, Assumption 1.3.1(iii) implies that $\{X_t\}$ is geometrically beta-mixing and geometrically rho-mixing.[4] Boundedness of the ratio of the joint to marginal densities in Assumption 1.3.1(iii) is violated if $\{X_t\}$ is constructed by stacking a higher-order Markov process into a first-order process as the joint distribution of (X_0, X_1) will be degenerate. Positivity of the SDF in Assumption 1.3.1(iv) is in line with the strict positivity of the SDF process assumed for identification in Hansen and Scheinkman (2009). Positivity of the SDF is satisfied in representative agent consumption-based asset pricing models for which

$$m(X_t, X_{t+1}) = \beta \frac{u'(C_{t+1})}{u'(C_t)}$$

provided the representative agent's marginal utility of consumption $u'(\cdot)$ is positive almost everywhere. Square integrability of the SDF is a standard assumption in asset pricing by no arbitrage (Hansen and Richard, 1987; Hansen and Renault, 2010).

Let $\mathscr{X}$ denote the Borel σ-algebra on $\mathcal{X}$ and let $L^2(Q) := L^2(\mathcal{X}, \mathscr{X}, Q)$ denote the space of all (equivalence classes of) measurable functions $\psi : \mathcal{X} \to \mathbb{R}$ for which $\|\psi\| := E[\psi(X)^2]^{1/2} < \infty$. The inner product $\langle \psi_1, \psi_2 \rangle := E[\psi_1(X)\psi_2(X)]$ makes $L^2(Q)$ a Hilbert space. Under Assumption 1.3.1, the pricing operator $\mathbb{M} : L^2(Q) \to L^2(Q)$ may be rewritten

[4] See section 2.4 for definitions of geometric beta- and rho-mixing.

as

$$\mathbb{M}\psi(x_0) = \int_{\mathcal{X}} m(x_0,x_1) \frac{f(x_0,x_1)}{q(x_0)q(x_1)} \psi(x_1)\,\mathrm{d}Q(x_1)\,. \tag{1.18}$$

Therefore $\mathbb{M}$ is an integral operator on $L^2(Q)$ of the form

$$\mathbb{M}\psi(x_0) = \int_{\mathcal{X}} \mathcal{K}(x_0,x_1)\psi(x_1)\,\mathrm{d}Q(x_1)$$

where the integral kernel $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is given by

$$\mathcal{K}(x_0,x_1) = m(x_0,x_1)\frac{f(x_0,x_1)}{q(x_0)q(x_1)}\,. \tag{1.19}$$

Assumption 1.3.1(ii)–(iv) implies that the kernel $\mathcal{K}$ is positive almost everywhere and

$$\int_{\mathcal{X}}\int_{\mathcal{X}} \mathcal{K}^2(x_0,x_1)\,\mathrm{d}Q(x_0)\,\mathrm{d}Q(x_1) < \infty\,.$$

Square-integrability of $\mathcal{K}$ implies $\mathbb{M}$ is Hilbert-Schmidt and therefore compact. The following identification and existence result is immediate by Theorems 2.2.1 and 2.2.2 in Section 2.2.

Theorem 1.3.1. *Under Assumption 1.3.1,*

(i) $\mathbb{M}$ *has a unique (to scale) eigenfunction* $\phi \in L^2(Q)$ *such that* $\phi > 0$ *(almost everywhere)*

(ii) ρ *is positive, has multiplicity one, and is the largest element of the spectrum of* $\mathbb{M}$.

There are a number of important implications of Theorem 1.3.1 beyond identification. First, that ρ has multiplicity one means that both ρ and ϕ are continuous with respect to small perturbations of $\mathbb{M}$. This continuity property is exploited in the derivation of the large sample properties of the estimators. Moreover, the fact that ρ is the largest eigenvalue of $\mathbb{M}$ is useful for estimation: if an estimator can be constructed that is close to $\mathbb{M}$ in an appropriate sense, then its maximum eigenvalue should be close to ρ.

1.3.2 Time reversal

Under Assumption 1.3.1 the time-reversed pricing operator $\mathbb{M}^* : L^2(Q) \to L^2(Q)$ is defined formally as the adjoint of $\mathbb{M}$ and is given by

$$\mathbb{M}^*\psi^*(x) = E[m(X_0, X_1)\psi^*(X_0)|X_1 = x] \,. \tag{1.20}$$

The reversed pricing operator might be interpreted as a pricing operator in the economy with time run backwards, but with the same SDF as if time were being run forwards.[5] Under Assumption 1.3.1, $\mathbb{M}^*$ has a unique (to scale) positive eigenfunction $\phi^* \in L^2(Q)$ such that

$$\mathbb{M}^*\phi^* = \rho\phi^* \tag{1.21}$$

(see Theorem 2.2.2 in Chapter 2). The adjoint eigenfunction ϕ^* is an important component of the asymptotic variance of the estimators, and also appears in the restatement of the long-term pricing result of Hansen and Scheinkman (2009) below.

Remark 1.3.1. *It is convenient to hereafter impose the normalizations* $E[\phi(X)^2] = 1$ *and* $E[\phi(X)\phi^*(X)] = 1$*, which define* ϕ *and* ϕ^* *uniquely.*

1.3.3 Asymptotic single-factor pricing

Hansen and Scheinkman (2009) and Hansen (2012) show that the positive eigenfunction ϕ captures state-dependence of the prices of long-horizon assets via the asymptotic single-factor pricing formula

$$\lim_{\tau\to\infty} \rho^{-\tau}\mathbb{M}_\tau\psi(X_t) = \widetilde{E}[\psi(X)/\phi(X)]\phi(X_t)$$

(see Section 7 of Hansen and Scheinkman (2009) and Section 6 of Hansen (2012) for precise statements of this result). Although Hansen and Scheinkman (2009) and Hansen (2012) define $\widetilde{E}[\cdot]$ as an expectation under a "twisted" probability measure associated with the

[5]See Rosenblatt (1971) for a discussion of time reversal for Markov processes.

permanent component, they do not show how to calculate the "twisted" probability measure.

The following theorem shows that an asymptotic pricing result holds for stationary discrete-time environments under Assumption 1.3.1. This theorem also shows how to calculate the twisted expectation $\widetilde{E}[\cdot]$ in stationary discrete-time environments.

Theorem 1.3.2. *Under Assumption 1.3.1, there exists a $c > 0$ such that*

$$\sup_{\psi \in L^2(Q): E[\psi(X)^2] \leq 1} \int_{\mathcal{X}} \left(\rho^{-\tau} \mathbb{M}^\tau \psi(x) - E[\psi(X)\phi^*(X)]\psi(x) \right)^2 \mathrm{d}Q(x) = O(e^{-c\tau})$$

as $\tau \to \infty$.

Theorem 1.3.2 shows that the scaled price $\rho^{-\tau}\mathbb{M}^\tau \psi(x)$ converges in mean square, uniformly over all payoff functions with unit norm, to $E[\psi(X)\phi^*(X)]\phi(x)$. Moreover, the approximation error vanishes exponentially quickly in the horizon τ. Let $\widetilde{Q}$ denote the twisted probability measure used to define the expectation $\widetilde{E}[\cdot]$. It follows by equating $E[\psi(X)\phi^*(X)]$ and $\widetilde{E}[\psi(X)/\phi(X)]$ that the Radon-Nikodym derivative of $\widetilde{Q}$ with respect to Q is

$$\frac{\mathrm{d}\widetilde{Q}(x)}{\mathrm{d}Q(x)} = \phi(x)\phi^*(x)$$

under Assumption 1.3.1. Theorem 1.3.2 is extended to other $L^p(Q)$ spaces in Chapter 2.2.

1.4 Estimation

This section introduces estimators of the positive eigenfunction, its eigenvalue, the long-term yield and the entropy of the permanent component of the SDF and presents the large sample properties of the estimators. It is assumed in this section that the SDF m is known and the researcher has available a time series $\{X_0, X_1, \ldots, X_n\}$ of data on the state process. Thus this section applies when the researcher is interested in investigating the long-run implications of a given SDF m. Extension to the case in which the SDF is first estimated from data is discussed in Section 1.5.1.

Sieve methods are used here to reduce the infinite-dimensional eigenfunction problem to a finite-dimensional matrix eigenvector problem. Implementation of the estimators is as simple as computing the eigenvectors and eigenvalues of two appropriately chosen matrices. The estimators introduced below may also be used to numerically compute the long-run implications of fully specified models for which analytical solutions are unavailable.

Chen, Hansen, and Scheinkman (2000) and Gobet, Hoffmann, and Reiss (2004) use sieve techniques to nonparametrically estimate an eigenfunction of the (selfadjoint) conditional expectation operator of a scalar diffusion process. In this chapter the operator $\mathbb{M}$ will typically be nonselfadjoint which introduces some additional technicalities.[6] For example, if $\mathbb{M}$ is selfadjoint then $\phi = \phi^*$. Moreover, if $\mathbb{M}$ is selfadjoint the pair (ρ, ϕ) are equivalently defined as the solution to an infinite-dimensional maximization problem. However, this equivalence does not hold in the nonselfadjoint case. The consistency and convergence rate calculations for the estimators of ρ and ϕ follow by simple modification of the arguments in Gobet, Hoffmann, and Reiss (2004). Estimation of the adjoint eigenfunction ϕ^*, derivation of the asymptotic distribution of the eigenvalue estimator and related estimators via a perturbation expansion, and the semiparametric efficiency bound calculations are all new.

1.4.1 Operator approximation

Let the sieve spaces $\{B_K : K \geq 1\} \subset L^2(Q)$ be a sequence of subspaces of $L^2(Q)$ of dimension K. For each K, let $b_{K1}, \ldots, b_{KK}$ denote the sieve basis functions that span B_K. Common examples of sieve basis functions include polynomial splines, wavelets, Fourier series and orthogonal polynomials (see Chen (2007) for an overview). Any function $\psi \in B_K$ may be written as

$$\psi(x) = b^K(x)' c_K(\psi)$$

[6] The operators in the other applications discussed in Section 1.1 will also typically be nonselfadjoint.

where $b^K(x) = (b_{K1}(x), \ldots, b_{KK}(x))'$ is a vector of basis functions and $c_K(\psi) \in \mathbb{R}^K$ is a vector of coefficients. Define the Gram matrix

$$\mathbf{G}_K = E[b^K(X)b^K(X)'] \,.$$

The relation $\psi \mapsto c_K(\psi)$ makes the space B_K isomorphic to $\mathbb{R}^K$ under the inner product induced by the Gram matrix because

$$E[\psi_1(X)\psi_2(X)] = c_K(\psi_1)'\mathbf{G}_K c_K(\psi_2)$$

for $\psi_1, \psi_2 \in B_K$.

The infinite-dimensional eigenfunction problem $\mathbb{M}\phi = \rho\phi$ in $L^2(Q)$ is approximated by a K-dimensional eigenfunction problem in B_K. Let $\Pi_K^b : L^2(Q) \to B_K$ denote the orthogonal projection onto B_K. Consider the eigenfunction problem

$$\Pi_K^b \mathbb{M}\phi_K = \rho_K \phi_K \tag{1.22}$$

where ρ_K is the largest eigenvalue of $\Pi_K^b\mathbb{M}$. Under the regularity conditions stated below, for all K sufficiently large the approximate eigenfunction ϕ_K will be unique (to scale) and ρ_K will be real-valued and positive. The approximate eigenfunction ϕ_K must belong to the space B_K. Consequently, ϕ_K can be written as

$$\phi_K(x) = b^K(x)'c_K$$

where $c_K = c(\phi_K)$ to simplify notation. Whenever $\mathbf{G}_K$ is invertible (this is guaranteed under the regularity conditions below) the approximate eigenvalue problem (1.22) may be rewritten as

$$\mathbf{G}_K^{-1}\mathbf{M}_K c_K = \rho_K c_K \tag{1.23}$$

where

$$\mathbf{M}_K = E[b^K(X_0)m(X_0, X_1)b^K(X_1)'] \tag{1.24}$$

and ρ_K is the largest eigenvalue of $\mathbf{G}_K^{-1}\mathbf{M}_K$.

Approximation of the adjoint positive eigenfunction ϕ^* is more subtle. When a solution to (1.22) exists with ρ_K real-valued, the adjoint of $\Pi_K^b\mathbb{M}$ has an eigenfunction ϕ_K^* with eigenvalue ρ_K. That is, there exists a ϕ_K^* such that

$$E[\phi_K^*(X)\Pi_K^b\mathbb{M}\psi(X)] = \rho_K E[\phi_K^*(X)\psi(X)]$$

for all $\psi \in L^2(Q)$. Let $\Pi_K^b\mathbb{M}|_{B_K}$ denote the restriction of $\Pi_K^b\mathbb{M}$ to the sieve space B_K. This restriction defines a linear operator $\Pi_K^b\mathbb{M}|_{B_K} : B_K \to B_K$. When a solution to (1.22) exists with ρ_K real-valued, the adjoint of $\Pi_K^b\mathbb{M}|_{B_K}$ has an eigenfunction $\phi_K^\star$ with eigenvalue ρ_K. That is,

$$E[\phi_K^\star(X)\Pi_K^b\mathbb{M}\psi_K(X)] = \rho_K E[\phi_K^\star(X)\psi_K(X)]$$

for all $\psi_K \in B_K$. The notation $\star$ in place of $*$ is used to denote that $\phi_K^\star$ is the eigenfunction of the adjoint of $\Pi_K^b\mathbb{M}|_{B_K}$ and that ϕ_K^* is the eigenfunction of the adjoint of $\Pi_K^b\mathbb{M}$. Although $\Pi_K^b\phi_K^* = \phi_K^\star$, it is not generally the case that $\phi_K^\star = \phi_K^*$. The approximate adjoint eigenfunction $\phi_K^\star$ belongs to the sieve space B_K. Therefore, $\phi_K^\star$ may be written as

$$\phi_K^\star(x) = b^K(x)'c_K^\star$$

where $c_K^\star = c_K(\phi_K^\star)$ to simplify notation. When ρ_K is real-valued and $\mathbf{G}_K$ is invertible, the vector $c_K^\star$ solves

$$\mathbf{G}_K^{-1}\mathbf{M}_K'c_K^\star = \rho_K c_K^\star\,. \tag{1.25}$$

where ρ_K is the largest eigenvalue of $\mathbf{G}_K^{-1}\mathbf{M}_K'$.

In summary, the infinite-dimensional eigenfunctions ϕ and ϕ^* are approximated by ϕ_K

and $\phi^\star_K$, where

$$\begin{aligned}\phi_K(x) &= b^K(x)' c_K \\ \phi^\star_K(x) &= b^K(x)' c^\star_K\end{aligned}$$

and c_K and $c^\star_K$ solve

$$\begin{aligned}\mathbf{G}_K^{-1}\mathbf{M}_K c_K &= \rho_K c_K \\ \mathbf{G}_K^{-1}\mathbf{M}_K' c^\star_K &= \rho_K c^\star_K\end{aligned}$$

where ρ_K is the largest eigenvalue of both $\mathbf{G}_K^{-1}\mathbf{M}_K$ and $\mathbf{G}_K^{-1}\mathbf{M}_K'$. Under the regularity conditions below, unique solutions to these eigenvector problems exist for all K sufficiently large. As, ϕ_K, ϕ^*_K and $\phi^\star_K$ are only defined up to scale, it is convenient to impose the sign normalizations $E[\phi_K(X)\phi(X)] \geq 0$, $E[\phi^*_K(X)\phi(X)^*] \geq 0$ and $E[\phi^\star_K(X)\phi(X)^*] \geq 0$ and the scale normalizations $E[\phi_K(X)^2] = 1$, $E[\phi_K(X)\phi^*_K(X)] = 1$ and $E[\phi_K(X)\phi^\star_K(X)] = 1$. These sign- and scale normalizations will be maintained hereafter, and define ϕ_K, ϕ^*_K and $\phi^\star_K$ uniquely.

1.4.2 Estimators

The matrices $\mathbf{G}_K$ and $\mathbf{M}_K$ can be estimated from data $\{X_0, X_1, \ldots, X_n\}$ by replacing the population expectations with their sample analogues, namely

$$\widehat{\mathbf{G}}_K = \frac{1}{n}\sum_{t=0}^{n-1} b^K(X_t) b^K(X_t)'$$

and

$$\widehat{\mathbf{M}}_K = \frac{1}{n}\sum_{t=0}^{n-1} b^K(X_t) m(X_t, X_{t+1}) b^K(X_{t+1})' \,.$$

The estimator $\widehat{\rho}$ of ρ is the largest eigenvalue of $\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{M}}_K$, i.e.

$$\widehat{\rho} = \lambda_{\max}(\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{M}}_K)\,.$$

When $\widehat{\rho}$ is real valued (which it is with probability approaching one under the regularity conditions below) let $\widehat{c}$ and $\widehat{c}^\star$ solve the matrix eigenvalue problems

$$\begin{aligned} \widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{M}}_K\widehat{c} &= \widehat{\rho}\,\widehat{c} && (1.26) \\ \widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{M}}'_K\widehat{c}^\star &= \widehat{\rho}\,\widehat{c}^\star\,. && (1.27) \end{aligned}$$

The estimators of ϕ and ϕ^* are

$$\begin{aligned} \widehat{\phi}(x) &= b^K(x)'\widehat{c} && (1.28) \\ \widehat{\phi}^\star(x) &= b^K(x)'\widehat{c}^\star\,. && (1.29) \end{aligned}$$

As $\widehat{\phi}$ and $\widehat{\phi}^\star$ are only defined up to sign and scale, impose the sign normalizations $E[\widehat{\phi}(X)\phi_K(X)] \geq 0$ and $E[\widehat{\phi}^\star(X)\phi_K^\star(X)] \geq 0$ and the scale normalizations $E[\widehat{\phi}(X)^2] = 1$ and $E[\widehat{\phi}(X)\widehat{\phi}^\star(X)] = 1$. The estimators $\widehat{\phi}$ and $\widehat{\phi}^\star$ are defined uniquely under these normalizations.

Recall that the long-term yield is $y = -\log\rho$ and the entropy of the permanent component of the SDF is $L = \log\rho - E[\log m(X_0, X_1)]$. In light of these definitions,

$$\widehat{y} = -\log\widehat{\rho} \qquad (1.30)$$

and

$$\widehat{L} = \log\widehat{\rho} - \frac{1}{n}\sum_{t=0}^{n-1}\log m(X_t, X_{t+1}) \qquad (1.31)$$

are natural estimators of y and L.

1.4.3 Regularity conditions and convergence rates

The following regularity conditions, in conjunction with Assumption 1.3.1, are sufficient to establish consistency and convergence rates of the estimators. Estimation of the positive eigenfunctions of a collection of operators under higher-level conditions is discussed in Section 2.3.

Let $\zeta_0(K) = \sup_x \|b^K(x)\|_2$ as in Newey (1997). For example, $\zeta_0(K) = O(\sqrt{K})$ for polynomial spline, Fourier series and wavelet bases and $\zeta_0(K) = O(K)$ for polynomial bases on appropriate domains (see, e.g., Newey (1997)). Let $\tilde{b}^K(x) = E[b^K(X)b^K(X)']^{-1}b^K(x)$ denote a vector of orthonormalized sieve basis functions. Define the orthonormalized estimators

$$
\begin{aligned}
\widehat{\widetilde{\mathbf{G}}}_K &= \frac{1}{n}\sum_{t=0}^{n-1}\tilde{b}^K(X_t)\tilde{b}^K(X_t)' \\
\widehat{\widetilde{\mathbf{M}}}_K &= \frac{1}{n}\sum_{t=0}^{n-1}\tilde{b}^K(X_t)m(X_t,X_{t+1})\tilde{b}^K(X_{t+1})'
\end{aligned}
$$

and their orthonormalized population counterparts

$$
\begin{aligned}
\widetilde{\mathbf{G}}_K &= E[\tilde{b}^K(X)\tilde{b}^K(X)'] \\
\widetilde{\mathbf{M}}_K &= E[\tilde{b}^K(X_0)m(X_0,X_1)\tilde{b}^K(X_1)']
\end{aligned}
$$

where $\widetilde{\mathbf{G}}_K = I_K$ (the $K \times K$ identity matrix) by virtue of orthonormalization. The orthonormalized estimators are infeasible in practice because Q is typically unknown; however it is convenient to define the regularity conditions in terms of these quantities.

Let $\|\cdot\|_2$ denote the matrix spectral norm when applied to matrices and the Euclidean norm when applied to vectors. That is, if $\mathbf{A}_K$ is a $K \times K$ matrix and $c = (c_1, \ldots, c_K)' \in \mathbb{R}_K$

then

$$\begin{aligned}\|\mathbf{A}_K\|_2 &= \sup\{\|\mathbf{A}_K c\|_2 : c \in \mathbb{R}^K, \|c\|_2 = 1\} \\ \|c\|_2 &= \left(\sum_{k=1}^{K} c_k^2\right)^{-1/2} .\end{aligned}$$

Recall that $\|\cdot\|$ denotes the $L^2(Q)$ norm when applied to functions in $L^2(Q)$. Let $\|\cdot\|$ also denote the operator norm when applied to linear operators on $L^2(Q)$. That is, if $A : L^2(Q) \to L^2(Q)$ is a linear operator then $\|A\| = \sup\{\|Af\| : f \in L^2(Q), \|f\| \leq 1\}$.

Define the K-vectors $\widetilde{c}_K$ and $\widetilde{c}_K^\star$ such that $\phi_K(x) = \widetilde{b}^K(x)'\widetilde{c}_K$ and $\phi_K^\star(x) = \widetilde{b}^K(x)'\widetilde{c}_K^\star$. Let $\{\bar{\eta}_{n,K}, \eta_{n,K} : n, K \geq 1\}$ be sequences of positive real numbers such that

$$\max\left\{\|\widehat{\widetilde{\mathbf{G}}}_K - I_K\|_2, \|\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K\|_2\right\} = O_p(\bar{\eta}_{n,K})$$

and

$$\begin{aligned}\max\Big\{&\|(\widehat{\widetilde{\mathbf{G}}}_K - \widetilde{\mathbf{G}}_K)\widetilde{c}_K\|_2, \|(\widehat{\widetilde{\mathbf{G}}}_K - \widetilde{\mathbf{G}}_K)\widetilde{c}_K^\star/\|\widetilde{c}_K^\star\|_2\|_2, \\ &\|(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)\widetilde{c}_K\|_2, \|(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)\widetilde{c}_K^\star/\|\widetilde{c}_K^\star\|_2\|_2\Big\} = O_p(\eta_{n,K}) .\end{aligned}$$

The inequality $\|\mathbf{A}_K c\|_2 \leq \|\|\mathbf{A}_K\|_2\|c\|_2$ holds by definition of $\|\cdot\|_2$ and implies $\eta_{n,K} = O(\bar{\eta}_{n,K})$. Different values of $\bar{\eta}_{n,K}$ and $\eta_{n,K}$ will be obtained depending on the number of moments of $m(X_t, X_{t+1})$.

Remark 1.4.1. *Section 2.4 provides further details as to how to calculate $\bar{\eta}_{n,K}$ and $\eta_{n,K}$. If Assumption 1.3.1 holds, then $\bar{\eta}_{n,K} = \zeta_0(K)^2/\sqrt{n}$ and $\eta_{n,K} = \zeta_0(K)^2/\sqrt{n}$. If, in addition, m is bounded and $\zeta_0(K)(\log n)/\sqrt{n} = o(1)$, then $\bar{\eta}_{n,K} = \zeta_0(K)(\log n)\sqrt{n}$ and $\eta_{n,K} = \zeta_0(K)/\sqrt{n}$.*

Assumption 1.4.1. *The following regularity conditions are satisfied:*

(i) $\|\Pi_K^b \mathbb{M} - \mathbb{M}\| = O(\delta_K)$ *where* $\delta_K = o(1)$ *as* $K \to \infty$

(ii) $\bar{\eta}_{n,K} = o(1)$ *as* $n, K \to \infty$

(iii) $\lambda_{\min}(\mathbf{G}_K) \geq \underline{\lambda} > 0$ *for each* $K \geq 1$

(iv) *there exists a sequence* $\{h_K^* : K \geq 1\}$ *with* $h_K^* \in B_K$ *such that* $\|\phi^* - h_K^*\| = O(\delta_K^*)$.

Assumption 1.4.1(i) requires that the range of $\mathbb{M}$ can be uniformly well approximated over the sieve space B_K, with the approximation error vanishing as the dimension of the sieve space increases. Assumption 1.4.1(i) also implies that $\|\Pi_K^b \phi - \phi\| = O(\delta_K)$. The weaker condition $\|\Pi_K^b \mathbb{M} - \mathbb{M}\| = o(1)$ and $\|\Pi_K^b \phi - \phi\| = O(\delta_K)$ suffices to calculate the following convergence rates for $\widehat{\rho}$ and $\widehat{\phi}$, however the stronger form presented in Assumption 1.4.1(i) is useful for derivation of the limit theory. Assumption 1.4.1(ii) is a condition on the maximum rate at which K can increase with n while maintaining consistency of the matrix estimators. Assumption 1.4.1(iii) is a standard condition for nonparametric estimation with a linear sieve space (see, e.g., Newey (1997); Chen and Pouzo (2012)). Assumption 1.4.1(iii) can be relaxed to allow $\lambda_{\min}(\mathbf{G}_K) \searrow 0$ as K increases, but this may slow the convergence rates. Assumption 1.4.1(iv) requires that ϕ^* can be approximated by a sequence of elements of the sieve space, with the approximation error vanishing as K increases. The condition on ϕ^* in Assumption 1.4.1(iv) can be dropped if $\mathbb{M}$ is selfadjoint (since $\phi = \phi^*$ in that case). When $\mathbb{M}$ is nonselfadjoint the separate treatment of ϕ and ϕ^* is required because Assumption 1.4.1(i) does not guarantee that $\mathbb{M}^*$, and therefore ϕ^*, can be approximated well over B_K. Assumptions 1.4.1(i) and (iv) can be motivated by imposing smoothness conditions on the kernel $\mathcal{K}$ and choosing an appropriate sieve, as in the example below.

The following theorem establishes consistency of $\widehat{\rho}$, and mean square convergence rates of $\widehat{\phi}$ and $\widehat{\phi}^\star$ as $n \to \infty$.

Theorem 1.4.1. *Under Assumptions 1.3.1 and 1.4.1, there is a set whose probability approaches one on which* $\widehat{\rho}$ *is real and positive and has multiplicity one, and*

(i) $|\widehat{\rho} - \rho| = O_p(\delta_K + \eta_{n,K})$

(ii) $\|\widehat{\phi} - \phi\| = O_p(\delta_K + \eta_{n,K})$

(iii) $\|\widehat{\phi}^\star/\|\widehat{\phi}^*\| - \phi^*/\|\phi^*\|\| = O_p(\delta_K^* + \eta_{n,K})$.

The rates of convergence in Theorem 1.4.1 exhibit the standard bias-variance tradeoff in nonparametric estimation. The "bias term" is $O(\delta_K)$ (or $O(\delta_K^*)$ for ϕ^*) which measures error in approximating ϕ and ϕ^* by their K-dimensional counterparts ϕ_K and $\phi_K^\star$. The "variance term" is $O_p(\eta_{n,K})$ which measures the difference between the estimators $\widehat{\phi}$ and $\widehat{\phi}^\star$ and their sample counterparts ϕ_K and $\phi_K^\star$.

Consistency and preliminary rates of convergence for $\widehat{y}$ and $\widehat{L}$ are established in the following Corollary. These estimators will be shown to be $\sqrt{n}$-consistent under stronger assumptions in Section 1.4.4.

Corollary 1.4.1. *Under the assumptions of Theorem 1.4.1,* $|\widehat{y} - y| = O_p(\delta_K + \eta_{n,K})$. *If, in addition,* $E[(\log m(X_0, X_1))^2] < \infty$, *then* $|\widehat{L} - L| = O_p(\delta_K + \eta_{n,K} + n^{-1/2})$.

Let $\|\cdot\|_\infty$ denote the sup norm. That is, if $f : \mathcal{X} \to \mathbb{R}$ then $\|f\|_\infty = \sup_x\{|f(x)| : x \in \mathcal{X}\}$. Sup-norm rates of convergence of $\widehat{\phi}$ and $\widehat{\phi}^\star$ follow from the $L^2(Q)$ rates by standard arguments for sieve estimation under a slight strengthening of Assumptions 1.4.1(i) and 1.4.1(iv). The sup-norm rates obtained in Corollary 1.4.2 below are useful for constructing estimators of the asymptotic variance of $\widehat{\rho}$, $\widehat{y}$, and $\widehat{L}$.

Assumption 1.4.2. *There exist sequences of functions* $\{g_K : K \geq 1\}$ *and* $\{g_K^* : K \geq 1\}$ *such that* $g_K \in B_K$ *and* $g_K^* \in B_K$ *for each* $K \geq 1$ *and:*

(i) $\|\phi - g_K\|_\infty = O(\delta_K)$

(ii) $\|\phi^* - g_K^*\|_\infty = O(\delta_K^*)$.

A sufficient condition for Assumption 1.4.1(i) and 1.4.2(i) is that $\mathbb{M}$ maps the $L^2(Q)$ unit ball to a subspace $\mathcal{S} \subset L^2(Q)$ over which the sieve has uniformly good approximation properties in sup-norm, i.e. $\{\mathbb{M}\psi : \|\psi\| \leq 1\} \subseteq \mathcal{S}$, and $\sup_{f\in\mathcal{S}} \inf_{b(f)\in B_K} \|f - b(f)\|_\infty = O(\delta_K)$. This condition can be motivated by imposing smoothness conditions on the integral kernel $\mathcal{K}$ and using an appropriate sieve, as in the example below. Assumption 1.4.2(ii) is a sufficient condition for Assumption 1.4.1(iv) by virtue of the relation $\|\phi^* - g_K^*\| \leq \|\phi^* - g_K^*\|_\infty$.

Corollary 1.4.2. *Under Assumptions 1.3.1, 1.4.1, and 1.4.2,*

(i) $\|\widehat{\phi} - \phi\|_\infty = O_p(\zeta_0(K)(\delta_K + \eta_{n,K}))$

(ii) $\|\widehat{\phi}^\star / \|\widehat{\phi}^\star\| - \phi^* / \|\phi^*\|\|_\infty = O_p(\zeta_0(K)(\delta_K^* + \eta_{n,K}))$.

Example: Smooth kernel

This example shows how to calculate δ_K and $\bar{\eta}_{n,K}$ under primitive smoothness conditions on the integral kernel $\mathcal{K}$ defined in expression (1.19). For any $p > 0$ let $[p]$ denote the maximum integer less than or equal to p. Let $C^{[p]}(\mathcal{X})$ denote the space of $[p]$-times continuously differentiable functions with support $\mathcal{X}$. Given a d-tuple $\alpha = (\alpha_1, \ldots, \alpha_d)$ of nonnegative integers, set $|\alpha| = \alpha_1 + \ldots + \alpha_d$ and let D^α denote the differential operator

$$D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}} .$$

Define the Hölder norm $\| \cdot \|_{\infty,p}$ on $C^{[p]}(\mathcal{X})$ by

$$\|f\|_{\infty,p} = \max_{|\alpha| \leq [p]} \sup_{x \in \mathcal{X}} |D^\alpha f(x)| + \max_{|\alpha| = [p]} \sup_{x,x' \in \mathcal{X}, x \neq x'} \frac{|D^\alpha f(x) - D^\alpha f(x')|}{\|x - x'\|^{p-[p]}} .$$

Let $\Lambda^p(\mathcal{X}) = \{f \in C^{[p]}(\mathcal{X}) \text{ such that } \|f\|_{\infty,p} < \infty\}$ denote the Hölder space of p-smooth functions and let $\Lambda_c^p(\mathcal{X}) = \{f \in \Lambda^p(\mathcal{X}) \text{ such that } \|f\|_{\infty,p} \leq c\}$ denote the Hölder ball of smoothness p and radius c.

Let Assumptions 1.3.1 and 1.4.1(iii) hold and assume additionally that (i) $\mathcal{X} \subset \mathbb{R}^d$ is compact, rectangular and has nonempty interior, (ii) q is continuous and uniformly bounded away from zero on $\mathcal{X}$, (iii) there is a $p > 0$ and finite constant C such that

$$\int_{\mathcal{X}} (D^\alpha \mathcal{K}(x_0, x_1))^2 \, \mathrm{d}Q(x_1) \leq C^2$$

for each $|\alpha| \leq [p]$ and

$$\int_{\mathcal{X}} (D^\alpha \mathcal{K}(x, x_1) - D^\alpha \mathcal{K}(x', x_1))^2 \, \mathrm{d}Q(x_1) \leq C^2 \|x - x'\|^{2(p-[p])}$$

for each $|\alpha| = [p]$, and (iv) B_K is a spanned by a (tensor product) of polynomial splines of degree $v > p$ with uniformly bounded mesh ratio (see Schumaker (2007)).

Conditions (ii) and (iii) imply that $\{\mathbb{M}f : \|f\| \leq 1\} \subset \Lambda^p_c(\mathcal{X})$ for some finite c and, in particular, that $\phi \in \Lambda^p_c(\mathcal{X})$. Assumptions 1.4.1(i) and 1.4.2(i) are satisfied with $\delta_K = O(K^{-p/d})$ for a polynomial spline sieve under conditions (i) and (iv) (Schumaker, 2007, Chapter 12). Condition (iv) implies that $\zeta_0(K) = O(\sqrt{K})$ (see, for example, Newey (1997)), so $\eta_{n,K} = O(K/\sqrt{n})$ if m is unbounded and $\eta_{n,K} = O(\sqrt{K}/\sqrt{n})$ if m is bounded. The following mean-square and sup-norm convergence rates obtain:

(a) If m is unbounded, choosing $K \asymp n^{d/(2p+2d)}$ yields $\|\widehat{\phi} - \phi\| = O_p(n^{-p/(2p+2d)})$. If $p > \frac{1}{2}d$ this choice of K yields a sup-norm rate of convergence of $\|\widehat{\phi} - \phi\|_\infty = O_p(n^{(d/2-p)/(2p+2d)})$.

(b) If m is bounded, choosing $K \asymp n^{d/(2p+d)}$ yields $\|\widehat{\phi} - \phi\| = O_p(n^{-p/(2p+d)})$. This is the same as the minimax optimal mean-square convergence rate for a nonparametric regression estimator of a p-smooth function of d variables (Stone, 1982). If $p > \frac{1}{2}d$ this choice of K yields a sup-norm rate of convergence of $\|\widehat{\phi} - \phi\|_\infty = O_p(n^{(d/2-p)/(2p+d)})$.

This example shows that reasonable mean-square convergence rates for $\widehat{\phi}$ can be obtained when $\mathcal{K}(x_0, x_1)$ is smooth in x_0. The kernel $\mathcal{K}$ does not necessarily need to be smooth in x_1 to attain these rates. For example, if $m(x_0, x_1) = m(x_1)$ then $\mathcal{K}$ may satisfy the above smoothness conditions provided $f(x_1|x_0)$ is sufficiently smooth in x_0 even if $m(x_1)$ is kinked or discontinuous in x_1. However, such kinks or discontinuities may affect how well ϕ^* can be approximated, because $\mathbb{M}^*$ is an integral operator with kernel $\mathcal{K}^*$ given by $\mathcal{K}^*(x_0, x_1) = \mathcal{K}(x_1, x_0)$.

1.4.4 Asymptotic inference

A feasible means of conducting asymptotic inference for the eigenvalue ρ, the long-term yield y, and the entropy of the permanent component of the SDF L is now provided. The asymptotic distribution of the estimators is derived via a perturbations expansion. This approach is distinct from the usual Taylor-series arguments used in the derivation of the limit distribution of extremum estimators.

Under the regularity conditions below, the estimator $\widehat{\rho}$ is asymptotically linear and its influence function is formed from m, ϕ, ϕ^*, and ρ, i.e.

$$\sqrt{n}(\widehat{\rho}-\rho) = \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\{\phi^*(X_t)m(X_t,X_{t+1})\phi(X_{t+1}) - \rho\phi^*(X_t)\phi(X_t)\} + o_p(1)\,. \tag{1.32}$$

Conveniently, the summands in expression (1.32) form a martingale difference sequence with respect to $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$. The asymptotic distribution for $\widehat{\rho}$ then follows by applying a central limit theorem for martingales with stationary and ergodic differences. The asymptotic distributions of $\widehat{y}$ and $\widehat{L}$ also follow straightforwardly from the expansion (1.32). An additional assumption is needed to ensure the representation (1.32) is valid and that the asymptotic variances of the estimators are well defined.

Assumption 1.4.3. *The following moment and rate conditions are satisfied:*

(i) Either (a) or (b) holds:

(a) m is bounded, $E[\phi(X)^4] < \infty$ and $E[\phi^*(X)^4] < \infty$

(b) $E[m(X_0,X_1)^6] < \infty$, $E[\phi(X)^6] < \infty$ and $E[\phi^*(X)^6] < \infty$

(ii) $\bar{\eta}_{n,K} = o(n^{-1/4})$, $\delta_K = o(n^{-1/2})$, and $\zeta_0(K)\max\{\delta^*_K, \delta_K, \bar{\eta}_{n,K}\} = o(1)$

(iii) $E[|\log m(X_0,X_1)|^2] < \infty$.

Assumption 1.4.3(i)(iii) guarantees that the asymptotic variance of the estimators are well defined. Assumption 1.4.3(ii) ensures the estimation and approximation errors vanish sufficiently quickly that the expansion (1.32) is valid. The condition $\bar{\eta}_{n,K} = o(n^{-1/4})$ is analogous to the requirement in semiparametric extremum estimation that the estimator of the nonparametric part converges at least as fast as $n^{-1/4}$ to obtain $\sqrt{n}$-consistency of the estimator of the parametric part. The condition $\delta_K = o(n^{-1/2})$ ensures the bias term $\rho_K - \rho$ vanishes sufficiently quickly that it does not affect the asymptotic distribution for $\widehat{\rho}$. The condition $\zeta_0(K)\max\{\delta_K^*, \delta_K, \bar{\eta}_{n,K}\} = o(1)$ is used, inter alia, to establish consistency of the asymptotic variance estimators introduced below.

If $\{Z_t\}$ is a real-valued stationary stochastic process, define the long-run variance of $\{Z_t\}$ as

$$\operatorname{lrvar}(Z_t) = \sum_{t=-\infty}^{\infty} E[Z_0 Z_t]\,.$$

Let

$$\begin{aligned} V_\rho &= E\left[\{\phi^*(X_0)m(X_0,X_1)\phi(X_1) - \rho\phi^*(X_0)\phi(X_0)\}^2\right] &(1.33)\\ V_L &= \operatorname{lrvar}(\rho^{-1}\phi^*(X_t)m(X_t,X_{t+1})\phi(X_{t+1}) - \phi^*(X_t)\phi(X_t) \\ &\qquad - \log m(X_t,X_{t+1}) + E[\log m(X_0,X_1)])\,. &(1.34) \end{aligned}$$

Assumptions 1.3.1 and 1.4.1 provide that V_ρ and V_L are well defined.

Theorem 1.4.2. *Under Assumptions 1.3.1, 1.4.1, 1.4.2, and 1.4.3, if $V_\rho > 0$ and $V_L > 0$, then*

(i) $\sqrt{n}(\widehat{\rho} - \rho) \to_d N(0, V_\rho)$

(ii) $\sqrt{n}(\widehat{y} - y) \to_d N(0, \rho^{-2}V_\rho)$

(iii) $\sqrt{n}(\widehat{L} - L) \to_d N(0, V_L)$.

The conditions $V_\rho > 0$ and $V_L > 0$ exclude cases in which the limit distributions of the estimators are degenerate. For example, if $m(x_0, x_1) = c$ for some positive constant c then $\phi(x) = 1$, $\phi^*(x) = 1$, and $\rho = c$ irrespective of the law of motion of the state variables, in which case $V_\rho = 0$ and $V_L = 0$.

With Theorem 1.4.2 in hand it remains to provide consistent variance estimators. The rates of convergence in Theorem 1.4.1 and 1.4.2 are for estimators whose scale has been normalized under the true distribution Q. Let $\widehat{\phi}^f$ and $\widehat{\phi}^{\star f}$ denote versions of $\widehat{\phi}$ and $\widehat{\phi}^\star$ normalized under the empirical measure, so that

$$\frac{1}{n}\sum_{t=0}^{n-1} \widehat{\phi}^f(X_t)^2 = 1, \quad \frac{1}{n}\sum_{t=0}^{n-1} \widehat{\phi}^f(X_t)\widehat{\phi}^{\star f}(X_t) = 1\,.$$

Corollary 1.4.3. *Under Assumptions 1.3.1 and 1.4.1,*

(i) $\|\widehat{\phi}^f - \phi\| = O_p(\delta_K + \bar{\eta}_{n,K})$

(ii) $\|\widehat{\phi}^{\star f} - \phi^*\| = O_p(\delta_K + \delta_K^* + \bar{\eta}_{n,K})\,.$

If, in addition, Assumption 1.4.2 holds, then

(iii) $\|\widehat{\phi}^f - \phi\|_\infty = O_p(\zeta_0(K)(\delta_K + \bar{\eta}_{n,K}))$

(iv) $\|\widehat{\phi}^{\star f} - \phi^*\|_\infty = O_p(\zeta_0(K)(\delta_K + \delta_K^* + \bar{\eta}_{n,K}))$.

The asymptotic variance estimators for $\widehat{\rho}$ and $\widehat{y}$ are constructed by replacing the population quantities in V_ρ and $\rho^{-2}V_\rho$ by feasible sample analogues. To simplify notation, for any $f : \mathcal{X} \to \mathbb{R}$ let $f_t = f(X_t)$, and let $m_{t,t+1} = m(X_t, X_{t+1})$. The estimator of V_ρ is

$$\widehat{V}_\rho = \frac{1}{n}\sum_{t=0}^{n-1}\left(\widehat{\phi}_t^{\star f} m_{t,t+1}\widehat{\phi}_{t+1}^f - \widehat{\rho}\widehat{\phi}_t^{\star f}\widehat{\phi}_t^f\right)^2\,. \tag{1.35}$$

No sample mean correction is required because $\sum_{t=0}^{n-1}(\widehat{\phi}_t^{\star f} m_{t,t+1}\widehat{\phi}_{t+1}^f - \widehat{\rho}\widehat{\phi}_t^{\star f}\widehat{\phi}_t^f) = 0$ by definition of $\widehat{\phi}$ and $\widehat{\phi}^\star$. The estimator of the asymptotic variance of $\widehat{y}$ is $\widehat{\rho}^{-2}\widehat{V}_\rho$.

Estimating V_L requires estimating a long-run variance. The following approach for conducting inference on L uses an orthogonal series long-run variance (OSLRV) estimator of Phillips (2005) in conjunction with fixed-bandwidth asymptotics as in Chen, Liao, and Sun (2014b).[7] The estimator will be asymptotically χ^2-distributed and therefore inconsistent.

[7]There is a large literature on consistent long-run variance estimation using kernel-based truncated lag

However, asymptotic inference for L can still be performed using this OSLRV estimator and the asymptotic distribution for $\widehat{L}$ developed in Theorem 1.4.2: the only difference is that Gaussian critical values are replaced by t critical values.

Let $\{h_j : j \geq 0\}$ be a continuously differentiable orthonormal basis for the space $L^2([0,1], \mathscr{B}([0,1]), Leb)$ (where $\mathscr{B}([0,1])$ denotes the Borel σ-algebra on $[0,1]$ and Leb is Lebesgue measure), such as a cosine basis or a Legendre polynomial basis. Let $h_0 = 1$, whence $\int_0^1 h_j(u)\,\mathrm{d}u = 0$ for each $j \geq 1$ by orthogonality. For each $j = 1, \ldots, J$, define

$$\widehat{\Lambda}_j = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right) \widehat{\Delta}_{t,t+1}\,.$$

where

$$\begin{aligned}
\widehat{\Delta}_{t,t+1} &= \widehat{\rho}^{-1}\left(\widehat{\phi}_t^{\star f} m_{t,t+1} \widehat{\phi}_{t+1}^{f} - \widehat{\rho}\widehat{\phi}_t^{\star f}\widehat{\phi}_t^{f}\right) - \left(\log m_{t,t+1} - \overline{\log m}_n\right) \\
\overline{\log m}_n &= n^{-1} \sum_{t=0}^{n-1} \log m_{t,t+1}\,.
\end{aligned}$$

The OSLRV estimator $\widehat{V}_{L,J}^{os}$ for V_L using J basis functions is defined as

$$\widehat{V}_{L,J}^{os} = \frac{1}{J} \sum_{j=1}^{J} \widehat{\Lambda}_j^2\,. \tag{1.36}$$

The estimator $\widehat{V}_{L,J}^{os}$ is, by definition, guaranteed to be non-negative.

An additional regularity condition is required for the derivation of the limit theory for the OSLRV estimator. To introduce this condition, define the shrinking neighborhood $N_K = \{(f, f^*) \in B_K \times B_K : \|f - \phi\| \leq (\delta_K + \bar{\eta}_{n,K}) \log(\log n)$ and $f^* \in B_K : \|f^* - \phi^*\| \leq$

estimators following Parzen (1957) (standard econometric references include Newey and West (1987) and Andrews (1991)). To ensure consistency of these estimators, the truncation lag is required to increase at an appropriate rate with the sample size. Recent research has shown that, in some circumstances, asymptotic inference using consistent kernel-based truncated-lag estimators can suffer considerable size and power distortions in finite samples. To this end, a literature has developed that explores inference under alternative bandwidth asymptotics (see, e.g., Kiefer, Vogelsang, and Bunzel (2000); Jansson (2004); Müller (2007); Sun, Phillips, and Jin (2008)). Preliminary Monte Carlo simulations (not reported) revealed that the coverage probabilities of asymptotic confidence intervals for L constructed using a consistent kernel-based truncated lag estimator were sensitive to both the choice of kernel and bandwidth.

$(\delta_K + \delta_K^* + \bar{\eta}_{n,K}) \log(\log n)\}$.

Assumption 1.4.4. *The following equicontinuity conditions are satisfied:*

(i) $\sup_{(f,f^*) \in N_K} \sum_{t=0}^{n-1} h_j(\frac{t+1}{n})\{\phi_t^* \phi_t - f_t^* f_t - E[\phi_t^* \phi_t - f_t^* f_t]\} = o_p(n^{1/2})$

(ii) $\sup_{(f,f^*) \in N_K} \sum_{t=0}^{n-1} h_j(\frac{t+1}{n})\{m_{t,t+1}(\phi_t^* \phi_{t+1} - f_t^* f_{t+1}) - E[m_{t,t+1}(\phi_t^* \phi_{t+1} - f_t^* f_{t+1})]\} = o_p(n^{1/2})$.

Assumption 1.4.4 is essentially Assumption 5.2(i) of Chen, Liao, and Sun (2014b) applied in this context. The definition of N_K and the convergence rates of $\widehat{\phi}^f$ and $\widehat{\phi}^{f\star}$ established in Corollary 1.4.3 ensure that $(\widehat{\phi}^f, \widehat{\phi}^{f\star}) \in N_K$ with probability approaching one.

Consistency of the asymptotic variance estimators for $\widehat{\rho}$ and $\widehat{y}$ are now established, together with a means of performing asymptotic inference for $\widehat{L}$ based on the t distribution using the OSLRV estimator $\widehat{V}_{L,J}^{os}$. Let χ_J^2 and t_J denote the χ^2 and t distributions with J degrees of freedom.

Theorem 1.4.3. *Under Assumptions 1.3.1, 1.4.1 1.4.2, and 1.4.3*

(i) $\widehat{V}_\rho \to_p V_\rho$

(ii) $(\widehat{\rho}^{-2})\widehat{V}_\rho \to_p \rho^{-2} V_\rho$.

If, in addition, Assumption 1.4.4 holds, $V_\rho > 0$ *and* $V_L > 0$*, then*

(iii) $\widehat{V}_{L,J}^{os} \to_d J^{-1} V_L \chi_J^2$

(iv) $\sqrt{n}(\widehat{V}_{L,J}^{os})^{-1/2}(\widehat{L} - L) \to_d t_J$.

Theorems 1.4.2 and 1.4.3 together provide a means with which to perform feasible asymptotic inference on ρ, y, and L.

Example: Smooth kernel (continued)

Here $\delta_K = O(K^{-p/d})$. Assume that ϕ^* belongs to a Hölder ball of smoothness s, so that $\delta_K^* = O(K^{-s/d})$.

(a) If m is unbounded then $\bar{\eta}_{n,K} = O(K/\sqrt{n})$ and K may be chosen so that the conditions $\bar{\eta}_{n,K} = o(n^{-1/4})$ and $\delta_K = o(n^{-1/2})$ are satisfied provided $p > 2d$.

(b) If m is bounded then $\bar{\eta}_{n,K} = O(\zeta_0(K)(\log n)/n)$ and K may be chosen so that the conditions $\bar{\eta}_{n,K} = o(n^{-1/4})$ and $\delta_K = o(n^{-1/2})$ are satisfied provided $p > d$.

In either case the remaining condition $\zeta_0(K)\max\{\delta_K^*, \delta_K\} = o(1)$ is satisfied if $s > \frac{1}{2}d$. If, for arguments sake, $\mathbb{M}$ is selfadjoint, then $\phi^* = \phi$ and the condition $\delta_K = o(n^{-1/2})$ can be relaxed to $\delta_K = o(n^{-1/4})$ (Gobet, Hoffmann, and Reiss, 2004, Remark 4.7), in which case it suffices that $p > d$ if m is unbounded and $p > \frac{1}{2}d$ if m is bounded.

1.4.5 Semiparametric efficiency

The semiparametric efficiency bounds for $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ are now derived, and it is shown that the estimators attain their efficiency bounds.[8] The efficiency bound derivations follow the arguments of Greenwood and Wefelmeyer (1995) and Wefelmeyer (1999) (see also Bickel and Kwon (2001)). A tangent space of admissible perturbations to the unknown transition distribution of the state process is first constructed. A nonparametric version of local asymptotic normality holds for the perturbed models. The parameters ρ, y and L are shown to be differentiable with respect to the perturbation of the transition density and their gradients are characterized. The efficient influence function of the estimators are determined by projecting their gradients onto the (closure of the) tangent space. The asymptotic variances of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ are shown to coincide with the second moment of their efficient influence functions, whence efficiency obtains.

Theorem 1.4.4. *Under Assumptions 1.3.1, 1.4.1, 1.4.2 and 1.4.3, the semiparametric efficiency bounds for ρ, y and L are V_ρ, $\rho^{-2}V_\rho$ and V_L, and are achieved by $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$.*

Now consider the somewhat artificial case in which the stationary distribution Q is known but the dynamics of $\{X_t\}$ are still unknown. In this setting the Gram matrix $\mathbf{G}_K$

[8] In practice the true SDF is unknown, so the term "limited information bound" may be more appropriate than "semiparametric efficiency bound".

is known but $\mathbf{M}_K$ is unknown. An alternative estimator for ρ is

$$\check{\rho} = \lambda_{\max}(\mathbf{G}_K^{-1}\widehat{\mathbf{M}}_K)\,.$$

One might expect the asymptotic variance of $\check{\rho}$ to be smaller than that of $\widehat{\rho}$ because $\check{\rho}$ appears to make use of the fact that Q is known. The following theorem shows otherwise.

Theorem 1.4.5. *Under Assumptions 1.3.1, 1.4.1, 1.4.2 and 1.4.3(i)(ii), if $V_\rho > 0$ then*

$$\sqrt{n}(\check{\rho} - \rho) \to_d N(0, V_\rho + W_\rho)$$

where $W_\rho = 2\rho^2 E[(\phi^*(X_0)\phi(X_0) - 1)^2] + \rho^2 \mathrm{lrvar}((\phi^*(X_t)\phi(X_t) - 1))$.

Clearly $W_\rho \geq 0$, and the inequality is strict if $\phi(x)\phi^*(x)$ is non-constant on a set of positive probability. Therefore, the estimator $\widehat{\rho}$ is relatively more efficient than $\check{\rho}$, even though $\check{\rho}$ appears to incorporate the fact that the density is known.[9]

1.5 Extensions

The estimators and large-sample theory presented in Section 1.4 is now extended to study (i) the long-term implications of estimated SDFs, (ii) the long-term implications of SDFs with additional roughness, and (iii) nonparametric sieve estimation of the marginal utility of consumption of a representative agent.

1.5.1 Plugging-in an estimated SDF

Consider the two-stage problem of first estimating a SDF from data, then extracting its long-term implications. Let the data consist of a time series $\{(X_0, R_0), \ldots, (X_n, R_n)\}$ where $R_t = (R_{i,t}, \ldots, R_{d_R,t})'$ is a vector of returns on d_R assets for each t. Assume that the researcher has estimated a SDF, say $\widehat{m}$, from the data. The SDF estimator $\widehat{m}$ could be

[9] If Q is known the semiparametric efficiency bound for ρ may be different from V_ρ. Consequently, $\widehat{\rho}$ may not be semiparametrically efficient when Q is known.

parametric or semi/nonparametric. An example of a parametric SDF estimator $\widehat{m}$ is the consumption CAPM SDF $m(X_t, X_1; \beta, \gamma) = \beta G_{t+1}^{-\gamma}$ evaluated at $(\widehat{\beta}, \widehat{\gamma})$ where $(\widehat{\beta}, \widehat{\gamma})$ are estimated from $\{(X_0, R_0), \ldots, (X_n, R_n)\}$. Semi/nonparametric estimators include the semiparametric consumption CAPMs studied in Gallant and Tauchen (1989) and Fleissig, Gallant, and Seater (2000), nonparametric nonlinear factor models (Bansal and Viswanathan, 1993), models with nonparametric habit formation (Chen and Ludvigson, 2009), and models with recursive preferences and unknown dynamics (Chen, Favilukis, and Ludvigson, 2013).

In this case the matrix $\mathbf{M}_K$ is estimated using

$$\widehat{\mathbf{M}}_K = \frac{1}{n} \sum_{t=0}^{n-1} b^K(X_t) \widehat{m}(X_t, X_{t+1}) b^K(X_{t+1})' . \tag{1.37}$$

The eigenvalue ρ and eigenfunctions ϕ and ϕ^* are estimated by solving the matrix eigenvalue problems (1.26) and (1.27) with $\widehat{\mathbf{M}}_K$ given by (1.37). The estimators of the long-term yield and entropy of the permanent component of the SDF are

$$\begin{aligned} \widehat{y} &= -\log \widehat{\rho} \\ \widehat{L} &= \log \widehat{\rho} - \frac{1}{n} \sum_{t=0}^{n-1} \log \widehat{m}(X_t, X_{t+1}) \end{aligned}$$

by analogy with (1.30) and (1.31). Consistency and convergence rates of the estimators follow under similar conditions to those described in Section 1.4.

Theorem 1.5.1. *Let Assumption 1.3.1 hold, and let Assumption 1.4.1 hold with $\widehat{\mathbf{M}}_K$ as in expression (1.37). Then there is a set whose probability approaches one on which $\widehat{\rho}$ is real and positive and has multiplicity one, and*

(i) $|\widehat{\rho} - \rho| = O_p(\delta_K + \eta_{n,K})$

(ii) $\|\widehat{\phi} - \phi\| = O_p(\delta_K + \eta_{n,K})$

(iii) $\|\widehat{\phi}^\star / \|\widehat{\phi}^*\| - \phi^* / \|\phi^*\|\| = O_p(\delta_K^* + \eta_{n,K})$.

The requirement that Assumption 1.4.1 hold with $\widehat{\mathbf{M}}_K$ as in expression (1.37) is an implicit condition on convergence of $\widehat{m}$ to m.

The asymptotic distribution for $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ will be distorted (relative to the case in which m is known) by the error introduced by replacing m with a first-stage estimator $\widehat{m}$. The form of the asymptotic distribution for $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ will therefore differ depending on the method used to construct $\widehat{m}$. The following high-level assumption is made to establish the asymptotic linear expansion for $\widehat{\rho}$ in this setting.

Assumption 1.5.1. $\sum_{t=0}^{n-1} |\widehat{m}(X_t, X_{t+1}) - m(X_t, X_{t+1})|(1+\phi(X_{t+1})+\phi^*(X_t)) = O_p(n^{1/2})$.

The following Theorem establishes the distortion to the limit distribution of $\widehat{\rho}$ that arises due to the first-stage estimator $\widehat{m}$. The limit distribution of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ can then be derived from this expansion on a case-by-case basis.

Theorem 1.5.2. *Let Assumption 1.3.1, 1.4.1, 1.4.2, and 1.4.3 hold with $\widehat{\mathbf{M}}_K$ as in expression (1.37), and let Assumption 1.5.1 hold. Then*

$$\begin{aligned}\sqrt{n}(\widehat{\rho} - \rho) &= \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \{\phi^*(X_t) m(X_t, X_{t+1}) \phi(X_{t+1}) - \rho \phi^*(X_t) \phi(X_t)\} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \phi^*(X_t)(\widehat{m}(X_t, X_{t+1}) - m(X_t, X_{t+1}))\phi(X_{t+1}) + o_p(1)\,.\end{aligned}$$

Comparing Theorem 1.5.2 with the expansion for $\widehat{\rho}$ when m is known shows that the limit distribution of $\widehat{\rho}$ will be distorted (relative to the known SDF case) by an additional functional of $(\widehat{m} - m)$. The following remark deals with the case in which $\widehat{m}$ is estimated parametrically.

Remark 1.5.1. *Let m be known up to a finite-dimensional parameter $\theta_0 \in \Theta \subseteq \mathbb{R}^{d_\theta}$ and let m be estimated by plugging in a first-stage estimator $\widehat{\theta}$ of θ_0, i.e.*

$$\begin{aligned}m(X_t, X_{t+1}) &= m(X_t, X_{t+1}; \theta_0) \\ \widehat{m}(X_t, X_{t+1}) &= m(X_t, X_{t+1}; \widehat{\theta})\,.\end{aligned}$$

If (a) $\sqrt{n}(\widehat{\theta}-\theta_0)=O_p(1)$, *(b)* $\theta_0 \in \mathrm{int}(\Theta)$, *(c) for all* $(x_0,x_1)\in\mathcal{X}^2$, $m(x_0,x_1;\theta)$ *is twice continuously differentiable in* θ *on a neighborhood* $\Theta_0 \subset \mathrm{int}(\Theta)$ *containing* θ_0, *(d)*

$$E\left[\left\|\frac{\partial m(X_0,X_1;\theta_0)}{\partial\theta}\right\|_2^2\right]<\infty \quad \textit{and} \quad E\left[\left|\frac{\partial m(X_0,X_1;\theta_0)}{\partial\theta_i}\right|\phi^*(X_0)\phi(X_1)\right]<\infty$$

for $i=1,\ldots,d_\theta$, *(e) there exists a* $g:\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ *such that*

$$\sup_{\theta\in\Theta_0}\left\|\frac{\partial^2 m(x_0,x_1;\theta)}{\partial\theta\partial\theta'}\right\|_2 \leq g(x_0,x_1)$$

with $E[g(X_0,X_1)^2]<\infty$ *and* $E[g(X_0,X_1)\phi^*(X_0)\phi(X_1)]<\infty$. *Then, Assumption 1.5.1 is satisfied and, under the remaining conditions of Theorem 1.5.2,*

$$\begin{aligned}\sqrt{n}(\widehat{\rho}-\rho) &= \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\{\phi^*(X_t)m(X_t,X_{t+1})\phi(X_{t+1})-\rho\phi^*(X_t)\phi(X_t)\} \\ &\quad +E\left[\phi^*(X_0)\phi(X_1)\frac{\partial m(X_0,X_1;\theta_0)}{\partial\theta}\right]\sqrt{n}(\widehat{\theta}-\theta_0)+o_p(1)\,.\end{aligned}$$

The limit distribution for $\widehat{\rho}$ when m is estimated semi/nonparametrically may be similarly derived using Theorem 1.5.2.

1.5.2 Roughing-up the SDF

Following Hansen and Scheinkman (2012b, 2013), consider a class of economy characterized by a discrete-time (first-order) Markov state process $\{(X_t,Y_t)\}$ defined on a complete probability space $(\Omega,\mathcal{F},\{\mathcal{F}_t\},\mathbb{P})$ where time is again indexed by $t\in\mathbb{Z}$ and where $\mathcal{F}_t=\sigma(X_t,Y_t,X_{t-1},Y_{t-1},\ldots)$. Assume that the distribution of (X_{t+1},Y_{t+1}) conditioned on (X_t,Y_t) is the same as the joint distribution of (X_{t+1},Y_{t+1}) conditioned on X_t. More compactly,

$$(X_{t+1},Y_{t+1})|(X_t,Y_t)=_d(X_{t+1},Y_{t+1})|X_t \tag{1.38}$$

for all t, where $=_d$ denotes equality in distribution. The "non-causality" condition (1.38) is convenient for dimension reduction: it allows for the SDF to be a function of (X_t,X_{t+1},Y_{t+1})

whilst restricting the class of eigenfunctions to be functions of X only (not functions of (X,Y)).

Let $\{X_t\}$ have support $\mathcal{X} \subseteq \mathbb{R}^d$ and let $\{Y_t\}$ have support $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$. Assume further that the date-t 1-period SDF is now $m(X_t, X_{t+1}, Y_{t+1})$ for some $m : \mathcal{X} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. Define $\mathbb{M} : L^2(Q) \to L^2(Q)$ as the 1-period pricing operator given by

$$\mathbb{M}\psi(x) \quad = \quad E[m(X_t, X_{t+1}, Y_{t+1})\psi(X_{t+1})|X_t = x]\,.$$

The adjoint operator $\mathbb{M}^*$ is defined as

$$\mathbb{M}^*\psi(x) \quad = \quad E[m(X_t, X_{t+1}, Y_{t+1})\psi(X_t)|X_{t+1} = x]\,.$$

The positive eigenfunction problems are again

$$\begin{aligned} \mathbb{M}\phi &= \rho\phi \\ \mathbb{M}^*\phi^* &= \rho\phi^* \end{aligned}$$

with ϕ and ϕ^* positive (almost everywhere). The following regularity conditions are a straightforward extension of Assumption 1.3.1.

Assumption 1.5.2. $\{(X_t, Y_t)\}$ *and m satisfy the following conditions:*

(i) $\{(X_t, Y_t)\}$ *is a strictly stationary and ergodic (first-order) Markov process which satisfies the non-causality condition (1.38), and which has support* $\mathcal{X} \times \mathcal{Y} \subseteq \mathbb{R}^d \times \mathbb{R}^{d_y}$

(ii) the stationary distributions Q of $\{X_t\}$ and Q_y of $\{Y_t\}$ have densities q and q_y (wrt Lebesgue measure) s.t. $q(x) > 0$ and $q_y(y) > 0$ almost everywhere

(iii) (X_0, X_1, Y_1) *has joint density f (wrt Lebesgue measure) s.t. $f(x_0, x_1, y_1) > 0$ almost everywhere and $f(x_0, x_1, y_1)/(q(x_0)q(x_1)q_y(y_1))$ is uniformly bounded away from infinity.*

(iv) $m : \mathcal{X} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ *has* $m(x_0, x_1, y_1) > 0$ *almost everywhere and* $E[m(X_0, X_1, Y_1)^2] < \infty$.

Nonparametric identification of ϕ and ϕ^* in this environment follows similarly.

Theorem 1.5.3. *Under Assumption 1.5.2,*

(i) $\mathbb{M}$ *and* $\mathbb{M}^*$ *have unique (to scale) eigenfunctions* $\phi \in L^2(Q)$ *and* $\phi^* \in L^2(Q)$ *such that* $\phi > 0$ *and* $\phi^* > 0$ *(almost everywhere)*

(ii) ρ *is positive, has multiplicity one, and is the largest element of the spectrum of* $\mathbb{M}$.

Given a candidate SDF m and a time series of data $\{X_0, (X_1, Y_1), \ldots, (X_n, Y_n)\}$, the positive eigenfunctions ϕ and ϕ^* and the eigenvalue ρ can be estimated by solving the matrix eigenvalue problems (1.26) and (1.27) as before, but with $\mathbf{M}_K$ and $\widehat{\mathbf{M}}_K$ given by

$$\mathbf{M}_K = E[b^K(X_0) m(X_0, X_1, Y_1) b^K(X_1)'] \tag{1.39}$$

$$\widehat{\mathbf{M}}_K = \frac{1}{n} \sum_{t=0}^{n-1} b^K(X_t) m(X_t, X_{t+1}, Y_{t+1}) b^K(X_{t+1})' \,. \tag{1.40}$$

The long-run yield and entropy of the permanent component of the SDF are estimated with

$$\begin{aligned} \widehat{y} &= -\log \widehat{\rho} \\ \widehat{L} &= \log \widehat{\rho} - \frac{1}{n} \sum_{t=0}^{n-1} \log m(X_t, X_{t+1}, Y_{t+1}) \end{aligned}$$

by analogy with expressions (1.30) and (1.31).

Consistency, convergence rates, and the asymptotic distribution of the estimators follow by arguments identical to the case dealt with in Section 1.4. However, Assumption 1.5.2 does not characterize the joint weak-dependence properties of $\{(X_t, Y_t)\}$ so an extra assumption is required to establish the limit distribution of $\widehat{L}$. For the remainder of this subsection, let V_ρ and V_L be defined as in expressions (1.33) and (1.34), but with

$m(X_t, X_{t+1}, Y_{t+1})$ in place of $m(X_t, X_{t+1})$. Let

$$\begin{aligned} \psi_L(X_t, X_{t+1}, Y_{t+1}) &= \rho^{-1}\phi^*(X_t)m(X_t, X_{t+1}Y_{t+1})\phi(X_{t+1}) - \phi^*(X_t)\phi(X_t) \\ &\quad - \log m(X_t, X_{t+1}, Y_{t+1}) + E[\log m(X_0, X_1, Y_1)] \\ V_L &= \mathrm{lrvar}(\psi_L(X_t, X_{t+1}, Y_{t+1})) \end{aligned}$$

The following high-level assumption is sufficient to establish the limit distribution of $\widehat{L}$.

Assumption 1.5.3. *The following regularity conditions hold:*

(i) $V_L < \infty$

(ii) $n^{-1/2}\sum_{t=0}^{n-1} \psi_L(X_t, X_{t+1}, Y_{t+1}) \to_d N(0, V_L)$.

Mean-square convergence rates of the eigenfunction estimators and $\sqrt{n}$-asymptotic normality of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ are now established.

Theorem 1.5.4. *Let Assumption 1.5.2 hold, and let Assumption 1.4.1 hold with* $\mathbf{M}_K$ *and* $\widehat{\mathbf{M}}_K$ *as in expressions (1.39) and (1.40). Then there is a set whose probability approaches one on which* $\widehat{\rho}$ *is real and positive and has multiplicity one, and*

(i) $|\widehat{\rho} - \rho| = O_p(\delta_K + \eta_{n,K})$

(ii) $\|\widehat{\phi} - \phi\| = O_p(\delta_K + \eta_{n,K})$

(iii) $\|\widehat{\phi}^\star/\|\widehat{\phi}^*\| - \phi^*/\|\phi^*\|\| = O_p(\delta_K^* + \eta_{n,K})$.

If, in addition, Assumptions 1.4.2 and 1.4.3 hold with $m(X_0, X_1, Y_1)$ *in place of* $m(X_0, X_1)$ *and* $V_\rho > 0$*, then*

(iv) $\sqrt{n}(\widehat{\rho} - \rho) \to_d N(0, V_\rho)$

(v) $\sqrt{n}(\widehat{y} - y) \to_d N(0, \rho^{-2}V_\rho)$.

If, in addition, Assumption 1.5.3 holds and $V_L > 0$*, then* *(vi)* $\sqrt{n}(\widehat{L} - L) \to_d N(0, V_L)$.

The asymptotic variances V_ρ, $\rho^{-2}V_\rho$ and V_L of $\widehat{\rho}$, $\widehat{y}$ and $\widehat{L}$ may be estimated analogously to the case dealt with in Section 1.4. That is, $\widehat{V}_\rho$ and $\widehat{V}^{os}_{L,J}$ are defined as in expression (1.35) and (1.36), respectively, but with $m(X_t, X_{t+1}, Y_{t+1})$ in place of $m(X_t, X_{t+1})$. The estimators $\widehat{V}_\rho$ and $\widehat{\rho}^{-2}\widehat{V}_\rho$ are consistent under the conditions of Theorem 1.5.4(iv)(v). Asymptotic inference based on $\widehat{V}^{os}_{L,J}$ follows under additional regularity.

1.5.3 Application: nonparametric Euler equation estimation

The results of Section 1.5.2 may be used to establish the large sample properties of nonparametric sieve estimators of the marginal utility of consumption of a representative agent, as outlined in Section 1.1.1. The sieve approach outlined below is an alternative to the kernel-based procedure analyzed by Linton, Lewbel, and Srisuma (2011). As in Linton, Lewbel, and Srisuma (2011), the process $\{(X_t, R_{i,t+1})\}$ is required to be stationary and ergodic. This requirement restricts the forms of utility compatible with this analysis. Consider, for example, MU of the form

$$MU_t = MU(C_t, C_{t-1}, Z_t) \tag{1.41}$$

where C_t is aggregate consumption at date t and $X_t = (C_t, C_{t-1}, Z_t)$. A conventional assumption is that aggregate consumption $\{C_t\}$ is nonstationary but growth in aggregate consumption $\{C_t/C_{t-1}\}$ is stationary (see, e.g., Hansen and Singleton (1982); Gallant and Tauchen (1989)). Under this assumption, MU of the form (1.41) is incompatible with the stationarity requirement. If MU in expression (1.41) is homogeneous of degree zero in its first two arguments, MU_t may be rewritten as

$$MU_t = MU(C_t/C_{t-1}, Z_t)\,. \tag{1.42}$$

Marginal utility of the form (1.42) may then be estimated as described below, provided the process $\{(C_t/C_{t-1}, Z_t, R_{i,t+1})\}$ is strictly stationary and ergodic.[10]

Assume $MU_t = MU(X_t)$ where $\{(X_t, R_{i,t})\}$ is strictly stationary and ergodic (here $\{(X_t, R_{i,t})\}$ does not need to be a Markov process). Given $\{X_0, (X_1, R_{i,1}), \ldots, (X_n, R_{i,n})\}$, β and MU can be estimated by solving

$$\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{T}}_{i,K}\widehat{c} = \widehat{\beta}^{-1}\widehat{c}$$

and setting $\widehat{MU}(x) = b^K(x)'\widehat{c}$, where $\widehat{\beta}^{-1}$ is the largest eigenvalue of $\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{T}}_{i,K}$ and

$$\widehat{\mathbf{T}}_{i,K} = \frac{1}{n}\sum_{t=0}^{n-1} b^K(X_t)R_{i,t+1}b^K(X_{t+1})'\,.$$

Let MU^* solve $E[R_{i,t+1}MU^*(X_t)|X_{t+1}] = \beta^{-1}MU^*(X_{t+1})$. Then MU^* may be estimated by solving $\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{T}}'_{i,K}\widehat{c}^\star = \widehat{\beta}^{-1}\widehat{c}^\star$ and setting $\widehat{MU}^\star(x) = b^K(x)'\widehat{c}^\star$.

The large sample properties of $\widehat{\beta}^{-1}$, $\widehat{MU}$ and $\widehat{MU}^\star$ follow from Theorem 1.5.4. Normalize MU and MU^* so that $E[MU(X)^2] = 1$, $E[MU(X)MU^*(X)] = 1$. Without confusion, let Q denote the stationary distribution of $\{X_t\}$. Replace Assumption 1.5.2 with (a) $\mathbb{T}_i : L^2(Q) \to L^2(Q)$ is Hilbert-Schmidt, and (b) $\mathbb{T}_i MU = \beta^{-1}MU$ where $MU \in L^2(Q)$ and $\beta^{-1} > 0$ is the largest eigenvalue of $\mathbb{T}_i$ and has multiplicity one. Also let Assumptions 1.4.1, 1.4.2 and 1.4.3 hold with $\mathbb{T}_i$, $\widehat{\mathbf{T}}_{i,K}$, $\mathbf{T}_{i,K} = E[b^K(X_t)R_{i,t+1}b^K(X_t)']$, MU, and MU^* in place of $\mathbb{M}$, $\widehat{\mathbf{M}}_K$, $\mathbf{M}_K$, ϕ and ϕ^*. Theorem 1.5.4(i)–(iii) establishes consistency and convergence rates of $\widehat{\beta}^{-1}$, $\widehat{MU}$ and $\widehat{MU}^\star$. The limit distribution for $\widehat{\beta}$ is more subtle. Let $\mathcal{H}_t = \sigma(X_t, R_{i,t}, X_{t-1}, R_{i,t-1}, \ldots)$. If $(X_{t+1}, R_{i,t+1})|X_t =_d (X_{t+1}, R_{i,t+1})|\mathcal{H}_t$ for all t, then

$$\sqrt{n}(\widehat{\beta} - \beta) \to_d N(0, \beta^4 E[\{MU(X_{t+1})R_{i,t+1}MU^*(X_t) - \beta^{-1}MU(X_t)MU^*(X_t)\}^2])$$

by Theorem 1.5.4(iv) and the delta method. Otherwise, simple modification of the proof

[10] Chen and Ludvigson (2009) use a similar homogeneity assumption to rewrite a semiparametric habit formation model in terms of consumption growth (rather than levels of consumption).

of Theorem 1.4.2 yields, under regularity,

$$\sqrt{n}(\widehat{\beta} - \beta) \to_d N(0, \beta^4 W_\beta)$$

where $W_\beta = \text{lrvar}(\{MU(X_{t+1})R_{i,t+1}MU^*(X_t) - \beta^{-1}MU(X_t)MU^*(X_t)\})$.

1.6 Monte Carlo simulation

The following Monte Carlo (MC) exercise explores the performance of the estimators when applied to a stylized consumption CAPM. The SDF is

$$m(X_t, X_{t+1}) = \beta \exp(-\gamma g_{t+1})$$

where β is the time preference parameter, γ is the risk aversion parameter, and g_{t+1} is log consumption growth from time t to $t+1$. The state variable is simply $X_t = g_t$. The data are constructed to be somewhat representative of U.S. real monthly aggregate consumption growth. Log consumption growth evolves as the Gaussian AR(1)

$$g_{t+1} - \mu = \kappa(g_t - \mu) + \sigma e_{t+1}$$

where the e_t are i.i.d. $N(0,1)$ random variables. Gaussianity of the disturbances ensures that the process $\{g_t\}_{t=-\infty}^{\infty}$ is time reversible (see, e.g., Weiss (1975)), which is used to obtain a closed-form solution for ϕ^*. The positive eigenfunction and adjoint positive eigenfunction are

$$\begin{aligned}
\phi(g) &= \exp\left(-\frac{\gamma\kappa}{1-\kappa}g + \frac{\mu\gamma\kappa}{1-\kappa} - \frac{\gamma^2\kappa^2\sigma^2}{(1-\kappa^2)(1-\kappa)^2}\right) \\
\phi^*(g) &= \exp\left(-\frac{\gamma}{1-\kappa}g + \frac{\mu\gamma}{1-\kappa} + \frac{\gamma^2\sigma^2}{(1-\kappa^2)(1-\kappa)^2}\left\{\kappa^2 - \frac{1}{2}(1+\kappa)^2\right\}\right)
\end{aligned}$$

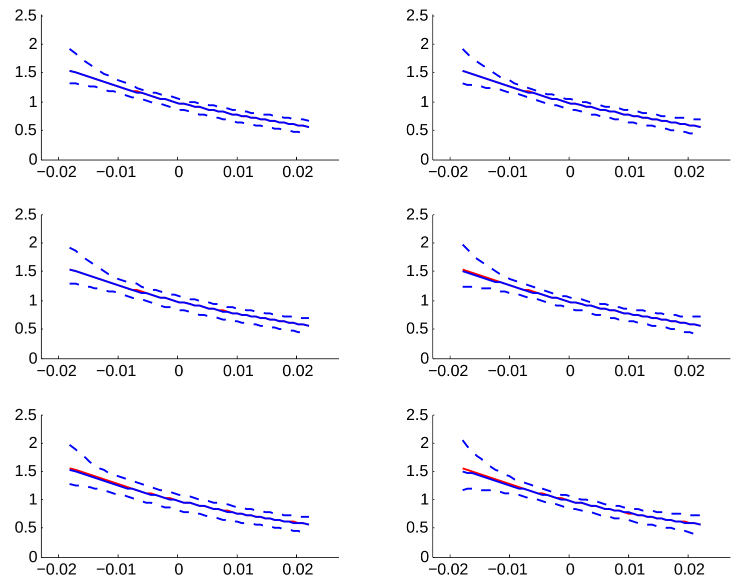

Figure 1.1: MC plots for $\widehat{\phi}^f$ with $\gamma = 25$. Each panel shows the true ϕ (solid red line), pointwise MC median (solid blue line), and pointwise MC 90% confidence bands (dashed lines). Results are presented for Hermite polynomial (left) and B-spline (right) sieves of dimension 6 (top), 10 (middle) and 14 (bottom).

where both ϕ and ϕ^* have been normalized so that $E[\phi^2(g)] = 1$ and $E[\phi(g)\phi^*(g)] = 1$. Their eigenvalue ρ is

$$\rho \;=\; \beta \exp\left(-\gamma\mu + \frac{1}{2}\frac{\gamma^2}{(1-\kappa)^2}\sigma^2\right) .$$

The parameters for the simulation are $\mu = 0.002$, $\kappa = 0.3$, and $\sigma = 0.01/\sqrt{1-\kappa^2}$, which are similar in magnitude to the parameters of the U.S. real per capita consumption growth series investigated in the next section. The sample length is set to 500, and 10000 simulations are performed. The time preference parameter β is set to 0.998, and γ is varied from 0 to 30. Two choices of sieve are used, namely Hermite polynomials and cubic B-splines, with dimension $K = 6$, 10 and 14. For each simulation, the Hermite polynomial sieve was centered and scaled by the sample mean and sample standard deviation of g, and the knots of the cubic B-spline sieve were placed at the empirical quantiles. Cosine bases of dimension $J = 10$ and 15 were used to compute the OSLRV estimator.

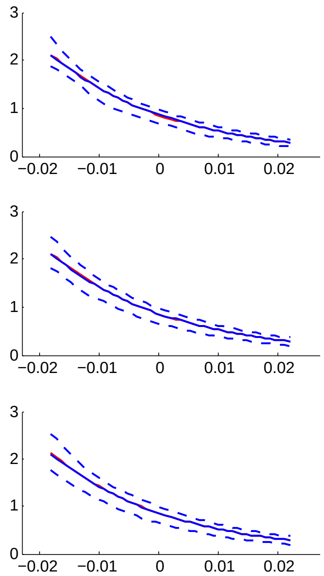

Figure 1.2: MC plots for $\widehat{\phi}^{\star f}$ with $\gamma = 25$. Each panel shows the true ϕ (solid red line), pointwise MC median (solid blue line), and pointwise MC 90% confidence bands (dashed lines). Results are presented for Hermite polynomial (left) and B-spline (right) sieves of dimension 6 (top), 10 (middle) and 14 (bottom).

MC results $\widehat{\phi}^{f}$ and $\widehat{\phi}^{\star f}$ for $\gamma = 25$ are presented in Figures 1.1 and 1.2, respectively. Each panel shows the true ϕ (or ϕ^*) for $g \in [\mu - 2\sigma, \mu + 2\sigma]$ (solid red lines) together with the pointwise MC median $\widehat{\phi}$ (or $\widehat{\phi}^{\star}$) (solid blue lines) and pointwise MC 90% confidence bands (the pointwise .05 and .95 quantiles of the estimator approximated by simulation; dashed lines). Both $\widehat{\phi}$ and $\widehat{\phi}^{\star}$ are normalized feasibly as in Corollary 1.4.3. The estimators have negligible bias. The width of the confidence bands increases with the sieve dimension K, which illustrates that the "variance term" $\bar{\eta}_{n,K}$ is increasing in K. Other simulations (not reported) show that increasing/decreasing γ also increases/decreases the width of the MC confidence bands.

Table 1.1 shows the MC coverage probabilities for 90% and 95% confidence intervals (CIs) for ρ and y, and Table 1.2 shows the MC coverage probabilities for 90% and 95% CIs for L. To construct the MC coverage probabilities, for each simulation ρ, y, and L were estimated and their 90% and 95% confidence intervals estimated using the variance estimators $\widehat{V}_{\rho}$, $\widehat{\rho}^{-2}\widehat{V}_{\rho}$, and $\widehat{V}^{os}_{L,J}$. Gaussian critical values were used for the CIs for ρ and y,

	90% CI for ρ			95% CI for ρ		
	$K=6$	$K=10$	$K=14$	$K=6$	$K=10$	$K=14$
$\gamma=5$	89.60	89.64	89.75	94.82	94.78	94.79
H-Pol $\gamma=15$	89.86	89.91	90.04	94.52	94.54	94.60
$\gamma=25$	89.43	89.53	89.39	93.63	93.59	93.61
$\gamma=5$	89.63	89.60	89.54	94.82	94.75	94.74
B-Spl $\gamma=15$	89.80	89.63	89.57	94.50	94.51	94.39
$\gamma=25$	89.28	88.64	88.41	93.51	93.27	92.95
	90% CI for y			95% CI for y		
	$K=6$	$K=10$	$K=14$	$K=6$	$K=10$	$K=14$
$\gamma=5$	89.70	89.66	89.74	94.81	94.80	94.77
H-Pol $\gamma=15$	89.91	89.90	90.02	94.62	94.60	94.61
$\gamma=25$	89.44	89.58	89.50	93.78	93.76	93.79
$\gamma=5$	89.68	89.58	89.39	94.80	94.68	94.53
B-Spl $\gamma=15$	89.78	89.53	89.42	94.53	94.50	94.18
$\gamma=25$	89.33	88.68	88.24	93.67	93.35	92.91

Table 1.1: Monte Carlo coverage probabilities for 90% and 95% asymptotic confidence intervals for ρ and y based on the asymptotic distribution in Theorem 1.4.2 and the consistent variance estimators $\widehat{V}_\rho$ and $\widehat{\rho}^{-2}\widehat{V}_\rho$. Results are presented for Hermite polynomial (H-Pol) and B-spline (B-Spl) sieves of varying dimension K.

and t_J critical values were used for the CIs for L. The MC coverage probabilities are the proportion of simulations for which the estimated CIs contained the true parameter values. Table 1.1 shows that the 90% and 95% CIs for ρ and y have MC coverage probabilities that are very close to their nominal coverage probabilities, for all sieve choices and all levels of γ. The MC coverage probabilities for L presented in Table 1.2 show that the CIs corresponding to a B-spline sieve are too narrow, especially at high values of γ. The coverage probabilities for the CIs corresponding to a Hermite polynomial sieve are close to their nominal values with both $J = 10$ and $J = 15$. The MC coverage probabilities for ρ, y and L appear generally robust to the dimension K of the sieve space, especially when a Hermite polynomial sieve is used.

1.7 Empirical illustration

The long-run implications of the consumption CAPM are now investigated using the tools introduced in this chapter. The consumption CAPM has been the basis for a vast amount of

		90% CI for L			95% CI for L		
		$K=6$	$K=10$	$K=14$	$K=6$	$K=10$	$K=14$
	$\gamma=5$	91.66	91.70	91.80	95.92	95.96	95.90
H-Pol, $J=10$	$\gamma=15$	90.35	90.34	90.30	94.91	94.96	94.98
	$\gamma=25$	88.69	88.47	88.28	93.67	93.53	93.18
	$\gamma=5$	92.95	93.06	93.03	96.74	96.80	96.84
H-Pol, $J=15$	$\gamma=15$	91.11	91.20	91.08	95.47	95.52	95.53
	$\gamma=25$	89.41	89.11	88.94	94.27	93.98	93.86
	$\gamma=5$	88.26	88.20	87.84	92.98	92.71	92.62
B-Spl, $J=10$	$\gamma=15$	86.69	86.12	85.87	91.55	91.16	90.80
	$\gamma=25$	84.47	83.63	83.09	89.36	88.96	88.51
	$\gamma=5$	89.72	89.67	89.51	93.99	93.83	93.70
B-Spl, $J=15$	$\gamma=15$	87.49	86.98	86.75	92.29	91.83	91.66
	$\gamma=25$	85.29	84.58	84.10	89.93	89.48	88.97

Table 1.2: Monte Carlo coverage probabilities for 90% and 95% asymptotic confidence intervals for L based on the asymptotic distribution in Theorem 1.4.3 and the OSLRV estimator $\widehat{V}^{os}_{L,J}$, which was computed with a cosine basis of dimension $J=10$ and $J=15$. Results are presented for Hermite polynomial (H-pol) and B-spline (B-Spl) sieves for of varying dimension K.

research, from the seminal works of Hansen and Singleton (1982) and Mehra and Prescott (1985) though to recent rare disasters-based investigations of Backus, Chernov, and Martin (2011) and Julliard and Ghosh (2012). The SDF to be investigated is

$$m(X_t, X_{t+1}; \beta, \gamma) = \beta \exp(-\gamma g_{t+1}) \tag{1.43}$$

where β is the time preference parameter, γ is the risk aversion parameter, and g_{t+1} is log consumption growth from time t to $t+1$. As shown in Bansal and Lehmann (1997), Hansen (2012) and Backus, Chernov, and Zin (2012), the SDF (1.43) has the same permanent component (and therefore entropy of the permanent component) and implies the same long-term yield as SDFs of the form

$$m(X_t, X_{t+1}; \beta, \gamma) = \beta \exp(-\gamma g_{t+1}) \frac{h(X_{t+1})}{h(X_t)} \tag{1.44}$$

where h is a positive function. For example, $h(X_t)$ may capture a limiting version of recursive preferences as in Hansen (2012). Alternatively, $h(X_t)$ may be an external habit formation component as in Chen and Ludvigson (2009). The following analysis therefore applies to a wider class of consumption-based asset pricing models than simply the consumption CAPM.

Three specifications of the state process are investigated, namely $X_t = g_t$, $X_t = (g_t, g_{e,t})'$ where $g_{e,t}$ denotes the growth in corporate earnings from time $t-1$ to time t, and $X_t = (g_t, r_{f,t})'$ where $r_{f,t} = \log R_{1,t+1}$ denotes the short-term risk-free rate at date t. Corporate earnings growth is included as a state variable in line with Hansen, Heaton, and Li (2008) who, in a different but related application, model log consumption and log corporate earnings jointly using a Gaussian vector autoregression.[11] The risk-free rate is included in the state process for comparison with Case I of Bansal and Yaron (2004), in which log consumption growth is modeled as

$$\begin{aligned} g_{t+1} &= \mu + x_t + \sigma_g e_{t+1} \\ x_{t+1} &= \rho x_t + \sigma_x \eta_{t+1} \end{aligned} \tag{1.45}$$

where x_t is a latent predictable component of consumption growth and e_{t+1} and η_{t+1} are mutually independent and i.i.d. $N(0,1)$. When $\rho \in (-1,1)$ the state vector $X_t = (g_t, x_t)'$ is a strictly stationary and ergodic first-order Markov process. The risk-free rate in Case I of Bansal and Yaron (2004) is an affine function of x_t. Therefore, in Case I of Bansal and Yaron (2004) the observable vector $(g_t, r_{f,t})'$ and the partially latent vector $(g_t, x_t)'$ contain the same information: one can simply rewrite x_t as an affine function of $r_{f,t}$. Both Hansen, Heaton, and Li (2008) and Bansal and Yaron (2004) assume a representative agent with Epstein-Zin-Weil recursive preferences. The SDF (1.43) is a restricted parameterization of the recursive preferences SDF used in these models (obtained by setting the elasticity of intertemporal substitution equal to γ^{-1}). As is common practice, it is assumed that the household decision interval coincides with the sampling interval.

[11] Hansen, Heaton, and Li (2008) model log consumption and log corporate earnings as a cointegrated, fifth-order Gaussian vector autoregression. By contrast, here log consumption growth and log earnings growth are modeled jointly as a first-order (possibly nonlinear) Markov process.

	g	r_f	g_e
Mean	0.0053	0.0030	0.0115
Std Dev	0.0055	0.0069	0.0733
Skewness	-0.4631	-0.4747	-0.1722
Kurtosis	4.2770	5.3039	8.2600
AR(1)	0.2859	0.7446	0.0016

Table 1.3: Summary statistics for quarterly U.S. per capita (log) growth in consumption of nondurables and services g, risk-free rate r_f, and quarterly growth in corporate earnings g_e. AR(1) denotes the first order autocorrelation coefficient. The data span 1947:Q2–2012:Q4.

1.7.1 Data

The data span 1947:Q2 to 2012:Q4 (263 observations). Data on aggregate consumption, corporate earnings, and population size were sourced from the National Income and Product Accounts (NIPA) tables. The consumption growth series is formed by taking seasonally adjusted consumption of nondurables and services data (NIPA Table 2.3.5), deflating by the implicit price deflator for personal consumption expenditures (PCE; NIPA Table 2.3.4), and then calculating per capita growth rates using the deflated series and population data (NIPA Table 2.1). After-tax corporate earnings data (NIPA Table 1.12) were used because dividends are paid out to stockholders on an after-tax basis (Longstaff and Piazzesi, 2004). The risk-free rate was taken as the three-month Treasury bill rate (from CRSP). The earnings growth and risk-free series were both converted to real rates using the PCE deflator data.

Summary statistics of the consumption growth, risk-free rate and corporate earnings growth series are presented in Table 1.3. All series exhibit negative skewness and excess kurtosis. The consumption growth and risk-free rate series are positively autocorrelated. Earnings growth exhibits little persistence, but is much more volatile than consumption growth.

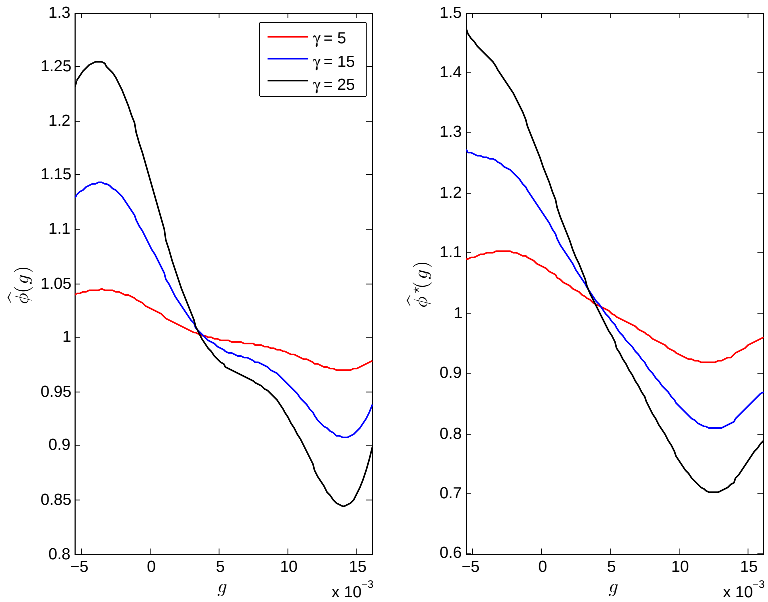

Figure 1.3: Estimated ϕ and ϕ^* for the consumption CAPM at different levels of risk aversion γ. The state variable is $X_t = g_t$, where g_t is quarterly real U.S. per capita (log) growth in consumption of nondurables and services.

1.7.2 Implementation

The time preference parameter was set at $\beta = 0.998^3$. The risk aversion parameter was varied between $\gamma = 0$ (risk neutrality) and $\gamma = 30$. Hermite polynomial bases were formed for each series (centering and scaling by the sample mean and standard deviation of each series). A sieve spaces of dimension $K = 8$ is used for $X_t = g_t$, and sieve spaces of dimension $K = 16$ are used for $X_t = (g_t, r_{f,t})'$ and $X_t = (g_t, g_{e,t})'$. The sieve spaces for bivariate state vectors are formed by taking the tensor product of two univariate bases of dimension four. The OSLRV estimator $\widehat{V}^{os}_{L,J}$ was implemented with a cosine basis of dimension $J = 10$. The estimates were insensitive to both the dimension of the sieve space and the dimension of the basis used to compute the OSLRV estimators.

1.7.3 Results

Figure 1.3 displays the estimates $\widehat{\phi}$ and $\widehat{\phi}^\star$ with $\gamma = 5$, $\gamma = 15$, and $\gamma = 25$ for the case $X_t = g_t$. The estimates $\widehat{\phi}$ and $\widehat{\phi}^\star$ are more acutely sloped for higher levels of γ.

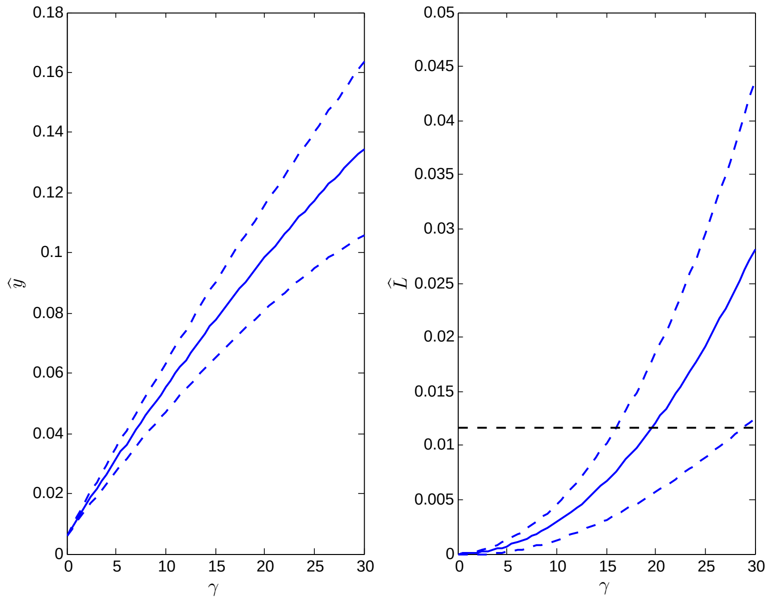

Figure 1.4: Estimated long-run yield $\widehat{y}$ and entropy of the permanent component of the SDF $\widehat{L}$ for the consumption CAPM at different levels of risk aversion γ, for $\beta = 0.998^3$ (solid blue lines). The state variable is $X_t = g_t$, where g_t is quarterly real U.S. per capita (log) consumption growth. Dashed blue lines are pointwise asymptotic 90% confidence bands. The dashed black line represents the estimated average quarterly excess return on equities relative to long-term bonds.

The estimated positive eigenfunctions are decreasing in g, which implies that the price of long-horizon assets is a decreasing function of aggregate consumption growth.

The estimated long-run yield $\widehat{y}$ and entropy of the permanent component $\widehat{L}$ are plotted in Figure 1.4 for $X_t = g_t$ and in Figure 1.5 for $X_t = (g_t, g_{e,t})'$. The entropy of the permanent component of the SDF is independent of β. The long-run yield depends on both β and γ. The solid blue lines present the pointwise estimates, and the dashed blue lines are 90% pointwise confidence bands. Comparison of Figures 1.4 and 1.5 show that very similar estimated long-run yields and permanent component entropies are obtained for $X_t = g_t$ or $X_t = (g_t, g_{e,t})'$. Similar results are also obtained for the specification $X_t = (g_t, r_{f,t})$ (not presented).

The entropy of the permanent component of the SDF is an upper bound for the average return on risky assets relative to long-term bonds (see equation (1.14)). An average excess return of 1.17% per quarter was estimated from the quarterly return on the

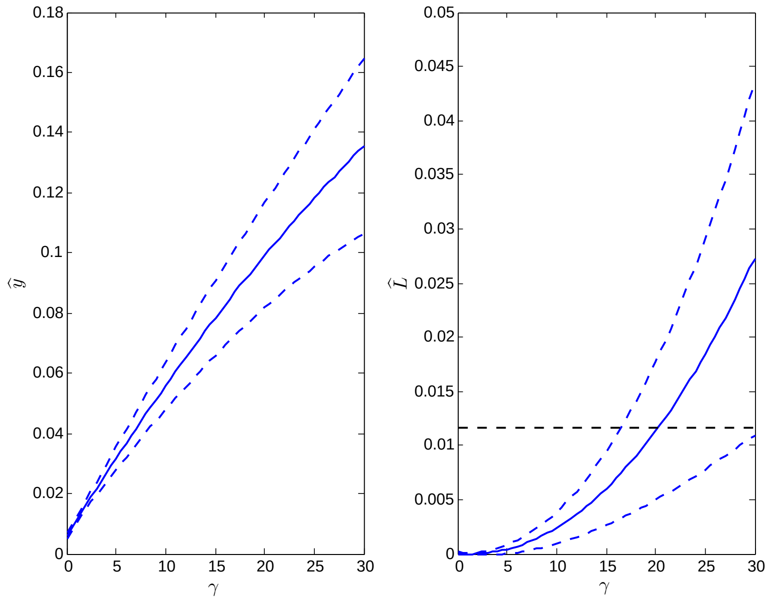

Figure 1.5: Estimated long-run yield $\widehat{y}$ and entropy of the permanent component of the SDF $\widehat{L}$ for the consumption CAPM at different levels of risk aversion γ, for $\beta = 0.998^3$ (solid blue lines). The state variable is $X_t = (g_t, g_{e,t})'$, where g_t is quarterly real U.S. per capita (log) consumption growth and $g_{e,t}$ is quarterly real U.S. corporate earnings growth. Dashed blue lines are pointwise asymptotic 90% confidence bands. The dashed black line represents the estimated average quarterly excess return on equities relative to long-term bonds.

NYSE/AMEX/NASDAQ combined index, including dividends, relative to the quarterly return on 30-year bonds over the sample period (both return series were sourced from CRSP).[12] The estimates presented in Figures 1.4 and 1.5 show that one requires $\gamma \geq 15$ for the estimated entropy of the permanent component of the consumption CAPM SDF to exceed 1.17%. As the bound (1.14) applies to all risky assets (not just the aggregate market), the lower bound for the entropy of the permanent component of the SDF would be at least as large as 1.17% if information on other assets was taken into account. A larger γ would then be required to generate and estimated entropy that was compatible with a higher bound. This analysis suggests that the level of risk aversion required for the

[12] The effect of coupon payments is ignored. Ignoring coupon payments is unlikely to have any substantial effect on the qualitative implications of these findings, however. The estimated quarterly premium for the combined index in excess of the three-month T-bill rate was 1.46% over the sample period. Historical data show that the term premium is small. For example, Backus, Chernov, and Zin (2012) suggest that the absolute value of the average yield spread is unlikely to exceed 0.1% monthly (see also Alvarez and Jermann (2005)).

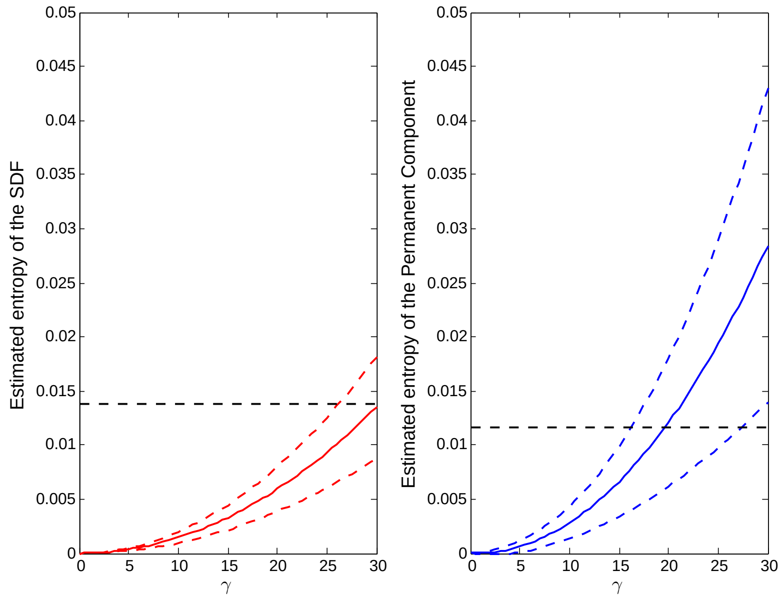

Figure 1.6: Estimated entropy of the SDF and entropy of the permanent component of the SDF for the consumption CAPM at different levels of risk aversion γ, for $\beta = 0.998^3$ (solid red and blue lines). The state variable is $X_t = g_t$, where g_t is quarterly real U.S. per capita (log) consumption growth. Dashed lines are pointwise asymptotic 90% confidence bands. The dashed black lines represent historical average quarterly excess returns on equities relative to short-term bonds (left panel) and relative to long-term bonds (right panel).

entropy of the permanent component of the consumption CAPM SDF to be compatible with historical average returns on equity to relative to long-term bonds is substantially higher than the threshold of 10 imposed by Mehra and Prescott (1985).

As Figures 1.4 and 1.5 show, the estimated long-term yield is much larger than historical average long-term yields when γ is set sufficiently high to rationalize the average return on equities relative to long-term bonds. For $\gamma \geq 15$ the estimated long-term *quarterly* yield is at least 6% per quarter. By contrast, the average real yield on the longest maturity Treasury bond over the period February 1959 to December 2012 was 0.76% per quarter.[13] Decreasing the time preference parameter β further increases the estimated long-term yields. Under the restriction $\gamma \geq 15$, estimates of the long-term yield in line with historical average yields on long-term bonds were only obtainable with $\beta > 1$. Again, very similar results are

[13]This yield is estimated by taking the maximum treasury constant maturity yield available each month (either 20 or 30 years) from the Federal Reserve H-15 release, adjusting to a quarterly yield, and deflating using the implicit price deflator for personal consumption expenditures in the NIPA tables.

obtained with $X_t = (g_t, r_{f,t})'$ (not presented)

These findings provide evidence of a long-term version of the equity premium and risk-free rate puzzles under the restrictions $\beta \in (0,1)$ and $\gamma \in [0,10]$ imposed by Mehra and Prescott (1985), at least to the extent that U.S. aggregate consumption growth can be described as a stationary Markov process with low-dimensional state vector. Similar qualitative results are obtained with monthly data.[14]

How does the entropy of the SDF in the consumption CAPM compare with the entropy of its permanent component? Figure 1.6 presents estimates of the entropy of the consumption CAPM SDF (left panel) and the entropy of its permanent component (right panel), together with their 90% pointwise confidence bands.[15] The dashed horizontal lines are the estimated average returns on equities relative to short-term bonds (left panel) and relative to long-term bonds (right panel) over the sample period.[16] The entropy of the SDF is an upper bound for the average return on risky assets relative to short-term bonds (see expression (1.17)). Figure 1.6 shows that the level of risk aversion required to generate an entropy of the SDF that rationalizes the historical average return on equities relative to short-term bonds is considerably larger than the level required to generate an entropy of the permanent component that rationalizes the historical average return on equities relative to long-term bonds. It may be possible to augment the consumption CAPM SDF by a term of the form $h(X_{t+1})/h(X_t)$ as in expression (1.44) so as to rationalize the premium relative to short term bonds at lower levels of risk aversion. However, such transitory distortions will not alter the permanent component of the SDF.

[14]Bakshi and Chabi-Yo (2012) estimate a monthly return premium in excess of long-term bonds of 0.41% per month, from U.S. market data spanning January 1932 to December 2010. With monthly data (spanning February 1959–December 2012), the level of γ required to generate an estimated entropy compatible with this bound was in excess of 20 for $X_t = g_t$ and $X_t = (g_t, r_{f,t})'$ (corporate earnings are not available at the monthly frequency). Moreover, the estimated long-run yields were at least 3% per month for values of γ larger than 20. By contrast, the average real yield on the longest maturity Treasury bond over the sample period is around 0.25% per month.

[15]The entropy of the SDF is estimated by replacing the expected values in expression (1.16) by their sample averages. Confidence bands for the estimated entropy of the SDF are computed using the OSLRV estimator with a cosine basis of dimension $J = 10$.

[16]The return relative to short-term bonds is estimated from the quarterly return on the NYSE/AMEX/NASDAQ combined index, including dividends, relative to the three-month T-bill rate.

1.8 Conclusion

The long-run implications of a dynamic asset pricing model are jointly determined by both the functional form of the SDF and the short-run dynamics, or law of motion, of the variables in the model. The econometric framework introduced in this chapter treats the dynamics as an unknown nuisance parameter. This chapter introduces nonparametric sieve estimators of the positive eigenfunction and its eigenvalue (which are used to decompose the SDF into its permanent and transitory components), the long-term yield, and the entropy of the permanent component of the SDF. The sieve estimators are particularly easy to implement, and may also be used to numerically compute the long-run implications of fully specified models for which analytical solutions are unavailable. Consistency and convergence rates of the estimators are established, together with a means of performing asymptotic inference on the eigenvalue, long-run yield and entropy of the permanent component of the SDF. The estimators of the eigenvalue, long-run yield and entropy of the permanent component are shown to be semiparametrically efficient. Nonparametric identification conditions are presented for the positive eigenfunction in stationary discrete-time environments, and a version of the long-run pricing result of Hansen and Scheinkman (2009) is shown to obtain under the identification conditions.

There are several ways in which the research reported in this chapter may be extended. One such extension is to study nonparametric identification and estimation in environments in which the state variable contains latent components or is measured with error. Data-driven methods for choosing the sieve space dimension would provide a more objective means for choosing the sieve space dimension than the ad hoc approach used in this chapter. Confidence bands for the estimated eigenfunction and the asymptotic distribution of functionals of the estimated eigenfunction would be useful for performing inference on the estimated eigenfunction. These extensions are currently being investigated by the author.

1.9 Proofs of main results

Proof of Theorem 1.4.1. First verify the conditions of Theorems 2.3.1 and 2.3.2 for $\mathbb{M}$. Assumption 2.3.1 is satisfied for $\mathbb{M}$ under Assumption 1.3.1 by Theorem 1.3.1. Assumption 2.3.2(i) and the part of Assumption 2.3.2(ii) pertaining to ϕ is trivially satisfied by Assumptions 1.4.1(i). The remaining condition in Assumption 2.3.2(ii) is satisfied by Assumption 1.4.1(iv) because

$$\|(\mathbb{M}^*\Pi_K^b - \mathbb{M}^*)\phi^*\| \leq \|\mathbb{M}^*\|\|\Pi_K^b\phi^* - \phi^*\| = \|\mathbb{M}\|\|\Pi_K^b\phi^* - \Pi_K^b h_K^* + h_K^* - \phi^*\| \leq 2\|\mathbb{M}\|\|\phi^* - h_K^*\|$$

which is $O(\delta_K^*)$.

Let $\|\cdot\|_{HS}$ denote the Hilbert-Schmidt norm and recall $\|\mathbb{M}\|_{HS} < \infty$ under Assumption 1.3.1. The bound

$$\|R(z, \mathbb{M})\| \leq \frac{1}{d(z, \sigma(\mathbb{M}))} \exp\left(\frac{1}{2} + 2\frac{\|\mathbb{M}\|_{HS}^2}{d(z, \sigma(\mathbb{M}))^2}\right)$$

obtains for any $z \in \mathbb{C} \setminus \sigma(\mathbb{M})$ (see, e.g. Bandtlow (2004)). Let $\{e_k : k \geq 1\}$ be an orthonormal basis for $L^2(Q)$ such that $\{e_k : 1 \leq k \leq K\}$ are an orthonormal basis for B_K. As Hilbert-Schmidt norms are invariant to the choice of basis,

$$\|\Pi_K^b \mathbb{M}|_{B_K}\|_{HS}^2 = \sum_{k=1}^{K} \|\Pi_K^b \mathbb{M} e_k\|^2 \leq \sum_{k=1}^{\infty} \|\Pi_K^b \mathbb{M} e_k\|^2 \leq \sum_{k=1}^{\infty} \|\mathbb{M} e_k\|^2 = \|\mathbb{M}\|_{HS}^2 .$$

Therefore, $\Pi_K^b \mathbb{M}|_{B_K}$ is Hilbert-Schmidt and the bound

$$\|R(z, \Pi_K^b \mathbb{M}|_{B_K})\| \leq \frac{1}{d(z, \sigma(\Pi_K^b \mathbb{M}|_{B_K}))} \exp\left(\frac{1}{2} + 2\frac{\|\mathbb{M}\|_{HS}^2}{d(z, \sigma(\Pi_K^b \mathbb{M}|_{B_K}))^2}\right)$$

obtains for any $z \in \mathbb{C} \setminus \sigma(\Pi_K^b \mathbb{M}|_{B_K})$. The function $r : (0, \infty) \to (0, \infty)$ given by

$$r(x) = \frac{1}{x} \exp\left(\frac{1}{2} + 2\frac{\|\mathbb{M}\|_{HS}}{x^2}\right)$$

is continuous and strictly decreasing, verifying Assumption 2.3.3.

Assumption 2.3.4(i) is trivially satisfied by Assumption 1.4.1(iii). Assumption 2.3.4(ii) and (iii) are satisfied by Assumption 1.4.1(ii) and definition of $\bar{\eta}_{n,K}$ and $\eta_{n,K}$.

Parts (i) and (ii) are straightforward applications of Theorems 2.3.1 and 2.3.2. For part (iii) it is enough to show that $\|\phi_K^*/\|\phi_K^*\| - \phi_K^\star/\|\phi_K^\star\|\| = O(\delta_K^*)$, which follows from Assumptions 1.4.1(iv). $\square$

Proof of Corollary 1.4.1. The rate of convergence of $\widehat{y}$ follows immediately from Theorem 1.4.1 by continuity of log on a neighborhood of ρ. For $\widehat{L}$, first write

$$|\widehat{L} - L| \leq |\widehat{y} - y| + \left| \frac{1}{n} \sum_{t=0}^{n-1} \log m(X_t, X_{t+1}) - E[\log m(X_0, X_1)] \right| .$$

It is enough to show that the second term on the right-hand side is $O_p(n^{-1/2})$. This follows by Chebychev's inequality, using the condition $E[\log m(X_0, X_1)^2] < \infty$ and Lemma 1.10.2. $\square$

Proof of Corollary 1.4.2. For any $f_K \in B_K$, the sup and $L^2(Q)$ norms are related by

$$\|f_K\|_\infty^2 \leq \underline{\lambda}^{-1} \zeta_0(K)^2 \|f_K\|^2 .$$

By Assumption 1.4.2(i) and the triangle inequality

$$\begin{aligned} \|\phi - \widehat{\phi}\|_\infty &\leq \|\phi - g_K\|_\infty + \|g_K - \widehat{\phi}\|_\infty \\ &\leq O(\delta_K) + \zeta_0(K)\underline{\lambda}^{-1/2}\|g_K - \widehat{\phi}\| \\ &\leq O(\delta_K) + \zeta_0(K)\underline{\lambda}^{-1/2}(\|g_K - \phi\| + \|\phi - \widehat{\phi}\|) \\ &= O_p(\zeta_0(K)(\delta_K + \eta_{n,K})) \end{aligned}$$

where the final line is by Theorem 1.4.1 and the fact that the sup norm dominates the $L^2(Q)$ norm. This proves part (i); the proof of part (ii) is similar. $\square$

Proof of Theorem 1.4.2. Part (i): Lemma 1.10.4 shows that the representation (1.32) is valid. Assumption 1.4.3(i) and the Hölder inequality imply V_ρ is finite. The central limit theorem for martingales with stationary and ergodic differences (Billingsley, 1961) then yields

$$\sqrt{n}(\widehat{\rho} - \rho) \to_d N(0, V_\rho)$$

whenever $V_\rho > 0$.

Part(ii): The asymptotic distribution for $\widehat{y}$ then follows by the delta method.

Part (iii): Let $\bar{\ell}_n = n^{-1} \sum_{t=1}^{n} \ell(X_t, X_{t+1})$ where

$$\ell(X_t, X_{t+1}) = \log m(X_t, X_{t+1}) - E[\log m(X_t, X_{t+1})]\,.$$

By definition of $\widehat{L}$ and the asymptotic linear expansion for $\widehat{\rho}$,

$$\begin{aligned} & \sqrt{n}(\widehat{L} - L) \\ = \; & \sqrt{n}(\log \widehat{\rho} - \log \rho - \bar{\ell}_n) \\ = \; & \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \left\{ \rho^{-1} \phi^*(X_t) m(X_t, X_{t+1}) \phi(X_{t+1}) - \phi^*(X_t)\phi(X_t) - \ell(X_t, X_{t+1}) \right\} + o_p(1)\,. \end{aligned}$$

The summands are strictly stationary geometrically phi-mixing random variables by Assumption 1.3.1 and Lemma 1.10.1, have mean zero, and have and finite second moment by Assumptions 1.4.3. Application of Lemma 1.10.2 provides that

$$V_L = \lim_{n\to\infty} \frac{1}{n} E\left[\left(\sum_{t=0}^{n-1} \left\{ \rho^{-1}\phi^*(X_t) m(X_t, X_{t+1}) \phi(X_{t+1}) - \phi^*(X_t)\phi(X_t) - \ell(X_t, X_{t+1}) \right\} \right)^2 \right]$$

is finite. The result follows by a CLT for strictly stationary phi-mixing sequences (Peligrad, 1985, Corollary 2.2). □

Proof of Corollary 1.4.3. Part (i): Write $\|\widehat{\phi}^f - \phi\| \leq \|\widehat{\phi}^f - \widehat{\phi}\| + \|\widehat{\phi} - \phi\|$ where $\|\widehat{\phi}\| = 1$. Theorem 1.4.1 gives $\|\widehat{\phi} - \phi\| = O_p(\delta_K + \eta_{n,K})$ so it remains to control $\|\widehat{\phi}^f - \widehat{\phi}\|$. Note that

$\widehat{\phi} = \widehat{c}'b^K$ where $\widehat{c}'\mathbf{G}_K\widehat{c} = 1$ and $\widehat{\phi}^f = \widehat{c}^{f\prime}b^K$ where $\widehat{c}^f = \widehat{c}/(\widehat{c}'\widehat{\mathbf{G}}_K\widehat{c})^{1/2}$. Therefore,

$$
\begin{aligned}
\|\widehat{\phi}^f - \widehat{\phi}\| &= \left|\frac{1}{(\widehat{c}'\widehat{\mathbf{G}}_K\widehat{c})^{1/2}} - 1\right| \\
&= \left|\frac{1}{(\widehat{c}'\mathbf{G}_K^{1/2}\widehat{\widehat{\mathbf{G}}}_K\mathbf{G}_K^{1/2}\widehat{c})^{1/2}} - 1\right| .
\end{aligned}
$$

The minimax characterization of eigenvalues of symmetric matrices (Kato, 1980, Section I.6.10) implies that

$$
\lambda_{\min}(\widehat{\widehat{\mathbf{G}}}_K) \leq (\widehat{c}'\mathbf{G}_K^{1/2}\widehat{\widehat{\mathbf{G}}}_K\mathbf{G}_K^{1/2}\widehat{c})^{1/2} \leq \lambda_{\max}(\widehat{\widehat{\mathbf{G}}}_K) .
$$

Moreover,

$$
\begin{aligned}
\max\{|\lambda_{\max}(\widehat{\widehat{\mathbf{G}}}_K) - 1|, |\lambda_{\min}(\widehat{\widehat{\mathbf{G}}}_K) - 1|\} &= \max\{|\lambda_{\max}(\widehat{\widehat{\mathbf{G}}}_K - I_K)|, |\lambda_{\min}(\widehat{\widehat{\mathbf{G}}}_K - I_K)|\} \\
&= \|\widehat{\widehat{\mathbf{G}}}_K - I_K\|_2 \\
&= O_p(\bar{\eta}_{n,K})
\end{aligned}
$$

by definition of $\bar{\eta}_{n,K}$. This proves $\|\widehat{\phi}^f - \widehat{\phi}\| = O_p(\bar{\eta}_{n,K})$.

Part (iii): By the relation between the $L^2(Q)$ norm and sup norm on B_K and Assumption 1.4.2(i),

$$
\begin{aligned}
& \|\widehat{\phi}^f - \phi\|_\infty \\
\leq\ & \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi}^f - \widehat{\phi}\| + \|\widehat{\phi} - \phi\|_\infty \\
\leq\ & \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi}^f - \widehat{\phi}\| + \|\widehat{\phi} - g_K\|_\infty + \|g_K - \phi\|_\infty \\
\leq\ & \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi}^f - \widehat{\phi}\| + \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi} - g_K\| + \|g_K - \phi\|_\infty \\
\leq\ & \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi}^f - \widehat{\phi}\| + \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi} - \phi\| + \underline{\lambda}^{-1/2}\zeta_0(K)\|\phi - g_K\| + \|g_K - \phi\|_\infty .
\end{aligned}
$$

The result follows by Part (i), Assumption 1.4.2(i), and Theorem 1.4.1.

Part (ii): Write $\|\widehat{\phi}^{\star f} - \phi^*\| \leq \|\widehat{\phi}^{\star f} - \widehat{\phi}^{\star}\| + \|\widehat{\phi}^{\star} - \phi^*\|$ where $E[\widehat{\phi}(X)\widehat{\phi}^{\star}(X)] = 1$. Theorems 2.3.1 and 2.3.2 show that $\|\widehat{\phi}^{\star} - \phi^*\| = O_p(\delta_K + \bar{\eta}_{n,K})$. It remains to control $\|\widehat{\phi}^{\star f} - \widehat{\phi}^{\star}\| = O_p(\bar{\eta}_{n,K})$. Note that $\widehat{\phi}^{\star} = \widehat{c}^{\star\prime} b^K$ where $\widehat{c}^{\star\prime}\mathbf{G}_K\widehat{c} = 1$ and $\widehat{\phi}^f = \widehat{c}^{\star f \prime} b^K$ where $\widehat{c}^{\star f} = \widehat{c}^{\star}(\widehat{c}'\widehat{\mathbf{G}}_K\widehat{c})^{1/2}/(\widehat{c}^{\star\prime}\widehat{\mathbf{G}}_K\widehat{c})$. Therefore,

$$\|\widehat{\phi}^{\star f} - \widehat{\phi}^{\star}\| = \left| \frac{(\widehat{c}'\widehat{\mathbf{G}}_K\widehat{c})^{1/2}}{\widehat{c}^{\star\prime}\widehat{\mathbf{G}}_K\widehat{c}} - 1 \right| \|\widehat{\phi}^{\star}\|$$

where $\|\widehat{\phi}^{\star}\| = O_p(1)$ by Theorem 1.4.1(ii), and the proof of Part (i) shows $(\widehat{c}'\widehat{\mathbf{G}}_K\widehat{c})^{1/2} = 1 + O_p(\bar{\eta}_{n,K})$. Moreover,

$$\begin{aligned}
\widehat{c}^{\star\prime}\widehat{\mathbf{G}}_K\widehat{c} &= \widehat{c}^{\star\prime}\mathbf{G}_K\widehat{c} + \widehat{c}^{\star\prime}(\widehat{\mathbf{G}}_K - \mathbf{G}_K)\widehat{c} \\
&= 1 + \widehat{c}^{\star\prime}\mathbf{G}_K^{1/2}(\widehat{\widehat{\mathbf{G}}}_K - I_K)\mathbf{G}_K^{1/2}\widehat{c} \\
&\leq 1 + \|\widehat{\phi}^{\star}\|\|\widehat{\widehat{\mathbf{G}}}_K - I_K\|_2 \\
&= 1 + O_p(1) \times O_p(\bar{\eta}_{n,K})
\end{aligned}$$

by Theorem 1.4.1(iii), definition of $\bar{\eta}_{n,K}$, and the normalization $\|\widehat{\phi}\| = 1$. Thus $\|\widehat{\phi}^{\star f} - \widehat{\phi}^{\star}\| = O_p(\bar{\eta}_{n,K})$.

Part (iv): Arguing as in the proof of Part (iii) yields

$$\|\widehat{\phi}^{\star f} - \phi^*\|_\infty \leq \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi}^{f\star} - \widehat{\phi}^{\star}\| + \underline{\lambda}^{-1/2}\zeta_0(K)\|\widehat{\phi}^{\star} - \phi^*\| + \underline{\lambda}^{-1/2}\zeta_0(K)\|\phi^* - g_K^*\| + \|g_K^* - \phi^*\|_\infty\,.$$

The result follows by Part (ii), Assumption 1.4.2(ii), and $\|\widehat{\phi}^{\star} - \phi^*\| = O_p(\delta_K + \bar{\eta}_{n,K})$. $\square$

Proof of Theorem 1.4.3. Part (i): By addition and subtraction of terms,

$$
\begin{aligned}
\widehat{V}_\rho - V_\rho \;=\;& \frac{1}{n}\sum_{t=0}^{n-1}\left(\widehat{\phi}_t^{\star f2} m_{t,t+1}^2 \widehat{\phi}_{t+1}^{f2} - \phi_t^{*2} m_{t,t+1}^2 \phi_{t+1}^2\right) \\
&+\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2} m_{t,t+1}^2 \phi_{t+1}^2 - E[\phi^*(X_0)^2 m(X_0,X_1)^2\phi(X_1)^2] \qquad (1.46)\\
&+\frac{1}{n}\sum_{t=0}^{n-1}\left(\widehat{\rho}^2\widehat{\phi}_t^{\star f2}\widehat{\phi}_t^{f2} - \rho^2\phi_t^{*2}\phi_t^2\right) \qquad (1.47)\\
&+\frac{1}{n}\sum_{t=0}^{n-1}\rho^2\phi_t^{*2}\phi_t^2 - \rho^2 E[\phi^*(X_0)\phi(X_0)] \qquad (1.48)\\
&-\frac{2}{n}\sum_{t=0}^{n-1}\left(\widehat{\rho}\widehat{\phi}_t^{\star f2} m_{t,t+1}\widehat{\phi}_t^f\widehat{\phi}_{t+1}^f - \rho\phi_t^{*2} m_{t,t+1}\phi_t\phi_{t+1}\right) \qquad (1.49)\\
&-\frac{2}{n}\sum_{t=0}^{n-1}\rho\phi_t^{*2} m_{t,t+1}\phi_t\phi_{t+1} + 2\rho E[\phi^*(X_0)^2 m(X_0,X_1)\phi(X_0)\phi(X_1)] \qquad (1.50)
\end{aligned}
$$

Terms (1.46), (1.48) and (1.50) are all $o_{a.s}(1)$ by the ergodic theorem (the expectations exist by Assumption 1.4.3(i)). For term (1.46),

$$
\begin{aligned}
(1.46) \;=\;& \frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2} m_{t,t+1}^2[\widehat{\phi}_{t+1}^{f2} - \phi_{t+1}^2] + \frac{1}{n}\sum_{t=0}^{n-1}[\widehat{\phi}_t^{\star f2} - \phi_t^{*2}] m_{t,t+1}^2[\widehat{\phi}_{t+1}^{f2} - \phi_{t+1}^2]\\
&+\frac{1}{n}\sum_{t=0}^{n-1}[\widehat{\phi}_t^{\star f2} - \phi_t^{*2}] m_{t,t+1}^2\phi_{t+1}^2\,.
\end{aligned}
$$

Using the relation $(a^2-b^2) = (a+b)(a-b)$, the triangle inequality, and the sup-norm convergence rates established in Corollary 1.4.3,

$$
\begin{aligned}
|(1.46)| \;\le\;& O_p(\zeta_0(K)(\delta_K^* + \delta_K + \bar{\eta}_{n,K}))\Big(\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2} m_{t,t+1}^2|\widehat{\phi}_{t+1}^f + \phi_{t+1}|\\
&+\frac{1}{n}\sum_{t=0}^{n-1}|\widehat{\phi}_t^{\star f} + \phi_t^*| m_{t,t+1}^2|\widehat{\phi}_{t+1}^f + \phi_{t+1}| + \frac{1}{n}\sum_{t=0}^{n-1}|\widehat{\phi}_t^{\star f} + \phi_t^*| m_{t,t+1}^2\phi_{t+1}^2\Big)\,.
\end{aligned}
$$

Writing

$$|\widehat{\phi}_{t+1}^f + \phi_{t+1}| \le 2\phi_{t+1} + \|\widehat{\phi}^f - \phi\|_\infty$$

and similarly for $\widehat{\phi}^{\star f}$, the condition $\zeta_0(K)(\delta_K^* + \delta_K + \bar{\eta}_{n,K}) = o(1)$ (by Assumption 1.4.3(ii)) and sup-norm convergence rates in Corollary 1.4.2 yield

$$
\begin{aligned}
|(1.46)| \quad \leq \quad & o_p(1)\Big(2\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2}m_{t,t+1}^2\phi_{t+1} + o_p(1)\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*2}m_{t,t+1}^2 + 4\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*}m_{t,t+1}^2\phi_{t+1} \\
& + o_p(1)\frac{1}{n}\sum_{t=0}^{n-1}m_{t,t+1}^2 + o_p(1)\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*}m_{t,t+1}^2 + o_p(1)\frac{1}{n}\sum_{t=0}^{n-1}m_{t,t+1}^2\phi_{t+1} \\
& + 2\frac{1}{n}\sum_{t=0}^{n-1}\phi_t^{*}m_{t,t+1}^2\phi_{t+1}^2 + o_p(1)\frac{1}{n}\sum_{t=0}^{n-1}m_{t,t+1}^2\phi_{t+1}^2\Big)\,.
\end{aligned}
$$

All sample averages in this display are of the form

$$
\frac{1}{n}\sum_{t=0}^{n-1}\phi^*(X_t)^k m(X_t, X_{t+1})^2\phi(X_{t+1})^l
$$

with $0 \leq k, l \leq 2$, and are therefore all $O_{a.s.}(1)$ by the ergodic theorem (all moments exist by Assumption 1.4.3(i)). Therefore term (1.46) is $o_p(1)$. Similar arguments show that terms (1.47) and (1.49) are both $o_p(1)$.

Part (ii): Immediate from Part (i) and consistency of $\widehat{\rho}$.

Parts (iii) and (iv): For each $j = 1, \ldots, J$, write

$$
\begin{aligned}
\widehat{\Lambda}_j \quad = \quad & \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}h_j\left(\frac{t+1}{n}\right)\Delta_{t,t+1} + \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}h_j\left(\frac{t+1}{n}\right)(\widehat{\Delta}_{t,t+1} - \Delta_{t,t+1}) \\
=: \quad & \Lambda_j + \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}h_j\left(\frac{t+1}{n}\right)(\widehat{\Delta}_{t,t+1} - \Delta_{t,t+1})
\end{aligned}
$$

where

$$
\Delta_{t,t+1} = \rho^{-1}\left(\phi_t^{\star f}m_{t,t+1}\phi_{t+1}^f - \rho\phi_t^{\star f}\phi_t^f\right) - (\log m_{t,t+1} - E[\log m(X_0, X_1)])
$$

Writing out term-by-term gives

$$\widehat{\Lambda}_j - \Lambda_j = \left(\frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)\right)\sqrt{n}(\overline{\log m}_n - E[\log m(X_0,X_1)]) \tag{1.51}$$

$$+\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)(\phi_t^*\phi_t - \widehat{\phi}_t^\star\widehat{\phi}_t) \tag{1.52}$$

$$+\widehat{\rho}^{-1}\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)(\widehat{\phi}_t^\star\widehat{\phi}_{t+1} - \phi_t^*\phi_{t+1})m_{t,t+1} \tag{1.53}$$

$$+\sqrt{n}(\widehat{\rho}^{-1} - \rho^{-1}) \times \frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)\phi_t^*\phi_{t+1}m_{t,t+1}\,. \tag{1.54}$$

Term (1.51) is $o_p(1)$ because $\frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right) = O(n^{-1/2})$ (by numerical integration, using $\int_0^1 h(u)\,\mathrm{d}u = 0$) and $\sqrt{n}(\overline{\log m}_n - E[\log m(X_0,X_1)]) = O_p(1)$ (by Markov's inequality, using the fact that $\{X_t\}$ is geometrically rho-mixing under Assumption 1.3.1, and that enough moments exist by Assumption 1.4.3(iii)).

For term (1.52), write

$$(1.52) = \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)\{\phi_t^*\phi_t - \widehat{\phi}_t^{\star f}\widehat{\phi}_t^f - E[\phi_t^*\phi_t - \widehat{\phi}_t^{\star f}\widehat{\phi}_t^f]\} + E[\phi_t^*\phi_t - \widehat{\phi}_t^{\star f}\widehat{\phi}_t^f]\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)$$

The first term in this display is $o_p(1)$ by Assumption 1.4.4(i). The second term in this display is $o_p(1)$ because $\frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right) = O(n^{-1/2})$ and

$$\begin{aligned} |E[\phi_t^*\phi_t - \widehat{\phi}_t^{\star f}\widehat{\phi}_t^f]| &\leq \|\phi^*\|\|\phi - \widehat{\phi}^f\| + \|\widehat{\phi}^f\|\|\phi^* - \widehat{\phi}^{\star f}\| \\ &= O_p(\delta_K + \delta_K^* + \bar{\eta}_{n,K}) \end{aligned}$$

with the first line by the triangle and Cauchy-Schwarz inequalities, and the second line by Corollary 1.4.3(i)(ii).

A similar argument shows term (1.53) is $o_p(1)$, using Assumption 1.4.4(ii) and Corollary 1.4.3(i)(ii).

For term (1.54), $\sqrt{n}(\widehat{\rho}^{-1} - \rho^{-1}) = O_p(1)$ by Theorem 1.4.2 and the delta method. For the remaining term, write

$$\begin{aligned}
& \frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)\phi_t^*\phi_{t+1}m_{t,t+1} \\
= \; & \frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)\{\phi_t^*\phi_{t+1}m_{t,t+1} - \rho E[\phi^*(X_0)\phi(X_0)]\} \qquad (1.55)\\
& +\rho E[\phi^*(X_0)\phi(X_0)] \times \frac{1}{n}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right). \qquad (1.56)
\end{aligned}$$

Term (1.56) is $O(n^{-1/2})$. Consider the process $\{\phi_t^*\phi_{t+1}m_{t,t+1} - \rho E[\phi^*(X_0)\phi(X_0)]\}$ and let V_ϕ denote its long-run variance. V_ϕ is finite by geometric rho-mixing of $\{X_t\}$ and the moment assumptions in Assumption 1.4.3(i). Moreover, straightforward calculation shows

$$V_\phi = V_\rho + 2\rho^2 E[(\phi^*(X_0)\phi(X_0)-1)^2] + \rho^2 \textstyle\sum_{t=-\infty}^{\infty} E[(\phi^*(X_0)\phi(X_0)-1)(\phi^*(X_t)\phi(X_t)-1)]$$

whence $V_\phi \geq V_\rho > 0$. Therefore, the process $\{\phi_t^*\phi_{t+1}m_{t,t+1} - \rho E[\phi^*(X_0)\phi(X_0)]\}$ satisfies an invariance principle under Assumption 1.3.1 (which implies the process is phi-mixing); see Corollary 2.2 of Peligrad (1985). Functional limit and Wiener integration arguments yield

$$\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} h_j\left(\frac{t+1}{n}\right)\{\phi_t^*\phi_{t+1}m_{t,t+1} - \rho E[\phi^*(X_0)\phi(X_0)]\} \to_d N(0, V_\phi)$$

since $\int_0^1 h(u)^2\,\mathrm{d}u = 1$. Therefore, term (1.55) is $O_p(n^{-1/2})$, and so term (1.54) is $o_p(1)$.

Finally, the process $\{\Delta_{t,t+1}\}$ satisfies an invariance principle under Assumption 1.3.1 (which implies the process is geometrically phi-mixing) and Assumption 1.4.3 (which guarantees enough moments); see Corollary 2.2 of Peligrad (1985). Therefore, by the functional limit and Wiener integration arguments in Phillips (2005),

$$\begin{aligned}
\begin{pmatrix} \sqrt{n}(\widehat{L}-L) & \widehat{\Lambda}_1 & \cdots & \widehat{\Lambda}_J \end{pmatrix}' &= \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\Delta_{t,t+1} & \Lambda_1 & \cdots & \Lambda_J \end{pmatrix}' + o_p(1) \\
&\to_d N(0, V_L \times I_{J+1})
\end{aligned}$$

and the result follows by definition of the χ^2_J and t_J distributions. $\square$

Proof of Theorem 1.4.4. Efficiency bound for ρ: The tangent space is first characterized as in Greenwood and Wefelmeyer (1995) (see also Wefelmeyer (1999); Greenwood, Schick, and Wefelmeyer (2001)). Let $\mathcal{BM}(\mathcal{X} \times \mathcal{X})$ denote the space of all real-valued bounded measurable functions on $\mathcal{X} \times \mathcal{X}$,[17] and define

$$\mathcal{T} = \{h \in \mathcal{BM}(\mathcal{X} \times \mathcal{X}) : \, E[h(X_0, X_1)|X_0 = x] = 0 \text{ for all } x \in \mathcal{X}\}\,.$$

Let $f(x_1|x_0)$ denote the true transition density of $X_1 = x_1$ given $X_0 = x_0$ (this exists by Assumption 1.3.1). For any $h \in \mathcal{T}$ there is $N_h \in \mathbb{N}$ such that for all $n \geq N_h$ the function

$$f_{n,h}(x_1|x_0) = f(x_1|x_0)\{1 + n^{-1/2}h(x_0, x_1)\}$$

is non-negative and integrates to 1 for every x_0, and is therefore a legitimate transition density.

Let $P_{n,h}$ denote the distribution of the sample $\{X_0, X_1, \ldots, X_n\}$ under the perturbed transition density $f_{n,h}$ and $P_{n,0}$ denote the distribution of the sample $\{X_0, X_1, \ldots, X_n\}$ under the true transition density f. A version of local asymptotic normality is known to obtain for this set of perturbed transition densities, i.e.

$$\log \frac{\mathrm{d}P_{n,h}}{\mathrm{d}P_{n,0}} = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} h(X_i, X_{i+1}) - \frac{1}{2} E[h(X_0, X_1)^2] + o_{P_{n,0}}(1)$$

(see Greenwood and Wefelmeyer (1995); Wefelmeyer (1999); Greenwood, Schick, and Wefelmeyer (2001)) where

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} h(X_i, X_{i+1}) \to_d N(0, E[h(X_0, X_1)^2])$$

[17] It suffices to consider bounded measurable functions as the are dense in the space $\{f : \mathcal{X} \times \mathcal{X} \to \mathbb{R} \text{ s.t. } f \text{ is measurable and } E[f(X_0, X_1)^2] < \infty\}$ (Wefelmeyer, 1999).

by a central limit theorem for martingales with stationary and ergodic differences (Billingsley, 1961).

The gradient of ρ is now characterized in terms of the transition density. For any $h \in \mathcal{T}$ define the perturbed pricing operator $\mathbb{M}_{n,h} : L^2(Q) \to L^2(Q)$ by

$$\mathbb{M}_{n,h}\psi(x) = \int_{\mathcal{X}} m(x,y)\frac{f_{n,h}(y|x)}{q(y)}\psi(y)\,\mathrm{d}Q(y)$$

and let its kernel be defined as

$$\mathcal{K}_{n,h}(y,x) = m(x,y)\frac{f_{n,h}(y|x)}{q(y)}\,.$$

Whenever $h \in \mathcal{T}$

$$\int_{\mathcal{X}}\int_{\mathcal{X}}(\mathcal{K}(y,x) - K_{n,h}(y,x))^2\,\mathrm{d}Q(x)\mathrm{d}Q(y) \leq Cn^{-1}E[m(X_0,X_1)^2h(X_0,X_1)^2]$$

for some finite positive constant C under Assumption 1.3.1. This implies that $\|\mathbb{M}-\mathbb{M}_{n,h}\| = O(n^{-1/2})$ since the Hilbert-Schmidt norm dominates the operator norm. Application of Lemma 2.5.2 shows that for n sufficiently large, the maximum eigenvalue $\rho_{n,h}$ of $\mathbb{M}_{n,h}$ is real and positive and

$$\rho_{n,h} = \rho + E[\phi^*(X_0)(\mathbb{M}_{n,h} - \mathbb{M})\phi(X_0)] + o(n^{-1/2})\,.$$

By this and the law of iterated expectations,

$$\sqrt{n}(\rho_{n,h} - \rho) = E[\phi^*(X_0)m(X_0,X_1)\phi(X_1)h(X_0,X_1)] + o(1)$$

where the expectation is finite for all $h \in \mathcal{T}$ under Assumption 1.4.3(i). The gradient of ρ

is $\phi^*(x_0)m(x_0,x_1)\phi(x_1)$ and its projection onto (the closure of) $\mathcal{T}$ is

$$\begin{aligned}\tilde{\psi}_\rho(x_0,x_1) &= \phi^*(x_0)m(x_0,x_1)\phi(x_1) - E[\phi^*(X_0)m(X_0,X_1)\phi(X_1)|X_0=x_0] \\ &= \phi^*(x_0)m(x_0,x_1)\phi(x_1) - \phi^*(x_0)\mathbb{M}\phi(x_0) \\ &= \phi^*(x_0)m(x_0,x_1)\phi(x_1) - \rho\phi^*(x_0)\phi(x_0)\,.\end{aligned}$$

Therefore $\tilde{\psi}_\rho(x_0,x_1)$ is the efficient influence function and $E[\tilde{\psi}_\rho(X_0,X_1)^2] = V_\rho$ is the efficiency bound for ρ. Theorem 1.4.2 shows $\widehat{\rho}$ attains this bound. Efficiency bound for y: follows (by continuity) from the efficiency bound for ρ.

Efficiency bound for L: As shown in Greenwood and Wefelmeyer (1995) and Wefelmeyer (1999),[18] the efficient influence function for estimating $E[\log m(X_0,X_1)]$ is

$$\begin{aligned}\tilde{\psi}_m(x_0,x_1) &= \log m(x_0,x_1) - E[\log m(X_0,X_1)|X_0=x_0] \\ &\quad + \sum_{t=1}^{\infty}\left(E[\log m(X_t,X_{t+1})|X_1=x_1] - E[\log m(X_t,X_{t+1})|X_0=x_0]\right)\,.\end{aligned}$$

The efficient influence function for L is therefore, by linearity and continuity of log,

$$\tilde{\psi}_L(x_0,x_1) = \rho^{-1}\tilde{\psi}_\rho(x_0,x_1) - \tilde{\psi}_m(x_0,x_1)\,.$$

Note that

$$V_L = \rho^{-2}V_\rho - 2\rho^{-1}C_{\rho m} + V_m'$$

where

$$\begin{aligned}V_m &= \textstyle\sum_{t=-\infty}^{\infty} E[(\log m(X_0,X_1) - E[\log m(X_0,X_1)])(\log m(X_t,X_{t+1}) - E[\log m(X_0,X_1)])] \\ C_{\rho m} &= E[(\phi^*(X_0)m(X_0,X_1)\phi(X_1) - \rho\phi^*(X_0)\phi(X_0))\log m(X_0,X_1)]\,.\end{aligned}$$

[18] Greenwood and Wefelmeyer (1995) prove efficiency of sample averages for estimating the expectation of bounded measurable functions of (X_0,X_1). Wefelmeyer (1999) extends this to the class of functions of (X_0,X_1) with finite second moment.

The efficiency bound for L is then

$$\begin{aligned} E[\tilde{\psi}_L(X_0,X_1)^2] &= \rho^{-2}V_\rho + E[\tilde{\psi}_m(X_0,X_1)^2] - 2\rho^{-1}E[\tilde{\psi}_\rho(X_0,X_1)\tilde{\psi}_m(X_0,X_1)] \\ &= \rho^{-2}V_\rho + V_m - 2\rho^{-1}E[\tilde{\psi}_\rho(X_0,X_1)\tilde{\psi}_m(X_0,X_1)] \end{aligned}$$

since $E[\tilde{\psi}_m(X_0,X_1)^2] = V_m$ (Wefelmeyer, 1999). Using the fact that $E[\tilde{\psi}_\rho(X_0,X_1)|X_0] = 0$,

$$\begin{aligned} & E[\tilde{\psi}_\rho(X_0,X_1)\tilde{\psi}_m(X_0,X_1)] \\ = \; & E[\log m(X_0,X_1)\tilde{\psi}_\rho(X_0,X_1)] \\ & + \sum_{t=1}^{\infty} E[E[\log m(X_t,X_{t+1})|X_1]E[\tilde{\psi}_\rho(X_0,X_1)|X_1]] \\ = \; & E[\log m(X_0,X_1)\{\phi^*(X_0)m(X_0,X_1)\phi(X_1) - \rho\phi^*(X_0)\phi(X_0)\}] \\ & + \sum_{t=1}^{\infty} \rho E[E[\log m(X_t,X_{t+1})|X_1](\phi^*(X_1)\phi(X_1) - E[\phi^*(X_0)\phi(X_0)|X_1])] \\ = \; & E[\log m(X_0,X_1)\phi^*(X_0)m(X_0,X_1)\phi(X_1)] - \rho \lim_{t\to\infty} E[\log m(X_t,X_{t+1})\phi^*(X_0)\phi(X_0)] \\ = \; & C_{\rho m} \end{aligned}$$

where the second equality is by definition of ϕ^*, the fourth is by telescoping series, and the fifth is by Lemma 1.10.2. Therefore, $E[\tilde{\psi}_m(X_0,X_1)^2] = V_L$. Theorem 1.4.2 shows $\widehat{L}$ attains this bound. □

Proof of Theorem 1.4.5. Application of Theorem 2.3.3 with $\widetilde{\mathbf{G}}_K$ in place of $\widehat{\widehat{\mathbf{G}}}_K$ yields

$$\check{\rho} - \rho_K = (\mathbf{G}_K^{1/2}c_K^\star)'(\widetilde{\mathbf{G}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)(\mathbf{G}_K^{1/2}c_K) + O_p(\bar{\eta}_{n,K}^2)\,.$$

Using the fact that $\widetilde{\mathbf{G}}_K^{-1} = I_K$ and $\mathbf{G}_K^{-1}\mathbf{M}_K c_K = \rho_K c_K$ yields

$$\check{\rho} - \rho_K = \frac{1}{n}\sum_{t=0}^{n-1} \phi_K^\star(X_t)m(X_t,X_{t+1})\phi_K(X_{t+1}) - \rho_K + O_p(\bar{\eta}_{n,K}^2)\,.$$

Applying the same arguments as in the proof of Theorem 1.4.2 yields

$$\sqrt{n}(\check{\rho} - \rho) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \{\phi_t^* m_{t,t+1} \phi_{t+1} - \rho\} + o_p(1)\,.$$

The summands are strictly stationary phi-mixing random variables by Assumption 1.3.1 and Lemma 1.10.1, have mean zero, and have and finite second moment by Assumption 1.4.3. It follows by Corollary 2.2 of Peligrad (1985) that

$$\sqrt{n}(\check{\rho} - \rho) \to_d N(0, \mathrm{lrvar}(\phi_t^* m_{t,t+1} \phi_{t+1} - \rho))\,.$$

By definition,

$$\begin{aligned} \mathrm{lrvar}(\phi_t^* m_{t,t+1} \phi_{t+1} - \rho) \quad &= \quad E[(\phi_0^* m_{0,1} \phi_1 - \rho)^2] \\ &\quad\; +2 \textstyle\sum_{t=1}^{\infty} E[(\phi_0^* m_{0,1} \phi_1 - \rho)(\phi_t^* m_{t,t+1} \phi_{t+1} - \rho)] \end{aligned} \tag{1.57}$$

where

$$E[(\phi_0^* m_{0,1} \phi_1 - \rho)^2] = V_\rho + \rho^2 E[(\phi^*(X_0)\phi(X_0) - 1)^2]$$

and, for each $t \geq 1$,

$$\begin{aligned} E[(\phi_0^* m_{0,1} \phi_1 - \rho)(\phi_t^* m_{t,t+1} \phi_{t+1} - \rho)] \quad &= \quad \rho^2 E[(\phi_1^* \phi_1 - 1)(\phi_t^* \phi_t - 1)] \\ &= \quad \rho^2 E[(\phi_0^* \phi_0 - 1)(\phi_{t-1}^* \phi_{t-1} - 1)]\,. \end{aligned}$$

Substituting into (1.57) yields

$$\mathrm{lrvar}(\phi_t^* m_{t,t+1} \phi_{t+1} - \rho) = V_\rho + 2\rho^2 E[(\phi_0^* \phi_0 - 1)^2] + \rho^2 \mathrm{lrvar}((\phi_t^* \phi_t - 1))$$

as required. □

Proof of Theorem 1.5.1. Follows identical arguments to the proofs of Theorem 1.4.1. □

Proof of Theorem 1.5.2. Repeating the arguments in Lemma 1.10.4 shows

$$\widehat{\rho} - \rho_K = \frac{1}{n}\sum_{t=0}^{n-1}\{\phi_K^\star(X_t)\widehat{m}(X_t, X_{t+1})\phi_K(X_{t+1}) - \rho_K\phi_K^\star(X_t)\phi_K(X_t)\} + O_p(\bar{\eta}_{n,K}^2)\,.$$

Therefore,

$$\begin{aligned} & \widehat{\rho} - \rho \\ = \ & \rho_K - \rho \\ & + \frac{1}{n}\sum_{t=0}^{n-1}\{\phi_K^\star(X_t)m(X_t, X_{t+1})\phi_K(X_{t+1}) - \rho_K\phi_K^\star(X_t)\phi_K(X_t)\} + O_p(\bar{\eta}_{n,K}^2) \quad (1.58) \\ & + \frac{1}{n}\sum_{t=0}^{n-1}\phi_K^\star(X_t)(\widehat{m}(X_t, X_{t+1}) - m(X_t, X_{t+1}))\phi_K(X_{t+1})\,. \quad (1.59) \end{aligned}$$

Expression (1.58) is controlled as in the proof of Lemma 1.10.4. It follows from the uniform convergence rates established in Corollary 1.4.2 and Assumption 1.5.1 that

$$(1.59) = \frac{1}{n}\sum_{t=0}^{n-1}\phi^*(X_t)(\widehat{m}(X_t, X_{t+1}) - m(X_t, X_{t+1}))\phi(X_{t+1}) + o_p(1)\,.$$

The result follows. □

Proof of Theorem 1.5.3. The proof is analogous to the proof of Theorem 1.3.1. Note that $\mathbb{M} : L^2(Q) \to L^2(Q)$ can be represented as an integral operator with integral kernel $\mathcal{K}(x_0, x_1)$ given by

$$\begin{aligned} \mathcal{K}(x_0, x_1) &= \left\{\int_{\mathcal{Y}} m(x_0, x_1, y_1)f(x_1, y_1|x_0)\,\mathrm{d}y_1\right\}\frac{1}{q(x_1)} \\ &= \left\{\int_{\mathcal{Y}} m(x_0, x_1, y_1)\frac{f(x_0, x_1, y_1)}{q(x_0)q(x_1)q_y(y_1)}\,\mathrm{d}Q_y(y_1)\right\}\,. \end{aligned}$$

The positivity conditions in Assumption 1.5.2 imply that $\mathcal{K}(x_0, x_1) > 0$ a.e.-$[Q \otimes Q]$.

To check square-integrability of $\mathcal{K}$, observe that

$$\begin{aligned}
& \int_{\mathcal{X}}\int_{\mathcal{X}} \mathcal{K}^2(x_0,x_1)\,\mathrm{d}Q(x_0)\,\mathrm{d}Q(x_1) \\
= \;& \int_{\mathcal{X}}\int_{\mathcal{X}} \left\{\int_{\mathcal{Y}} m(x_0,x_1,y_1)\frac{f(x_0,x_1,y_1)}{q(x_0)q(x_1)q_y(y_1)}\,\mathrm{d}Q_y(y_1)\right\}^2 \mathrm{d}Q(x_0)\,\mathrm{d}Q(x_1) \\
\leq \;& \int_{\mathcal{X}}\int_{\mathcal{X}}\int_{\mathcal{Y}} m(x_0,x_1,y_1)^2\frac{f(x_0,x_1,y_1)^2}{q(x_0)q(x_1)q_y(y_1)}\,\mathrm{d}y_1\,\mathrm{d}x_0\,\mathrm{d}x_1 \\
\leq \;& C\int_{\mathcal{X}}\int_{\mathcal{X}}\int_{\mathcal{Y}} m(x_0,x_1,y_1)^2 f(x_0,x_1,y_1)\,\mathrm{d}y_1\,\mathrm{d}x_0\,\mathrm{d}x_1 \\
= \;& CE[m(X_0,X_1,Y_1)^2] \quad < \quad \infty
\end{aligned}$$

for some finite positive C, by virtue of boundedness of $f(x_0,x_1,y_1)/(q(x_0)q(x_1)q_y(y_1))$, and Assumption 1.5.2(iv). The result follows by Theorem 2.2.1. □

Proof of Theorem 1.5.4. Follows identical arguments to the proofs of Theorems 1.4.1 and 1.4.2, noting that in this case $\mathbb{M}$ may be rewritten as

$$\mathbb{M}\psi(x) = E[E[m(X_0,X_1,Y_1)|X_0,X_1]\psi(X_1)|X_0 = x]$$

where clearly $E[m(X_0,X_1,Y_1)|X_0,X_1]$ is a function of (X_0,X_1). □

1.10 Supplementary lemmas and their proofs

Weak-dependence properties of $\{X_t\}$ are first established under Assumption 1.3.1. Section 2.4 provides definitions of the relevant mixing conditions.

Lemma 1.10.1. *Under Assumption 1.3.1(i)–(iii), $\{X_t\}$ is geometrically phi-mixing.*

Proof of Lemma 1.10.1. Let $\mathbb{E}$ denote the conditional expectation operator associated with $\{X_t\}$, i.e. $\mathbb{E}\psi(x) = E[\psi(X_1)|X_0 = x]$. A sufficient condition for $\{X_t\}$ to be geometrically phi-mixing is

$$\sup_{\psi\in L^\infty(Q):\psi\neq 0, E[\psi(X)]=0} \frac{\|\mathbb{E}^\tau\psi\|_{L^\infty(Q)}}{\|\psi\|_{L^\infty(Q)}} \to 0$$

as $\tau \to \infty$ (Doukhan, 1994, pp. 88–89). The inequality

$$\sup_{\psi \in L^\infty(Q):\psi \neq 0, E[\psi(X)]=0} \frac{\|\mathbb{E}^\tau \psi\|_{L^\infty(Q)}}{\|\psi\|_{L^\infty(Q)}} \leq \sup_{\psi \in L^\infty(Q):\psi \neq 0} \frac{\|\mathbb{E}^\tau \psi - E[\psi(X)]\|_{L^\infty(Q)}}{\|\psi\|_{L^\infty(Q)}} \tag{1.60}$$

is immediate. It is therefore sufficient to establish that the right-hand side of (1.60) is $O(e^{-c\tau})$ for some $c > 0$.

Theorem 2.2.4 will be applied to $\mathbb{E}$ (by setting $m(x_0, x_1) \equiv 1$). It is enough to show that $\mathbb{E}$ is compact. First observe that $\mathbb{E} : L^\infty(Q) \to L^\infty(Q)$ may be continuously extended to have domain $L^1(Q)$. For $\psi \in L^1(Q)$,

$$\begin{aligned} \|\mathbb{E}\psi\|_{L^\infty(Q)} &\leq \sup_{x_0,x_1} \left| \frac{f(x_0,x_1)}{q(x_0)q(x_1)} \right| \int_{\mathcal{X}} |\psi(x_1)| \, \mathrm{d}Q(x_1) \\ &\leq C\|\psi\|_{L^1(Q)} \end{aligned}$$

for some finite positive constant C, under Assumption 1.3.1(i)–(iii). Therefore $\mathbb{E} : L^1(Q) \to L^\infty(Q)$ is continuous, and so $\mathbb{E} : L^\infty(Q) \to L^\infty(Q)$ is compact (Schaefer, 1974, p. 337).

Let $f(x) = 1$ for all $x \in \mathcal{X}$. The function $f \in L^\infty(Q)$ and is an eigenfunction of $\mathbb{E}$ with eigenvalue 1 because $\mathbb{E}f = f$. The functional $e^* : L^\infty(Q) \to \mathbb{R}$ defined by $e^*(\psi) = E[\psi(X)] = \int_{\mathcal{X}} \psi(x) \, \mathrm{d}Q(x)$ is clearly bounded and linear. Let $x^* : L^\infty(Q) \to \mathbb{R}$ be a bounded linear functional. The adjoint $\mathbb{E}^*$ of $\mathbb{E}$ is defined by

$$(\mathbb{E}^* x^*)(\psi) = x^*(\mathbb{E}\psi)$$

for all $\psi \in L^\infty(Q)$ and all x^* in the dual space of $L^\infty(Q)$. By iterated expectations

$$(\mathbb{E}^* e^*)(\psi) = e^*(\mathbb{E}\psi) = \int_{\mathcal{X}} \mathbb{E}\psi(x) \, \mathrm{d}Q(x) = E[\psi(X)] = e^*(\psi)$$

for all $\psi \in L^\infty(Q)$. Therefore e^* is an eigenfunction of $\mathbb{E}^*$ with eigenvalue 1. Define $P : L^\infty(Q) \to L^\infty(Q)$ by $P\psi(x) = E[\psi(X)]$ for all $x \in \mathcal{X}$. Theorem 2.2.4 applied to $\mathbb{E} : L^\infty(Q) \to L^\infty(Q)$ yields the desired result. $\square$

Remark 1.10.1. *It follows from Lemma 1.10.1 and equation (2.4) that* $\{X_t\}$ *is also geometrically alpha-, rho- and beta-mixing under Assumption 1.3.1(i)–(iii).*

Lemma 1.10.2. *Under Assumption 1.3.1, there exists a* $\delta > 0$ *such that for any* $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *and* $g : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *with* $E[f(X_0, X_1)] = E[g(X_0, X_1)] = 0$*,* $E[f(X_0, X_1)^2] < \infty$ *and* $E[g(X_0, X_1)^2] < \infty$*, and any* $t \geq 1$*,*

$$|E[f(X_0, X_1)g(X_t, X_{t+1})]| \leq e^{-\delta(t-1)} E[f(X_0, X_1)^2]^{1/2} E[g(X_0, X_1)^2]^{1/2} .$$

Proof of Lemma 1.10.2. $\{X_t\}$ is geometrically rho-mixing under Assumption 1.3.1 (see Remark 1.10.1), so there exists a $\delta > 0$ such that

$$|\mathrm{Cov}(f_1(X_t), g_1(X_{t+\tau}))| \leq \exp^{-\delta\tau} E[f_1(X_0)^2]^{1/2} E[g_1(X_0)^2]^{1/2}$$

for each $\tau \geq 1$ and each $f_1 : \mathcal{X} \to \mathbb{R}$ and $g_1 : \mathcal{X} \to \mathbb{R}$ with $E[f_1(X_0)^2] < \infty$ and $E[g_1(X_0)^2] < \infty$ by the covariance inequality for rho-mixing random variables (Doukhan, 1994, p. 9)

Let $\mathcal{G}_t = \sigma(X_t, X_{t+1}, \ldots)$. By the Markov property,

$$\begin{aligned} E[f(X_0, X_1)g(X_t, X_{t+1})] &= E[f(X_0, X_1)E[g(X_t, X_{t+1})|\mathcal{F}_t]] \\ &= E[g(X_0, X_1)E[g(X_t, X_{t+1})|X_t]] \\ &= E[E[f(X_0, X_1)|\mathcal{G}_1]E[g(X_t, X_{t+1})|X_t]] \\ &= E[E[f(X_0, X_1)|X_1]E[g(X_t, X_{t+1})|X_t]] \end{aligned}$$

Therefore, by the covariance inequality,

$$|E[f(X_0, X_1)g(X_t, X_{t+1})]| = e^{-\delta(t-1)} E[E[f(X_0, X_1)|X_1]^2]^{1/2} E[E[g(X_0, X_1)|X_0]^2]^{1/2}$$

and the result follows by Jensen's inequality. □

Several lemmas are first required before proving Theorem 1.4.2. Define the remainder term

$$\begin{aligned}
\tau_{n,K} &= \frac{1}{n}\sum_{t=0}^{n-1} \xi_{K,t} - \xi_t \qquad (1.61)\\
\xi_{K,t} &= \phi_K^\star(X_t)\phi_K(X_{t+1})m(X_t,X_{t+1}) - \rho_K\phi_K^\star(X_t)\phi_K(X_t)\\
\xi_t &= \phi^*(X_t)\phi(X_{t+1})m(X_t,X_{t+1}) - \rho\phi^*(X_t)\phi(X_t)\,.
\end{aligned}$$

Lemma 1.10.3. *Under Assumptions 1.3.1, 1.4.1, 1.4.2, and 1.4.3(i)(ii),* $\tau_{n,K} = O_p(\zeta_0(K)(\delta_K^* + \delta_K)/\sqrt{n})$.

Proof of Lemma 1.10.3. First write

$$\tau_{n,K} = \frac{1}{\sqrt{n}} S_{n,K}$$

where $S_{n,K} = \sqrt{n}\tau_{n,K}$. Note that the summands in $S_{n,K}$ have mean zero and finite second moment. By Lemma 1.10.2 and the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, there exists a finite positive constant C such that

$$\begin{aligned}
E[S_{n,K}^2] &\leq CE[(\xi_{K,0} - \xi_0)^2]\\
&\leq 2CE\left[(\phi_K^\star(X_0)\phi_K(X_1) - \phi^*(X_0)\phi(X_1))^2 m(X_0,X_1)^2\right]\\
&\quad +2CE[(\rho_K\phi_K^\star(X_0)\phi_K(X_0) - \rho\phi^*(X_0)\phi(X_0))^2]\\
&\leq 4C\big(E[\{(\phi_K^\star(X_0) - \phi^*(X_0))^2\phi_K(X_1)^2 + (\phi_K(X_1)-\phi(X_1))^2\phi^*(X_0)^2\}m(X_0,X_1)^2]\\
&\quad +2\rho_K^2 E[(\phi_K^\star(X_0) - \phi^*(X_0))^2\phi_K(X_1)^2] + 2\rho_K^2 E[\phi^*(X_0)^2(\phi_K(X_1)-\phi(X_1))^2] \qquad (1.62)\\
&\quad +2(\rho_K-\rho)^2 E[\phi^*(X_0)^2\phi(X_1)^2]\big)\,.
\end{aligned}$$

Assumptions 1.3.1 and 1.4.1 are sufficient to apply Theorem 2.3.1 to $\mathbb{M}$ (see the proof of Theorem 1.4.1). This yields $\rho_K - \rho = O(\delta_K)$, $\|\phi_K - \phi\| = O(\delta_K)$ and $\|\phi_K^* - \phi^*\| = O(\delta_K)$.

Under Assumption 1.4.2(ii),

$$\|\phi^\star_K - \phi^*_K\| = \|\Pi^b_K(\phi^*_K - g^*_K) + g^*_K - \phi^*_K\| \leq 2\|\phi^*_K - g^*_K\| \leq 2\|\phi^*_K - \phi^*\| + 2\|\phi^* - g^*_K\|$$

which is $O(\delta_K + \delta^*_K)$ by Assumption 1.4.2(ii) and Theorem 2.3.1. It follows by Assumption 1.4.2 using similar arguments to the proof of Corollary 1.4.2 that $\|\phi^\star_K - \phi^*\|_\infty = O(\zeta_0(K)(\delta^*_K + \delta_K))$ and $\|\phi_K - \phi\|_\infty = O(\zeta_0(K)\delta_K)$. Plugging these rates into (1.62) and using the Hölder inequality yields

$$E[S^2_{n,K}] = O(\zeta_0(K)^2(\delta^*_K + \delta_K)^2)\,.$$

The result follows by Chebychev's inequality. $\square$

Lemma 1.10.4. *Under Assumptions 1.3.1, 1.4.1, 1.4.2, and 1.4.3(i)(ii),*

$$\sqrt{n}(\widehat{\rho} - \rho) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \{\phi^*(X_t) m(X_t, X_{t+1}) \phi(X_{t+1}) - \rho \phi^*(X_t)\phi(X_t)\} + o_p(1)\,.$$

Proof of Lemma 1.10.4. Assumptions 1.3.1 and 1.4.1 are sufficient for the assumptions of Theorem 2.3.3. Application of Theorem 2.3.3 yields

$$\widehat{\rho} - \rho_K = (\mathbf{G}_K^{1/2} c^\star_K)'(\widehat{\widetilde{\mathbf{G}}}_K^{-1} \widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)(\mathbf{G}_K^{1/2} c_K) + O_p(\bar{\eta}^2_{n,K})\,. \tag{1.63}$$

First, using the fact that $\widehat{\widetilde{\mathbf{G}}}_K^{-1} = I_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)$ gives

$$\begin{aligned}
& \widehat{\widetilde{\mathbf{G}}}_K^{-1} \widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K \\
= \; & \widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)\widetilde{\mathbf{M}}_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K) \\
= \; & \widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K - (I_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K))(\widehat{\widetilde{\mathbf{G}}}_K - I_K)\widetilde{\mathbf{M}}_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K) \\
= \; & \widehat{\widetilde{\mathbf{M}}}_K - \widehat{\widetilde{\mathbf{G}}}_K \widetilde{\mathbf{M}}_K + \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)^2 \widetilde{\mathbf{M}}_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)\,.
\end{aligned}$$

The leading term in (1.63) is then

$$
\begin{aligned}
(\mathbf{G}_K^{1/2}c_K^\star)'(\widehat{\widetilde{\mathbf{M}}}_K - \widehat{\widetilde{\mathbf{G}}}_K\widetilde{\mathbf{M}}_K)(\mathbf{G}_K^{1/2}c_K) &= c_K^{\star\prime}\widehat{\mathbf{M}}_K c_K - c_K^{\star\prime}\widehat{\mathbf{G}}_K\mathbf{G}_K^{-1}\mathbf{M}_K c_K \\
&= c_K^{\star\prime}\widehat{\mathbf{M}}_K c_K - \rho_K c_K^{\star\prime}\widehat{\mathbf{G}}_K c_K
\end{aligned}
$$

where the second line is by equation (1.23). It remains to show that the remaining part of expression (1.63) is $O_p(\bar{\eta}_{n,K}^2)$. By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
& |(\mathbf{G}_K^{1/2}c_K^\star)'(\widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)^2\widetilde{\mathbf{M}}_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K))(\mathbf{G}_K^{1/2}c_K)| \\
\leq\ & \|\phi_K^\star\|\|\phi_K\|\left\|\widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)^2\widetilde{\mathbf{M}}_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1}(\widehat{\widetilde{\mathbf{G}}}_K - I_K)(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)\right\|_2 \\
=\ & (\|\phi^*\| + o(1))(\|\phi\| + o(1)) \times O_p(\bar{\eta}_{n,K}^2)
\end{aligned}
$$

where the final line is because $\|\phi_K - \phi\| = o(1)$ and $\|\phi_\star^K - \phi_K^*\| = o(1)$ (see the proof of Lemma 1.10.3), and the $O_p(\bar{\eta}_{n,K}^2)$ term follows by the same arguments as the proof of Lemma 2.5.4. The expansion (1.63) may therefore be reexpressed as

$$
\widehat{\rho} - \rho_K = \frac{1}{n}\sum_{t=0}^{n-1}\{\phi_K^\star(X_t)m(X_t, X_{t+1})\phi_K(X_{t+1}) - \rho_K\phi_K^\star(X_t)\phi_K(X_t)\} + O_p(\bar{\eta}_{n,K}^2)\,.
$$

Rearranging yields

$$
\widehat{\rho} - \rho = \rho_K - \rho + \frac{1}{n}\sum_{t=0}^{n-1}\{\phi^*(X_t)m(X_t, X_{t+1})\phi(X_{t+1}) - \rho\phi^*(X_t)\phi(X_t)\} + \tau_{n,K} + O_p(\bar{\eta}_{n,K}^2)\,.
$$

where $\tau_{n,K}$ is defined in expression (1.61). Lemma 1.10.3 and Assumption 1.4.3(ii) together imply that $\sqrt{n}(\tau_{n,K} + O_p(\bar{\eta}_{n,K}^2)) = o_p(1)$. Finally, $|\rho_K - \rho| = O(\delta_K) = o(n^{-1/2})$ by the proof of Lemma 1.10.3 and the condition $\delta_K = o(n^{-1/2})$. □

Chapter 2

Further results on identification and estimation

This chapter is an addendum to Chapter 1, extending the identification, estimation, and long-run pricing results presented therein. In Section 2.2 the nonparametric identification conditions presented in Chapter 1 are both weakened and extended to more general function spaces. Versions of the long-term pricing result of Hansen and Scheinkman (2009) are also shown to obtain in more general function spaces under the weaker identification conditions. High-level conditions for consistency and convergence rates for nonparametric sieve estimators of the positive eigenfunctions of a collection of nonselfadjoint operators and their adjoints are derived in Section 2.3. Convergence of rates for the matrix estimators are derived in Section 2.4 using results from random matrix theory, which are useful in the verification of these high-level conditions. In addition, Section 2.1 briefly reviews relevant concepts and definitions related to spectral theory (Section 2.1.1) and Banach lattices (Section 2.1.2). All proofs are contained in Section 2.5.

Notation: Let $(\mathcal{X}, \mathscr{X}, \mu)$ be a measure space. The space $L^p(\mu) := L^p(\mathcal{X}, \mathscr{X}, \mu)$ with $1 \leq p < \infty$ consists of all (equivalence classes of) measurable functions $f : \mathcal{X} \to \mathbb{R}$ such that $\int |f|^p \, \mathrm{d}\mu < \infty$. The space $L^\infty(\mu) := L^\infty(\mathcal{X}, \mathscr{X}, \mu)$ consists of all (equivalence classes of) measurable functions $f : \mathcal{X} \to \mathbb{R}$ such that $\operatorname{ess\,sup}|f| < \infty$. Let $\|\cdot\|_{L^p(\mu)}$ denote the $L^p(\mu)$ norm when applied to functions and the operator norm when applied to linear operators on the space $L^p(\mu)$. Let a.e.-$[\mu]$ denote almost everywhere with respect to the measure μ and a.e.-$[\mu \otimes \mu]$ denote almost everywhere with respect to the product measure $\mu \otimes \mu$. Let $\Gamma(\delta, \lambda)$ denote the positively oriented circle (in the complex plane) centered at λ with radius δ. Let $B(\delta, \lambda)$ denote the open ball (in the complex plane) centered at λ with radius δ. Finally, let $d(z, A) = \inf_{\zeta \in A} |z - \zeta|$ for $z \in \mathbb{C}$ and $A \subset \mathbb{C}$.

2.1 Brief review of spectral theory and Banach lattices

2.1.1 Spectral theory

Definitions of relevant concepts from the spectral theory of operators on a Banach space are now briefly reviewed. Let E be a Banach space and $T : E \to E$ be a bounded linear operator.[1] The following definitions are as in Kato (1980) unless stated otherwise.

The resolvent set $\operatorname{res}(T) \subseteq \mathbb{C}$ of T is the set of all $z \in \mathbb{C}$ for which the resolvent operator $R(T, z) := (T - zI)^{-1}$ exists as a bounded linear operator on E (where $I : E \to E$ is the identity operator given by $Ix = x$ for all $x \in E$). The spectrum $\sigma(T)$ is defined as the complement in $\mathbb{C}$ of $\operatorname{res}(T)$, i.e. $\sigma(T) := (\mathbb{C} \setminus \operatorname{res}(T))$. If $S : E \to E$ is another bounded linear operator and $z \in \operatorname{res}(T) \cup \operatorname{res}(S)$ then the so-called second resolvent equation obtains:

$$R(S, z) - R(T, z) = R(S, z)(T - S)R(T, z)\,. \tag{2.1}$$

The spectral radius $\operatorname{spr}(T)$ of T is $\operatorname{spr}(T) = \sup\{|\lambda| : \lambda \in \sigma(T)\}$. The Gelfand formula shows that $\operatorname{spr}(T) = \lim_{n\to\infty} \|T^n\|^{1/n}$. The point spectrum $\pi(T) \subseteq \sigma(T)$ of T is the set of all $z \in \mathbb{C}$ for which the null space of $(T - zI)$ is nonzero. When $\pi(T)$ is nonempty, each $\lambda \in \pi(T)$ is an eigenvalue of T and any nonzero ψ in the null space of $(T - \lambda I)$ is an

[1]The definitions are presented in the case that E is a Banach space over $\mathbb{C}$. When E is a Banach space over $\mathbb{R}$, the complex extension $T(x + iy) = T(x) + iT(y)$ for $x, y \in E$ of T is defined on the complexification $E + iE$ of E. The spectrum and associated quantities of T when E is a Banach space over $\mathbb{R}$ are obtained by applying the complex Banach space definitions to the complex extension of T.

eigenvector of T corresponding to λ. The dimension of the null space of $(T - \lambda I)$ is the geometric multiplicity of the eigenvalue λ.

An eigenvalue $\lambda \in \pi(T)$ is said to be isolated if $\inf_{z\in\sigma(T):z\neq\lambda} |z-\lambda| \geq 2\epsilon$ for some $\epsilon > 0$, in which case the spectral projection of T corresponding to λ can be written as

$$P = \frac{-1}{2\pi i} \int_{\Gamma(\epsilon,\lambda)} R(T, z)\, \mathrm{d}z\,. \tag{2.2}$$

The dimension of PE is the algebraic multiplicity of λ. The algebraic multiplicity is the order of the pole of $R(T, z)$ at $z = \lambda$ and is at least as large as the geometric multiplicity of λ (Chatelin, 1983). The term multiplicity is used for eigenvalues whose algebraic and geometric multiplicity are equal. Suppose that λ is an isolated real eigenvalue of T. Then λ is an isolated real eigenvalue of the adjoint T^* of T, and the algebraic and geometric multiplicities of λ are the same for T and T^*. If, in addition, λ has multiplicity one, then $P = x\otimes x^*$ with where $(x\otimes x^*)\psi = x^*(\psi)x$ and where x and x^* are eigenvectors of T and T^* corresponding to λ, and $x^*(x) = 1$ (Chatelin, 1983, p. 113). If, in addition, E is a Hilbert space then $P = (x \otimes x^*)$ is given by $(x \otimes x^*)\psi = \langle\psi, x^*\rangle x$ where x^* is an eigenvector of T^* corresponding to λ, $\|x\| = 1$ and $\langle x, x^*\rangle = 1$. In this case $\|P\| = \|x^*\| \geq 1$, with $x^* = x$ and $\|P\| = 1$ if P is an orthogonal projection (which is the case when T is selfadjoint).

2.1.2 Banach lattices and positive operators

Definitions of relevant concepts from the theory of Banach lattices, which are special types of Banach spaces for which there is a well developed theory of positive operators, are now briefly reviewed. The following definitions are as in Schaefer (1999).

A vector space E over the real field $\mathbb{R}$ endowed with an order relation $\leq$ is an ordered vector space if $f \leq g$ implies $f + h \leq g + h$ for all $f, g, h \in E$ and $f \leq g$ implies $\lambda f \leq \lambda g$ for all $f, g \in E$ and $\lambda \geq 0$. If, in addition, $\sup\{f, g\} \in E$ and $\inf\{f, g\} \in E$ for all $f, g \in E$ then E is a vector lattice. If there exists a norm $\|\cdot\|$ on an ordered vector lattice E under which E is complete and $\|\cdot\|$ satisfies the lattice norm property, namely $f, g \in E$ and

$|f| \leq |g|$ implies $\|f\| \leq \|g\|$, then E is a Banach lattice.

Let E_+ denote the positive cone of E defined with respect to the order relation on E. If $E = L^p(\mu)$ with $1 \leq p \leq \infty$ then $E_+ = \{f \in E : f \geq 0 \text{ a.e.-}[\mu]\}$. If $E = C_b(\mathcal{X})$ or $C(\mathcal{X})$ then $E_+ = \{f \in E : f(x) \geq 0 \text{ for all } x \in \mathcal{X}\}$. An element $f \in E_+$ belongs to the quasi-interior E_{++} of E_+ if $\{g \in E : 0 \leq g \leq f\}$ is a total subset of E. For example, if $E = L^p(\mu)$ with $1 \leq p < \infty$ and $f \in E$ is such that $f > 0$ a.e.-$[\mu]$ then $f \in E_{++}$. If $E = L^\infty(\mu)$ and $f \in E$ is such that $\operatorname{ess\,inf} f > 0$ then $f \in E_{++}$. If $E = C_b(\mathcal{X})$ or $C(\mathcal{X})$ and $f \in E$ is such that $\inf_{x \in \mathcal{X}} f(x) > 0$ then $f \in E_{++}$.

Let E^* denote the dual space of E (the set of all bounded linear functionals on E). For $f \in E$, $f^* \in E^*$, define the evaluation $\langle f, f^* \rangle := f^*(f)$. The dual cone $E^*_+ := \{f^* \in E^* : \langle f, f^* \rangle \geq 0 \text{ whenever } f \in E_+\}$ is the set of positive linear functionals on E. An element $f^* \in E^*_+$ is strictly positive if $f \in E_+$ and $f \neq 0$ implies $\langle f, f^* \rangle > 0$. The set of all strictly positive elements of E^*_+ is denoted E^*_{++}.

A bounded linear operator $T : E \to E$ is said to be positive if $T : E_+ \to E_+$ and irreducible if $TR(T, z) : (E_+ \setminus \{0\}) \to E_{++}$ for each $z \in (\operatorname{spr}(T), \infty)$.

2.2 Further results on identification and long-term pricing

This section provides primitive sufficient conditions for the identification of the positive eigenfunction and adjoint eigenfunction in stationary discrete-time environments for which $\mathbb{M}$ may be represented as an integral operator with positive kernel. Existence is achieved by application of the Perron-Frobenius theorem for positive integral operators. Identification is then established by a type of Kreĭn-Rutman theorem. A restatement of the long-term pricing result of Hansen and Scheinkman (2009) is shown to obtain under these conditions.

Basic regularity conditions are first introduced. Let $\mathscr{X}$ be the Borel σ-algebra on $\mathcal{X}$. The following conditions are sufficient for existence and nonparametric identification of ϕ.

Assumption 2.2.1. $\{X_t\}$ *is a strictly stationary and ergodic (first-order) Markov process supported on a Borel set* $\mathcal{X} \subseteq \mathbb{R}^d$*, and its stationary distribution* Q *has density* q *with* *respect to Lebesgue measure, with $q(x) > 0$ for almost every $x \in \mathcal{X}$.*

Assumption 2.2.2. $\mathbb{M} : L^p(Q) \to L^p(Q)$ *is bounded and* $\mathbb{M}_\tau$ *is compact for some* $\tau \geq 1$.

Assumption 2.2.3. $\mathbb{M}$ *may be written as in (1.18) with integral kernel* $\mathcal{K}$ *as in (1.19) such that:*

(i) $\mathcal{K}(x, y) \geq 0$ *a.e.-*$[Q \otimes Q]$

(ii) $\int_{A^c} \int_A m(x, y) f(x, y) \,\mathrm{d}x \,\mathrm{d}y > 0$ *for every* $A \in \mathscr{X}$ *with* $0 < Q(A) < 1$.

Assumption 2.2.1 is the same as Assumption 1.3.1(i). Assumptions 2.2.2 and 2.2.3 are higher-level conditions in place of Assumptions 1.3.1(ii)–(iv). Assumption 2.2.2 only requires some power of $\mathbb{M}$ to be compact and is weaker than requiring $\mathbb{M}$ itself to be compact.[2] Assumption 2.2.3 is trivially satisfied if $\mathcal{K}(x, y) > 0$ a.e.-$[Q \otimes Q]$. Assumption 1.3.1 is sufficient for Assumptions 2.2.1, 2.2.2, and 2.2.3 for $p = 2$.

Let $\mathrm{spr}(\mathbb{M})$ denote the spectral radius of $\mathbb{M}$ (see Section 2.1.1). Existence of ϕ follows by Theorem V.6.6 of Schaefer (1974).

Theorem 2.2.1 (Schaefer (1974))**.** *If Assumptions 2.2.1, 2.2.2 and 2.2.3 hold for some* $p \in [1, \infty]$, *then there exists a* $\phi \in L^p(Q)$ *with* $\phi(x) > 0$ *a.e.-*$[Q]$ *such that* $\mathbb{M}\phi = \rho\phi$ *with* $\rho = \mathrm{spr}(\mathbb{M}) > 0$, *and* ϕ *is the unique (to scale) eigenfunction of* $\mathbb{M}$ *corresponding to the eigenvalue* ρ. *If, in addition,* $\mathcal{K} > 0$ *a.e.-*$[Q \otimes Q]$ *then any other eigenvalue* λ *of* $\mathbb{M}$ *has modulus* $|\lambda| < \rho$.

If $1 \leq p < \infty$ let the dual index p' for $L^p(Q)$ be defined as $p^{-1} + p'^{-1} = 1$ with $p' = \infty$ if $p = 1$. The dual space of $L^p(Q)$ can be identified with the space $L^{p'}(Q)$ under the evaluation $E[\psi(X)\psi^*(X)]$ for $\psi \in L^p(Q)$, $\psi^* \in L^{p'}(Q)$. The adjoint operator $\mathbb{M}^* : L^{p'}(Q) \to L^{p'}(Q)$ is defined such that

$$E[\psi(X)\mathbb{M}^*\psi^*(X)] = E[\psi^*(X)\mathbb{M}\psi(X)]$$

[2]There are several sufficient conditions for this compactness condition. For $1 < p < \infty$ this is satisfied if there is a $\tau \geq 1$ such that $\mathbb{M}_\tau$ maps $L^p(Q)$ into $L^\infty(Q)$, for $p = 1$ if there is a $\tau \geq 1$ such that $\mathbb{M}_\tau$ maps $L^1(Q)$ into $L^r(Q)$ for some $r > 1$, and for $p = \infty$ if there is a $\tau \geq 1$ such that $\mathbb{M}_\tau$ has a continuous extension that maps $L^r(Q)$ into $L^\infty(Q)$ for some $r < \infty$ (Schaefer, 1974, p. 337).

for $\psi \in L^p(Q)$ and $\psi^* \in L^{p'}(Q)$. Let $(\mathcal{X}, \mathscr{X}_1, Q_1)$ denote the completion of $(\mathcal{X}, \mathscr{X}, Q)$ as described on p. 296 of Dunford and Schwartz (1958). The dual space of $L^\infty(Q)$ can be identified with the space $\mathrm{ba}(\mathcal{X}, \mathscr{X}_1, Q_1)$ of signed measures on $(\mathcal{X}, \mathscr{X}_1)$ which are absolutely continuous with respect to Q_1, under the evaluation $\psi^*(\psi) = \int_{\mathcal{X}} \psi \,\mathrm{d}\nu_{\psi^*}$ for $\psi \in L^\infty(Q)$ and $\nu_{\psi^*} \in \mathrm{ba}(\mathcal{X}, \mathscr{X}_1, Q_1)$ (Dunford and Schwartz, 1958, p. 296).

It follows from Theorem 2.2.1 by a version of the Kreĭn-Rutman theorem due to Schaefer (1960) that ϕ is the unique non-negative eigenfunction of $\mathbb{M}$.

Theorem 2.2.2. *If Assumptions 2.2.1, 2.2.2 and 2.2.3 hold for some $p \in [1, \infty]$, then ρ is an eigenvalue of $\mathbb{M}$ of multiplicity one, ϕ is the unique (to scale) eigenfunction of $\mathbb{M}$ with $\phi(x) \geq 0$ a.e.-$[Q]$ and:*

(i) If $p \in [1, \infty)$ there exists a $\phi^* \in L^{p'}(Q)$ with $\phi^*(x) > 0$ a.e.-$[Q]$ such that $\mathbb{M}^*\phi^* = \rho\phi^*$, and ϕ^* is the unique (to scale) eigenfunction of $\mathbb{M}^*$ with $\phi^*(x) \geq 0$ a.e.-$[Q]$

(ii) If $p = \infty$ there exists a unique (to scale) nonzero $\Phi_1^* \in \mathrm{ba}(\mathcal{X}, \mathscr{X}_1, Q_1)$ such that $\Phi_1^*(A) \geq 0$ for all $A \in \mathscr{X}$ with $Q(A) > 0$ and

$$\int_{\mathcal{X}} \mathbb{M}\psi(x) \,\mathrm{d}\Phi_1^*(x) = \rho \int \psi(x) \,\mathrm{d}\Phi_1^*(x)$$

for all $\psi \in L^\infty(Q)$.

Versions of the long-run pricing result

$$\lim_{\tau \to \infty} \rho^{-\tau} \mathbb{M}_\tau \psi(X_t) = \widetilde{E}[\psi(X)/\phi(X)]\phi(X_t)$$

of Hansen and Scheinkman (2009) hold under the identification conditions just presented. First consider the case $1 \leq p < \infty$. The positive eigenfunction and adjoint positive eigenfunction exist under Assumptions 2.2.1, 2.2.2 and 2.2.3. Impose the normalizations $E[\phi(X)^p] = 1$ and $E[\phi(X)\phi^*(X)] = 1$, and let $P : L^p(Q) \to L^p(Q)$ be defined as

$$P\psi(x) = E[\psi(X)\phi^*(X)]\phi(x)\,.$$

Theorem 2.2.3. *If Assumptions 2.2.1, 2.2.2 and 2.2.3 hold for some $p \in [1, \infty)$ with $\mathcal{K}(x, y) > 0$ a.e.-$[Q \otimes Q]$, then there exists $c > 0$ such that $\|\rho^{-\tau}\mathbb{M}^{\tau} - P\|_{L^p(Q)} = O(e^{-c\tau})$ as $\tau \to \infty$.*

Now consider the space $L^{\infty}(Q)$. Under Assumptions 2.2.1, 2.2.2 and 2.2.3 the positive eigenfunction exists, together with a nonzero measure $\Phi_1^* \in \mathrm{ba}(\mathcal{X}, \mathscr{X}_1, Q_1)$ such that $\Phi_1^*(A) \geq 0$ for all $A \in \mathscr{X}$ with $Q(A) > 0$. Normalize ϕ and Φ_1^* so that $\Phi_1^*(\mathcal{X}) = 1$ (making Φ_1^* a probability measure) and $\int_{\mathcal{X}} \phi \, \mathrm{d}\Phi_1^* = 1$. Let $P : L^{\infty}(Q) \to L^{\infty}(Q)$ be defined as

$$P\psi(x) = \left(\int_{\mathcal{X}} \psi \, \mathrm{d}\Phi_1^* \right) \phi(x) \,.$$

Theorem 2.2.4. *If Assumptions 2.2.1, 2.2.2 and 2.2.3 hold for $p = \infty$ with $\mathcal{K}(x, y) > 0$ a.e.-$[Q \otimes Q]$, then there exists $c > 0$ such that $\|\rho^{-\tau}\mathbb{M}^{\tau} - P\|_{L^{\infty}(Q)} = O(e^{-c\tau})$.*

Both Theorem 2.2.3 and Theorem 2.2.4 show that versions of the long-run pricing result hold in operator norm with the approximation error vanishing exponentially with τ. The theorems also show how to calculate the twisted probability measure $\widetilde{Q}$ used to calculate the unconditional expectation $\widetilde{E}$ when $1 \leq p < \infty$. Specifically, the Radon-Nikodym derivative of the twisted measure $\widetilde{Q}$ with respect to Q is

$$\frac{\mathrm{d}\widetilde{Q}(x)}{\mathrm{d}Q(x)} = \phi(x)\phi^*(x) \,. \tag{2.3}$$

Therefore, the twisted expectation $\widetilde{E}$ may be recovered by solving the eigenfunction problems $\mathbb{M}\phi = \rho\phi$ and $\mathbb{M}^*\phi^* = \rho\phi^*$.

2.3 Estimation of the positive eigenfunctions of a collection of operators

Let $\{X_t\}$ be a strictly stationary (not necessarily Markov) process with stationary distribution Q and whose support is a Borel set $\mathcal{X} \subseteq \mathbb{R}^d$. Consider a set of operators $\{\mathbb{M}_{\alpha} : \alpha \in \mathcal{A}\}$

indexed by an arbitrary parameter $\alpha \in \mathcal{A}$, where each $\mathbb{M}_\alpha : L^2(Q) \to L^2(Q)$ is given by

$$\mathbb{M}_\alpha \psi(x) = E[m(X_t, X_{t+1}; \alpha)\psi(X_{t+1})|X_t = x]$$

for some $m(\cdot, \cdot; \alpha) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. This setup trivially nests the case dealt with in Chapter 1 when $\mathcal{A}$ is a singleton. Suppose each $\mathbb{M}_\alpha$ has a isolated eigenvalue $\rho_\alpha = \mathrm{spr}(\mathbb{M}_\alpha)$ and unique positive eigenfunction ϕ_α corresponding to ρ_α (so each $\mathbb{M}_\alpha^*$ has a unique adjoint eigenfunction ϕ_α^*). Uniform (in α) convergence rates of nonparametric sieve estimators of $(\rho_\alpha, \phi_\alpha, \phi_\alpha^*)$ are now established.

The following analysis is conducted in $L^2(Q)$ as in Chapter 1. Let B_K be the sieve space spanned by the basis functions $\{b_{K1}, \ldots, b_{KK}\}$ and let Π_K^b be the orthogonal projection onto B_K. Under regularity conditions, for K sufficiently large the largest eigenvalue $\rho_{\alpha,K}$ of $\Pi_K^b \mathbb{M}_\alpha$ will be real and positive and have multiplicity one for all $\alpha \in \mathcal{A}$. Let $\phi_{\alpha,K} \in B_K$ be an eigenfunction of $\Pi_K^b \mathbb{M}_\alpha$ corresponding to $\rho_{\alpha,K}$. Similarly, the adjoint in $L^2(Q)$ of $\Pi_K^b \mathbb{M}_\alpha$ will have a unique eigenfunction $\phi_{\alpha,K}^*$ corresponding to $\rho_{\alpha,K}$, and the adjoint in B_K of $\Pi_K^b \mathbb{M}_\alpha|_{B_K} : B_K \to B_K$ will have an eigenfunction $\phi_{\alpha,K}^\star \in B_K$ corresponding to $\rho_{\alpha,K}$. As all quantities are defined up to sign and scale, impose the sign normalizations $E[\phi_{\alpha,K}(X)\phi_\alpha(X)] \geq 0$, $E[\phi_{\alpha,K}^*(X)\phi_\alpha^*(X)] \geq 0$ and $E[\phi_{\alpha,K}^\star(X)\phi_\alpha^*(X)] \geq 0$ and the scale normalizations $E[\phi_{\alpha,K}(X)^2] = 1$, $E[\phi_{\alpha,K}(X)\phi_{\alpha,K}^*(X)] = 1$ and $E[\phi_{\alpha,K}(X)\phi_{\alpha,K}^\star(X)] = 1$.

Let the Gram matrix $\mathbf{G}_K$ and its estimator $\widehat{\mathbf{G}}_K$ be as in Chapter 1. For each $\alpha \in \mathcal{A}$ let $\mathbf{M}_{\alpha,K}$ be as defined as in (1.24) with $m(\cdot, \cdot; \alpha)$ in place of $m(\cdot, \cdot)$ and let $\widehat{\mathbf{M}}_{\alpha,K}$ be a $K \times K$ matrix estimator of $\mathbf{M}_{\alpha,K}$ (i.e. a measurable function of the sample data). Under regularity conditions, with probability approaching one $\widehat{\mathbf{G}}_K$ is invertible and for each $\alpha \in \mathcal{A}$ the eigenvector problems

$$\begin{aligned} \widehat{\mathbf{G}}_K^{-1} \widehat{\mathbf{M}}_{\alpha,K} \widehat{c}_\alpha &= \widehat{\rho}_\alpha \widehat{c}_\alpha \\ \widehat{\mathbf{G}}_K^{-1} \widehat{\mathbf{M}}'_{\alpha,K} \widehat{c}_\alpha^\star &= \widehat{\rho}_\alpha \widehat{c}_\alpha^\star \end{aligned}$$

are solvable, where $\widehat{\rho}_\alpha = \lambda_{\max}(\widehat{\mathbf{G}}_K^{-1}\widehat{\mathbf{M}}_{\alpha,K})$ is real and positive. Then for each $\alpha \in \mathcal{A}$, $\widehat{\rho}_\alpha$ is the estimator of ρ_α, $\widehat{\phi}_\alpha = b^K(x)'\widehat{c}_\alpha$ is the estimator of ϕ_α, and $\widehat{\phi}^\star_\alpha = b^K(x)'\widehat{c}^\star_\alpha$ is the estimator of ϕ^*_α. As these eigenfunction estimators are only defined up to scale, impose the sign normalizations $E[\widehat{\phi}_\alpha(X)\phi_{\alpha,K}(X)] \geq 0$ and $E[\widehat{\phi}^\star_\alpha(X)\phi^\star_{\alpha,K}(X)] \geq 0$ and the scale normalizations $\|\widehat{\phi}_\alpha\| = 1$ and $E[\widehat{\phi}_\alpha(X)\widehat{\phi}^\star_\alpha(X)] = 1$.

Some definitions are required before introducing the regularity conditions. As in Section 1.4.2, let $\widetilde{b}^K$ denote the vector of orthonormalized basis functions and let

$$\widehat{\widetilde{\mathbf{G}}}_K = \frac{1}{n}\sum_{t=0}^{n-1}\widetilde{b}^K(X_t)\widetilde{b}^K(X_t)'\,.$$

For each $\alpha \in \mathcal{A}$ define

$$\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} = \mathbf{G}_K^{-1/2}\widehat{\mathbf{M}}_{\alpha,K}\mathbf{G}_K^{-1/2}\,.$$

The orthonormalized estimators $\widehat{\widetilde{\mathbf{G}}}_K$ and $\widehat{\widetilde{\mathbf{M}}}_{\alpha,K}$ are infeasible and do not actually need to be constructed, but it makes the asymptotic arguments easier to work with them in place of $\widehat{\mathbf{G}}_K$ and $\widehat{\mathbf{M}}_{\alpha,K}$. Note that any $\psi \in B_K$ can be written as $\psi_K(x) = \widetilde{c}_K(\psi)'\widetilde{b}^K(x)$ for some $\widetilde{c}_K(\psi) \in \mathbb{R}^K$. The space B_K is therefore isomorphic to $\mathbb{R}^K$ endowed with the Euclidean inner (dot) product, since

$$E[\psi_1(X)\psi_2(X)] = \widetilde{c}_K(\psi_1)'E[\widetilde{b}^K(x)\widetilde{b}^K(x)']\widetilde{c}_K(\psi_2) = \widetilde{c}_K(\psi_1)'\widetilde{c}_K(\psi_2)\,.$$

Therefore the matrix spectral norm $\|\cdot\|_2$ when applied to the orthonormalized matrices in $\mathbb{R}^{K\times K}$ is isomorphic to the operator norm for linear operators on B_K.

Let $\widetilde{c}_{\alpha,K}, \widetilde{c}^\star_{\alpha,K} \in \mathbb{R}^K$ be such that $\widetilde{b}^K(x)'\widetilde{c}_{\alpha,K} = \phi_{\alpha,K}$ and $\widetilde{b}^K(x)'\widetilde{c}^\star_{\alpha,K} = \phi^\star_{\alpha,K}$. Let $\{\delta_K, \delta^*_K, \bar{\eta}_{1,n,K}, \bar{\eta}_{2,n,K}, \eta_{1,n,K}, \eta_{2,n,K} : n, K \geq 1\}$ be sequences of positive real values to be defined in the following assumptions. In the event of measurability issues, outer probabilities are used below implicitly in place of probabilities.

Assumption 2.3.1. *The set of operators* $\{\mathbb{M}_\alpha : \alpha \in \mathcal{A}\}$ *satisfies:*

(i) for each $\alpha \in \mathcal{A}$*,* $\mathbb{M}_\alpha : L^2(Q) \to L^2(Q)$ *is a bounded linear operator and* $\rho_\alpha = \mathrm{spr}(\mathbb{M}_\alpha)$ *is an isolated eigenvalue of* $\mathbb{M}_\alpha$ *of multiplicity one*

(ii) $\sup_{\alpha\in\mathcal{A}} \|\mathbb{M}_\alpha\| < \infty$ *and* $\bar{\epsilon} := \inf_{\alpha\in\mathcal{A}} d(\rho_\alpha, \sigma(\mathbb{M}_\alpha) \setminus \{\rho_\alpha\}) > 0$*.*

Assumption 2.3.2. *The sieve approximation error satisfies:*

(i) $\sup_{\alpha\in\mathcal{A}} \|\Pi_K^b \mathbb{M}_\alpha - \mathbb{M}_\alpha\| = O(\bar{\delta}_K)$ *where* $\bar{\delta}_K = o(1)$ *as* $K \to \infty$

(ii) $\sup_{\alpha\in\mathcal{A}} \|(\Pi_K^b \mathbb{M}_\alpha - \mathbb{M}_\alpha)\phi_\alpha\| = O(\delta_K)$*,* $\sup_{\alpha\in\mathcal{A}} \|(\mathbb{M}_\alpha^* \Pi_K^b - \mathbb{M}_\alpha^*)\phi_\alpha^* / \|\phi_\alpha^*\|\| = O(\delta_K^*)$*.*

Assumption 2.3.3. *There exists a continuous decreasing function* $r : (0,\infty) \to (0,\infty)$ *such that for each* $\alpha \in \mathcal{A}$*:*

(i) $\|R(\mathbb{M}_\alpha, z)\| \leq r(d(z, \sigma(\mathbb{M}_\alpha)))$ *for all* $z \in (B(\bar{\epsilon}, \rho_\alpha) \setminus \sigma(\mathbb{M}_\alpha))$

(ii) $\|R(\Pi_K^b \mathbb{M}_\alpha|_{B_K}, z)\| \leq r(d(z, \sigma(\Pi_K^b \mathbb{M}_\alpha|_{B_K})))$ *for all* $z \in (B(\bar{\epsilon}, \rho_\alpha) \setminus \sigma(\Pi_K^b \mathbb{M}_\alpha|_{B_K}))$*.*

Assumption 2.3.4. *The matrix estimators and their population counterparts are such that:*

(i) $\lambda_{\min}(\mathbf{G}_K) > 0$ *for every* $K \geq 1$

(ii)

$$\begin{aligned} \|\widehat{\widetilde{\mathbf{G}}}_K - I_K\|_2 &= O_p(\bar{\eta}_{1,n,K}) \\ \sup_{\alpha\in\mathcal{A}} \|\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K}\|_2 &= O_p(\bar{\eta}_{2,n,K}) \end{aligned}$$

where $\bar{\eta}_{n,K} = \max\{\bar{\eta}_{1,n,K}, \bar{\eta}_{2,n,K}\} = o(1)$ *as* $n, K \to \infty$*.*

(iii)

$$\begin{aligned}
\sup_{\alpha\in\mathcal{A}} \|(\widehat{\widetilde{\mathbf{G}}}_K - \widetilde{\mathbf{G}}_K)\widetilde{c}_{\alpha,K}\|_2 &= O_p(\eta_{1,n,K}) \\
\sup_{\alpha\in\mathcal{A}} \|(\widehat{\widetilde{\mathbf{G}}}_K - \widetilde{\mathbf{G}}_K)\widetilde{c}^{\star}_{\alpha,K}/\|\widetilde{c}^{\star}_{\alpha,K}\|_2\|\|_2 &= O_p(\eta_{1,n,K}) \\
\sup_{\alpha\in\mathcal{A}} \|(\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K})\widetilde{c}_{\alpha,K}\|_2 &= O_p(\eta_{2,n,K}) \\
\sup_{\alpha\in\mathcal{A}} \|(\widehat{\widetilde{\mathbf{M}}}'_{\alpha,K} - \widetilde{\mathbf{M}}'_{\alpha,K})\widetilde{c}^{\star}_{\alpha,K}/\|\widetilde{c}^{\star}_{\alpha,K}\|_2\|_2 &= O_p(\eta_{2,n,K})\,.
\end{aligned}$$

Assumption 2.3.1(i) ensures the positive eigenfunction of $\mathbb{M}_\alpha$ exists and is identified for each $\alpha \in \mathcal{A}$. Part (ii) of Assumption 2.3.1 ensures the operators are uniformly bounded and the eigenvalues $\{\rho_\alpha : \alpha \in \mathcal{A}\}$ are uniformly well separated from the rest of the spectrum of $\{\mathbb{M}_\alpha : \alpha \in \mathcal{A}\}$. Assumption 2.3.1(ii) is implicitly satisfied by Assumption 2.3.1(i) if $\mathcal{A}$ has finite cardinality. Assumption 2.3.2 ensures the ranges of the operators $\mathbb{M}_\alpha$ are uniformly well approximated over the sieve space as K increases. Assumption 2.3.3 is required to ensure the spectrum of each $\Pi^b_K\mathbb{M}_\alpha$ remains sufficiently continuous as the dimension of the sieve space increases, and is trivially satisfied with $r(x) = x^{-1}$ if $\mathbb{M}_\alpha$ and $\Pi^b_K\mathbb{M}_\alpha|_{B_K}$ are normal or selfadjoint operators. Bounds are also available for common classes of compact operators, such as Hilbert-Schmidt and other Schatten-class operators (see Bandtlow (2004)). If T is a linear operator on a Hilbert space the lower bound $\|R(T,z)\| \geq d(z,\sigma(T))^{-1}$ for all $z \in \mathrm{res}(T)$ obtains generically. Assumption 2.3.4(i) is a standard condition for nonparametric estimation with a linear sieve space and is made to ensure that $\mathbf{G}_K$ is invertible uniformly in K. Assumption 2.3.4(ii) defines the rate of convergence of the matrix estimators. Assumption 2.3.4(ii) is sufficient for Assumption 2.3.4(iii) with $\eta_{1,n,K} = \bar{\eta}_{1,n,K}$ and $\eta_{2,n,K} = \bar{\eta}_{2,n,K}$ (by the relation between the spectral and Euclidean norms) but may lead to improved rates of convergence for $\widehat{\phi}_\alpha$ and $\widehat{\phi}^*_\alpha$ in certain circumstances.

The following two Theorems calculate the "bias" and "variance" components of the rates of convergence separately. These are proved by extending arguments in Gobet, Hoffmann, and Reiss (2004) to estimate the eigenfunction and adjoint eigenfunctions of nonselfadjoint operators.

Theorem 2.3.1. *Under Assumptions 2.3.1, 2.3.2, and 2.3.3(i), there exists* $\bar{K}$ *sufficiently large such that for each* $K \geq \bar{K}$, $\rho_{\alpha,K}$ *is real and positive and has multiplicity one and* $\phi_{\alpha,K}$ *and* $\phi^*_{\alpha,K}$ *are unique for each* $\alpha \in \mathcal{A}$, *and*

(i) $\sup_{\alpha\in\mathcal{A}} |\rho_\alpha - \rho_{\alpha,K}| = O(\delta_K)$

(ii) $\sup_{\alpha\in\mathcal{A}} \|\phi_\alpha - \phi_{\alpha,K}\| = O(\delta_K)$

(iii) $\sup_{\alpha\in\mathcal{A}} \|\phi^*_\alpha/\|\phi^*_\alpha\| - \phi^*_{\alpha,K}/\|\phi^*_{\alpha,K}\|\| = O(\delta^*_K)$

(iv) $\sup_{\alpha\in\mathcal{A}} \|\phi^*_\alpha - \phi^*_{\alpha,K}\| = O(\bar{\delta}_K)$.

Theorem 2.3.2. *Under Assumptions 2.3.1, 2.3.2, 2.3.3 and 2.3.4, there is a set whose probability approaches one on which* $\widehat{\rho}_\alpha$ *is real and positive and* $\widehat{\phi}_\alpha$ *and* $\widehat{\phi}^*_\alpha$ *are unique for each* $\alpha \in \mathcal{A}$, *and*

(i) $\sup_{\alpha\in\mathcal{A}} |\widehat{\rho}_\alpha - \rho_{\alpha,K}| = O_p(\eta_{n,K})$

(ii) $\sup_{\alpha\in\mathcal{A}} \|\widehat{\phi}_\alpha - \phi_{\alpha,K}\| = O_p(\eta_{n,K})$

(iii) $\sup_{\alpha\in\mathcal{A}} \|\widehat{\phi}^\star_\alpha/\|\widehat{\phi}^\star_\alpha\| - \phi^\star_{\alpha,K}/\|\phi^\star_{\alpha,K}\|\| = O_p(\eta_{n,K})$

(iv) $\sup_{\alpha\in\mathcal{A}} \|\widehat{\phi}^\star_\alpha - \phi^\star_{\alpha,K}\| = O_p(\bar{\eta}_{n,K})$.

The assumptions of Theorem 2.3.2 are sufficient to establish a uniform asymptotic expansion of the eigenvalue estimators $\widehat{\rho}_\alpha$.

Theorem 2.3.3. *Under Assumptions 2.3.1, 2.3.2, 2.3.3 and 2.3.4,*

$$\sup_{\alpha\in\mathcal{A}} \left| \widehat{\rho}_\alpha - \rho_{\alpha,K} - \widetilde{c}^{\star\prime}_{\alpha,K} (\widehat{\widehat{\mathbf{G}}}_K^{-1} \widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K}) \widetilde{c}_{\alpha,K} \right| = O_p(\bar{\eta}^2_{n,K}) \,.$$

2.4 Convergence of the matrix estimators

The following Lemmas are useful to verify Assumption 2.3.4 when $\mathcal{A}$ has finite cardinality. This includes the case in which $\mathcal{A}$ is a singleton, as studied in Chapter 1.

Recall that $\mathcal{F}_t = \sigma(\ldots, X_{t-1}, X_t)$ and $\mathcal{G}_t = \sigma(X_t, X_{t+1}, \ldots)$. The alpha-, rho-, beta-, and phi-mixing coefficients of $\{X_t\}$ are defined as

$$\begin{aligned}
\alpha(\tau) &= \sup_{t\in\mathbb{Z}} \sup_{A\in\mathcal{F}_t, B\in\mathcal{G}_{t+\tau}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \\
\varrho(\tau) &= \sup_{t\in\mathbb{Z}} \sup_{X\in\mathcal{L}^2(\mathcal{F}_t), Y\in\mathcal{L}^2(\mathcal{G}_{t+\tau})} |\mathrm{Corr}(X,Y)| \\
\beta(\tau) &= \frac{1}{2}\sup_{t\in\mathbb{Z}} \sup_{A_1,\ldots,A_I\in\mathcal{F}_t, B_1,\ldots,B_J\in\mathcal{G}_{t+\tau}} \sum_{(i,j)\in I\times J} |\mathbb{P}(A_i\cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| \\
\varphi(\tau) &= \sup_{t\in\mathbb{Z}} \sup_{A\in\mathcal{F}_t, B\in\mathcal{G}_{t+\tau}:\mathbb{P}(A)\neq 0} \left| \mathbb{P}(B) - \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)} \right|
\end{aligned}$$

for $\tau \in \mathbb{N}$ (see, for example, Doukhan (1994); Bradley (2005)) where $\mathcal{L}^2(\mathcal{F}_t)$ denotes (the equivalence class of) random variables measurable with respect to $\mathcal{F}_t$ with finite second moment, Corr denotes correlation, and the second supremum in the definition of $\beta(\tau)$ is taken over all pairs of partitions $(A_1, \ldots, A_I)$ and $(B_1, \ldots, B_J)$ of Ω with $I < \infty$, $J < \infty$. These mixing coefficient are related by the inequalities

$$\begin{aligned}
2\alpha(\tau) &\leq \beta(\tau) \leq \varphi(\tau) \\
4\alpha(\tau) &\leq \rho(\tau) \leq 2\sqrt{\varphi(\tau)}\,.
\end{aligned} \tag{2.4}$$

(see Bradley (2005)). The process $\{X_i\}_{i=-\infty}^{\infty}$ is said to be *algebraically beta-mixing* (at rate γ) if $\tau^\gamma \beta(\tau) = o(1)$ for some $\gamma > 1$, and *geometrically beta-mixing* if $\beta(\tau) \leq c\exp(-\gamma\tau)$ for some $\gamma > 0$ and $c \geq 0$. Forms of phi-, rho- and alpha-mixing are similarly defined.

The results are presented under different weak-dependence conditions and different assumptions on the number of moments of $m(X_0, X_1)$.

Assumption 2.4.1. *$\lambda_{\min}(\mathbf{G}_K) \geq \underline{\lambda} > 0$ for all $K \geq 1$.*

Assumption 2.4.2. *$\{X_t\}_{t\in\mathbb{Z}}$ is strictly stationary and geometrically beta-mixing.*

Assumption 2.4.3. *$\{X_t\}_{t\in\mathbb{Z}}$ is strictly stationary and geometrically rho-mixing.*

Assumption 2.4.1 is a standard assumption in nonparametric estimation with a linear sieve. Assumptions 2.4.2 and 2.4.3 are standard weak dependence conditions. Lemma 1.10.1 provides primitive sufficient conditions under which both of these assumptions are satisfied under the assumptions on $\{X_t\}$ in Chapter 1.

The results for beta-mixing data use an exponential inequality for sums of weakly-dependent random matrix random matrices developed in Chapter 4. The results for rho-mixing data follow arguments similar to Gobet, Hoffmann, and Reiss (2004) with the necessary modifications.

Let

$$\widehat{\widetilde{\mathbf{M}}}_K = \frac{1}{n}\sum_{t=0}^{n-1}\widetilde{b}^K(X_t)m(X_t,X_{t+1})\widetilde{b}^K(X_{t+1})'\,.$$

Lemma 2.4.1. *Under Assumptions 2.4.1 and 2.4.2, if m is bounded, then*

$$\|\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K\|_2 = O_p\left(\frac{\zeta_0(K)\log n}{\sqrt{n}}\right)$$

provided $\zeta_0(K)\log n/\sqrt{n} = o(1)$.

Lemma 2.4.2. *Under Assumptions 2.4.1 and 2.4.3, if $E[m(X_0,X_1)^2] < \infty$, then*

$$\|\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K\|_2 = O_p\left(\frac{\zeta_0(K)^2}{\sqrt{n}}\right)\,.$$

If, in addition, m is bounded, then

$$\|\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K\|_2 = O_p\left(\frac{\zeta_0(K)\sqrt{K}}{\sqrt{n}}\right)\,.$$

Lemma 2.4.3. *Under Assumptions 2.4.1 and 2.4.3, if $E[m(X_0,X_1)^2] < \infty$ and $\{v_K :$*

$K \geq 1\}$ *is a sequence of deterministic constants with* $v_K \in \mathbb{R}^K$ *and* $\sup_K \|v_K\|_2 < \infty$*, then*

$$\|(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)v_K\|_2 = O_p\left(\frac{\zeta_0(K)^2}{\sqrt{n}}\right).$$

If, in addition, m *is bounded, then*

$$\|(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)v_K\|_2 = O_p\left(\frac{\zeta_0(K)}{\sqrt{n}}\right).$$

Moreover, the same rates obtain for $\|(\widehat{\widetilde{\mathbf{M}}}'_K - \widetilde{\mathbf{M}}'_K)v_K\|_2$.

Rates for the estimator of the Gram matrix are also available. These are proved using arguments similar to Lemmas 2.4.1 and 2.4.2 so their proofs are omitted.

Lemma 2.4.4. *Under Assumptions 2.4.1 and 2.4.2,*

$$\|\widehat{\widetilde{\mathbf{G}}}_K - I_K\|_2 = O_p\left(\frac{\zeta_0(K)\log n}{\sqrt{n}}\right)$$

provided $\zeta_0(K)\log n/\sqrt{n} = o(1)$.

Lemma 2.4.5. *Under Assumptions 2.4.1 and 2.4.3, if* $\{v_K : K \geq 1\}$ *is a sequence of deterministic constants with* $v_K \in \mathbb{R}^K$ *and* $\sup_K \|v_K\|_2 < \infty$*, then*

$$\|(\widehat{\widetilde{\mathbf{G}}}_K - I_K)v_K\|_2 = O_p\left(\frac{\zeta_0(K)}{\sqrt{n}}\right).$$

2.5 Proofs

In what follows, unique means unique up to scale. Let $E = L^p(Q)$. Recall that the adjoint $\mathbb{M}^* : E^* \to E^*$ of $\mathbb{M}$ is defined as $\mathbb{M}^*(f^*) = f^* \circ \mathbb{M}$, i.e. $f^* \circ \mathbb{M} : E \to \mathbb{R}$ is a bounded linear functional for each $f^* \in E^*$.

Proof of Theorem 2.2.2. Part (i) $1 \leq p < \infty$: $\mathbb{M}$ is positive (by Assumption 2.2.3(i)), $\text{spr}(T)$ is a pole of the resolvent of $\mathbb{M}$ since $\mathbb{M}$ is power compact (by Assumption 2.2.2; see

Theorem 6, p. 579 of Dunford and Schwartz (1958)), and $\mathbb{M}$ is irreducible (by Assumption 2.2.3(ii); see the proof of Theorem V.6.6 of Schaefer (1974)). The result is immediate by Theorem 3.3.1 and Corollary 3.3.1.

Part (ii) $p = \infty$: the preadjoint of $\mathbb{M}$ on $L^1(Q)$ is irreducible (by the proof of Theorem V.6.6 in Schaefer (1974)). The proof for $L^1(Q)$ applied to the preadjoint of $\mathbb{M}$ provides that $\phi \in E_{++}$, ϕ is the unique eigenfunction of $\mathbb{M}$ belonging to E_+, and that ρ is an eigenvalue of $\mathbb{M}$ of multiplicity one. $\square$

Proof of Theorem 2.2.3. The positive eigenfunction and adjoint eigenfunction are unique (by Theorems 2.2.1 and 2.2.2), and ρ is an eigenvalue of $\mathbb{M}$ of multiplicity one.

Let $\overline{\mathbb{M}} = \rho^{-1}\mathbb{M}$. The condition $\mathcal{K}(x, y) > 0$ a.e.-$[Q \otimes Q]$ implies that any eigenvalue λ of $\overline{\mathbb{M}}$ with $\lambda \neq 1$ has $|\lambda| < 1$ (by Theorem 2.2.1). By construction, P is the spectral projection of $\overline{\mathbb{M}}$ corresponding to the eigenvalue 1.

Consider the bounded linear operator $\overline{\mathbb{M}} - P$. Let $\epsilon = \inf_{z \in \sigma(\overline{\mathbb{M}}):z\neq 1} |z - 1|$ and note that $\epsilon > 0$. Define $g : \mathbb{C} \to \mathbb{R}$ such that $g(z) = 1$ for all $z \in \mathbb{C}$ with $|z - 1| \leq \epsilon/2$ and $g(z) = 0$ for all $z \in \mathbb{C}$ with $|z - 1| > \epsilon/2$. Let $f : \mathbb{C} \to \mathbb{R}$ be given by $f(z) = z - g(z)$. Then

$$\begin{aligned} \overline{\mathbb{M}} - P &= \frac{-1}{2\pi i} \int_{\Gamma(1+\epsilon,0)} zR(\overline{\mathbb{M}}, z)\,\mathrm{d}z - \frac{-1}{2\pi i} \int_{\Gamma(\frac{1}{2}\epsilon,1)} R(\overline{\mathbb{M}}, z)\,\mathrm{d}z \\ &= \frac{-1}{2\pi i} \int_{\Gamma(1+\epsilon,0)} f(z)R(\overline{\mathbb{M}}, z)\,\mathrm{d}z \end{aligned}$$

(Dunford and Schwartz, 1958, Theorem 10, p. 560) since the only singularity of $R(\mathbb{M}, z)$ within $\Gamma(\frac{1}{2}\epsilon, 1)$ is at $z = 1$. By the spectral mapping theorem (Dunford and Schwartz, 1958, Theorem 11, p. 569) $\sigma(\overline{\mathbb{M}} - P) = f(\sigma(\overline{\mathbb{M}}))$. Therefore $\mathrm{spr}(\overline{\mathbb{M}} - P) < 1$ because $f(1) = 0$ and $f(\lambda) = \lambda$ for any $\lambda \in \sigma(\overline{\mathbb{M}})$ with $\lambda \neq 1$.

By the Gelfand formula (Dunford and Schwartz, 1958, p. 567),

$$\mathrm{spr}(\overline{\mathbb{M}} - P) = \lim_{\tau\to\infty} \|(\overline{\mathbb{M}} - P)^\tau\|_{L^p(Q)}^{1/\tau} < 1\,. \tag{2.5}$$

Let $\{\tau_k : k \geq 1\} \subseteq \mathbb{N}$ be the maximal subset of $\mathbb{N}$ for which $\|(\overline{\mathbb{M}} - P)^{\tau_k}\|_{L^p(Q)} > 0$ for each

τ_k. If this subsequence is finite then the proof is complete for suitable choice of c. If this subsequence is infinite, then by expression (2.5),

$$\limsup_{\tau_k \to \infty} \frac{\log \|(\overline{\mathbb{M}} - P)^{\tau_k}\|_{L^p(Q)}}{\tau_k} < 0\,.$$

Therefore, there exists a finite positive constant C such that for all τ_k large enough,

$$\log \|(\overline{\mathbb{M}} - P)^{\tau_k}\|_{L^p(Q)} \leq -C\tau_k\,.$$

Finally observe that $(\overline{\mathbb{M}} - P)^{\tau} = \overline{\mathbb{M}}^{\tau} - P = \overline{\mathbb{M}}_{\tau} - P$ since $\overline{\mathbb{M}}$ and P commute (Kato, 1980, pp. 178–179) and $\overline{\mathbb{M}} P\psi = E[\psi(X)\phi^*(X)]\overline{\mathbb{M}}\phi = P\psi$ for all $\psi \in L^p(Q)$. □

Proof of Theorem 2.2.4. Follows the same arguments as the proof of Theorem 2.2.3. □

Several lemmas are needed first before Theorems 2.3.1, 2.3.2 and 2.3.3 are proved. Parts (i) and (ii) of the following Lemma are a straightforward modification of two results in Gobet, Hoffmann, and Reiss (2004); parts (iii) and (iv) deal with estimation of the adjoint eigenfunction and are new. Lemma 2.5.2 is the key lemma from which the asymptotic linear expansion is derived.

Lemma 2.5.1. *Let $\{T_\alpha, T_{\alpha,\epsilon} : \alpha \in \mathcal{A}\}$ be a collection of linear operators on a real Hilbert space with norm $\|\cdot\|$ and inner product $\langle \cdot, \cdot \rangle$ such that T_α has an isolated real eigenvalue λ_α of multiplicity one for each $\alpha \in \mathcal{A}$. Let f_α denote the eigenfunction corresponding to λ_α normalized so that $\|f_\alpha\| = 1$. Suppose there exists a $\bar{T} < \infty$ such that $\sup_{\alpha \in \mathcal{A}} \|T_\alpha\| \leq \bar{T}$ and there exists a $\delta > 0$ such that $\inf_{z \in \sigma(T_\alpha) : z \neq \lambda_\alpha} |z - \lambda_\alpha| > \delta$ for all $\alpha \in \mathcal{A}$. Let $\bar{r} = (\sup_{\alpha \in \mathcal{A}} \sup_{z \in \Gamma(\delta, \lambda_\alpha)} \|R(T_\alpha, z)\|)^{-1}$. If $\bar{r} < \infty$ and $\sup_{\alpha \in \mathcal{A}} \|T_\alpha - T_{\alpha,\epsilon}\| < \frac{1}{2}\bar{r}$, then*

(i) The only element of $\sigma(T_{\alpha,\epsilon})$ within $\Gamma(\delta, \lambda_\alpha)$ is an real eigenvalue $\lambda_{\alpha,\epsilon}$ of multiplicity one, and $\sup_{\alpha \in \mathcal{A}} |\lambda_\alpha - \lambda_{\alpha,\epsilon}| \leq ((\bar{T} + \frac{1}{2}\bar{r})\sqrt{8}\bar{r}^{-1} + 1) \sup_{\alpha \in \mathcal{A}} \|(T_\alpha - T_{\alpha,\epsilon})f_\alpha\|$

(ii) Each $T_{\alpha,\epsilon}$ has an eigenfunction $f_{\alpha,\epsilon}$ corresponding to $\lambda_{\alpha,\epsilon}$ normalized so that $\|f_{\alpha,\epsilon}\| = 1$, and $\sup_{\alpha \in \mathcal{A}} \|f_\alpha - f_{\alpha,\epsilon}\| \leq \sqrt{8}\bar{r}^{-1} \sup_{\alpha \in \mathcal{A}} \|(T_\alpha - T_{\alpha,\epsilon})f_\alpha\|$

(iii) *Each* $T^*_{\alpha,\epsilon}$ *has an eigenfunction* $f^*_{\alpha,\epsilon}$ *corresponding to* $\lambda_{\alpha,\epsilon}$ *normalized so that* $\langle f^*_{\alpha,\epsilon}, f_{\alpha,\epsilon}\rangle = 1$, *and* $\sup_{\alpha\in\mathcal{A}} \|f^*_\alpha/\|f^*_\alpha\| - f^*_{\alpha,\epsilon}/\|f^*_{\alpha,\epsilon}\|\| \leq \sqrt{8}\bar{r}^{-1}\sup_{\alpha\in\mathcal{A}}\|(T^*_\alpha - T^*_{\alpha,\epsilon})f^*_\alpha/\|f^*_\alpha\|\|$

(iv) *Moreover,* $\sup_{\alpha\in\mathcal{A}}\|f^*_\alpha - f^*_{\alpha,\epsilon}\| \leq 2\delta\bar{r}^{-1}(\sqrt{2}\bar{r}^{-1}+1)\sup_{\alpha\in\mathcal{A}}\|T_\alpha - T_{\alpha,\epsilon}\|$.

Proof of Lemma 2.5.1. Parts (i) and (ii) are a straightforward modification of Proposition 4.2 and Corollary 4.3 of Gobet, Hoffmann, and Reiss (2004). Note that

$$\sup_{\alpha\in\mathcal{A}}\sup_{z\in\Gamma(\delta,\lambda_\alpha)}\|R(T_\alpha,z)\|\|T_\alpha - T_{\alpha,\epsilon}\| \leq \frac{1}{2}$$

holds. This implies that $\Gamma(\delta,\lambda_\alpha)$ contains precisely one eigenvalue of $T_{\alpha,\epsilon}$ by Theorem IV.3.18 of Kato (1980) and the discussion in Section IV.3.5 of Kato (1980). Therefore each $\lambda_{\alpha,\epsilon}$ must be real-valued (if it were complex-valued its conjugate would also be in $\Gamma(\delta,\lambda_\alpha)$, which would contradict there being only one eigenvalue of $T_{\alpha,\epsilon}$ in $\Gamma(\delta,\lambda_\alpha)$).

For part (iii), existence of the $f^*_{\alpha,\epsilon}$ follows from the fact that each $\lambda_{\alpha,\epsilon}$ is an eigenvalue of multiplicity one. Applying part (ii) with T^*_α in place of T_α (using the fact that an operator and its adjoint have the same norm, and that $R(T^*_\alpha,z) = R(T_\alpha,\bar{z})$) yields part (iii).

For part (iv), let P_α denote the spectral projection of T_α corresponding to λ_α and let $P_{\alpha,\epsilon}$ be the projection of $T_{\alpha,\epsilon}$ corresponding to $\lambda_{\alpha,\epsilon}$. Note that both of these may be expressed as a contour integral over $\Gamma(\delta,\lambda_\alpha)$ (cf. equation (2.2)). Therefore,

$$\begin{aligned}\sup_{\alpha\in\mathcal{A}}\|P_\alpha - P_{\alpha,\epsilon}\| &= \sup_{\alpha\in\mathcal{A}}\left\|\frac{-1}{2\pi i}\int_{\Gamma(\delta,\lambda_\alpha)} R(T_\alpha,z) - R(T_{\alpha,\epsilon},z)\,\mathrm{d}z\right\| \\ &\leq \frac{1}{2\pi}2\pi\delta\sup_{\alpha\in\mathcal{A}}\sup_{z\in\Gamma(\delta,\lambda_\alpha)}\|R(T_\alpha,z) - R(T_{\alpha,\epsilon},z)\| \\ &\leq \delta\sup_{\alpha\in\mathcal{A}}\sup_{z\in\Gamma(\delta,\lambda_\alpha)}\|R(T_\alpha,z)\|\|T_\alpha - T_{\alpha,\epsilon}\|\|R(T_{\alpha,\epsilon},z)\|\end{aligned}$$

where the second inequality is by expression (2.1). The condition $\sup_{\alpha\in\mathcal{A}}\|T_\alpha - T_{\alpha,\epsilon}\| < \frac{1}{2}\bar{r}$, together with expression (2.1), also implies that

$$\|R(T_{\alpha,\epsilon},z)\| \leq 2\|R(T_\alpha,z)\|$$

and so

$$\sup_{\alpha\in\mathcal{A}} \|P_\alpha - P_{\alpha,\epsilon}\| \le 2\delta\bar{r}^{-1} \sup_{\alpha\in\mathcal{A}} \|T_\alpha - T_{\alpha,\epsilon}\| \, . \tag{2.6}$$

Note that $P_\alpha = f_\alpha \otimes f_\alpha^*$ and $P_{\alpha,\epsilon} = f_{\alpha,\epsilon} \otimes f_{\alpha,\epsilon}^*$ with $\|P_\alpha\| = \|f_\alpha^*\|$ and $\|P_{\alpha,\epsilon}\| = \|f_{\alpha,\epsilon}^*\|$. It follows by the forward and reverse triangle inequalities and (2.6) and part (iii) that

$$\begin{aligned}
\|f_\alpha^* - f_{\alpha,\epsilon}^*\| &\le \|f_\alpha^*\| \left\| \frac{f_\alpha^*}{\|f_\alpha^*\|} - \frac{f_{\alpha,\epsilon}^*}{\|f_{\alpha,\epsilon}^*\|} \right\| + |\|f_\alpha^*\| - \|f_{\alpha,\epsilon}^*\|| \\
&\le \left(\sup_{\alpha\in\mathcal{A}} \|f_\alpha^*\| \right) \sqrt{8}\bar{r}^{-1} \sup_{\alpha\in\mathcal{A}} \|(T_\alpha^* - T_{\alpha,\epsilon}^*) f_\alpha^* / \|f_\alpha^*\|\| + 2\delta\bar{r}^{-1} \sup_{\alpha\in\mathcal{A}} \|T_\alpha - T_{\alpha,\epsilon}\| \\
&\le 2\delta\bar{r}^{-1}(\sqrt{2}\bar{r}^{-1} + 1) \sup_{\alpha\in\mathcal{A}} \|T_\alpha - T_{\alpha,\epsilon}\|
\end{aligned}$$

where the final line uses the fact that $\|f_\alpha^*\| = \|P_\alpha\| \le \delta\bar{r}^{-1}$ and the definition of the operator norm. □

Lemma 2.5.2. *Let $\{T_\alpha, T_{\alpha,\epsilon} : \alpha \in \mathcal{A}\}$ be a collection of bounded linear operators on a real Hilbert space such that T_α has an isolated real eigenvalue λ_α of multiplicity one for each $\alpha \in \mathcal{A}$. Let f_α and f_α^* denote the eigenfunctions of T_α and T_α^* corresponding to λ_α normalized so that $\|f_\alpha\| = 1$ and $\langle f_\alpha, f_\alpha^* \rangle = 1$. Suppose there exists a $\delta > 0$ such that $\inf_{z\in\sigma(T_\alpha):z\neq\lambda_\alpha} |z - \lambda_\alpha| > \delta$ for each $\alpha \in \mathcal{A}$. Let $\bar{r} = \inf_{\alpha\in\mathcal{A}} \inf_{z\in\Gamma(\delta,\lambda_\alpha)} (\|R(T_\alpha, z)\| \|T_\alpha - T_{\alpha,\epsilon}\|)^{-1}$. If $\bar{r} > 1$ then the only element of $\sigma(T_{\alpha,\epsilon})$ within $\Gamma(\delta, \lambda_\alpha)$ is a real eigenvalue $\lambda_{\alpha,\epsilon}$ of multiplicity one, and*

$$\sup_{\alpha\in\mathcal{A}} |\lambda_{\alpha,\epsilon} - \lambda_\alpha - \langle f_\alpha^*, (T_{\alpha,\epsilon} - T_\alpha) f_\alpha \rangle| \le \frac{\delta}{\bar{r}(\bar{r} - 1)} \, .$$

Proof of Lemma 2.5.2. For each $x \in \mathbb{R}$ and $\alpha \in \mathcal{A}$ define $T_{\alpha,\epsilon}(x) = T_\alpha + x(T_{\alpha,\epsilon} - T_\alpha)$. By the discussion on p. 379 of Kato (1980), the unique element of $\sigma(T_{\alpha,\epsilon}(x))$ within $\Gamma(\delta, \lambda_\alpha)$ is an eigenvalue of multiplicity one, say $\lambda_{\alpha,\epsilon}(x)$, for each $|x| < \bar{r}$, for each $\alpha \in \mathcal{A}$. Let P_α denote the spectral projection of T_α corresponding to λ_α. By the error estimates in Section

II.3.1 of Kato (1980),

$$|\lambda_{\alpha,\epsilon}(x) - \lambda_\alpha - x\mathrm{tr}\{(T_{\alpha,\epsilon} - T_\alpha)P_\alpha\}| \le \frac{|x|^2 \sup_{z\in\Gamma(\delta,\lambda_\alpha)} |z-\lambda_\alpha|}{\bar r(\bar r - |x|)} = \frac{|x|^2\delta}{\bar r(\bar r - |x|)}$$

for each $|x| < \bar r$, for each $\alpha \in \mathcal{A}$ (each T_α and $T_{\alpha,\epsilon}$ are bounded so the results from finite-dimensional perturbation theory can be applied, see Section VII.3.2 of Kato (1980)). The result follows by setting $\delta = 1$ and using the relation $\mathrm{tr}\{(T_{\alpha,\epsilon} - T_\alpha)P_\alpha\} = \{(T_{\alpha,\epsilon} - T_\alpha)(f_\alpha \otimes f_\alpha^*)\} = \langle f_\alpha^*, (T_{\alpha,\epsilon} - T_\alpha) f_\alpha\rangle$. $\square$

Lemma 2.5.3. *Under Assumptions 2.3.1, 2.3.2, and 2.3.3(i), there exists $\bar K$ sufficiently large such that for each $K \ge \bar K$, $\inf_{\alpha\in\mathcal{A}} \inf_{z\in\sigma(\Pi_K^b \mathbb{M}_\alpha): z\neq\rho_{\alpha,K}} |z - \rho_{\alpha,K}| \ge \frac{1}{2}\bar\epsilon$.*

Proof of Lemma 2.5.3. By Theorem 2.3.1 there is K_0 sufficiently large that $\sup_{\alpha\in\mathcal{A}} |\rho_\alpha - \rho_{\alpha,K}| \le \frac{1}{4}\bar\epsilon$ for all $K \ge K_0$. By Theorem IV.3.18 of Kato (1980) and the discussion in Section IV.3.5 of Kato (1980), each $\Gamma(\frac{3}{4}, \rho_\alpha)$ encloses precisely one eigenvalue of $\Pi_K^b \mathbb{M}_\alpha$ provided

$$\sup_{\alpha\in\mathcal{A}} \sup_{z\in\Gamma(\frac{3}{4}\bar\epsilon,\rho_\alpha)} \|R(\mathbb{M}_\alpha, z)\| \|\Pi_K^b \mathbb{M}_\alpha - \mathbb{M}_\alpha\| < 1\,.$$

By Assumptions 2.3.1(ii), 2.3.3(i) and 2.3.2

$$\sup_{\alpha\in\mathcal{A}} \sup_{z\in\Gamma(\frac{3}{4}\bar\epsilon,\rho_\alpha)} \|R(\mathbb{M}_\alpha, z)\| \|\Pi_K^b \mathbb{M}_\alpha - \mathbb{M}_\alpha\| \le r(\tfrac{1}{4}\bar\epsilon) \times o(1) = o(1)$$

and so $\sup_{\alpha\in\mathcal{A}} \sup_{z\in\Gamma(\frac{3}{4}\bar\epsilon,\rho_\alpha)} \|R(\mathbb{M}_\alpha, z)\| \|\Pi_K^b \mathbb{M}_\alpha - \mathbb{M}_\alpha\| < 1$ for all $K \ge K_1$ for some K_1. The result follows by setting $\bar K = \max\{K_0, K_1\}$. $\square$

Lemma 2.5.4. *Under Assumptions 2.3.1(ii) and 2.3.4(ii),*

$$\begin{aligned}
\sup_{\alpha\in\mathcal{A}} \left\| \hat{\widetilde{\mathbf{G}}}_K^{-1} \hat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K} \right\|_2 &= O_p(\bar\eta_{n,K}) \\
\sup_{\alpha\in\mathcal{A}} \left\| \hat{\widetilde{\mathbf{G}}}_K^{-1} \hat{\widetilde{\mathbf{M}}}'_{\alpha,K} - \widetilde{\mathbf{M}}'_{\alpha,K} \right\|_2 &= O_p(\bar\eta_{n,K})\,.
\end{aligned}$$

Proof of Lemma 2.5.4. First note that since $\widetilde{\mathbf{M}}_{\alpha,K}$ is isomorphic to $\Pi_K^b \mathbb{M}_\alpha|_{B_K}$,

$$\|\widetilde{\mathbf{M}}_{\alpha,K}\|_2 = \|\Pi_K^b \mathbb{M}_\alpha|_{B_K}\| \le \|\Pi_K^b \mathbb{M}_\alpha\| \le \|\mathbb{M}_\alpha\|\,.$$

Therefore, $\sup_{\alpha\in\mathcal{A}} \|\widetilde{\mathbf{M}}_{\alpha,K}\|_2$ is bounded uniformly in K by Assumption 2.3.1(ii).

The condition $\|\widehat{\widetilde{\mathbf{G}}}_K - I_K\|_2 = o_p(1)$ implies the eigenvalues of $\widehat{\widetilde{\mathbf{G}}}_K$ are bounded between $\frac{1}{2}$ and 2 on a set whose probability is approaching one. Working on this set,

$$\begin{aligned}
& \widehat{\widetilde{\mathbf{G}}}_K^{-1} \widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K} \\
= & \left(I_K - \widehat{\widetilde{\mathbf{G}}}_K^{-1} (\widehat{\widetilde{\mathbf{G}}}_K - I_K)\right) \widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K} \\
= & \widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K} - \widehat{\widetilde{\mathbf{G}}}_K^{-1} (\widehat{\widetilde{\mathbf{G}}}_K - I_K)\widetilde{\mathbf{M}}_{\alpha,K} - \widehat{\widetilde{\mathbf{G}}}_K^{-1} (\widehat{\widetilde{\mathbf{G}}}_K - I_K)(\widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K})
\end{aligned}$$

for each $\alpha \in \mathcal{A}$. The result follows by the triangle inequality and Assumption 2.3.4(ii), noting that $\|\widehat{\widetilde{\mathbf{G}}}_K^{-1}\|_2 \le 2$ whenever the eigenvalues of $\widehat{\widetilde{\mathbf{G}}}_K$ are bounded between $\frac{1}{2}$ and 2.

The proof for $\widehat{\widetilde{\mathbf{G}}}_K^{-1} \widehat{\widetilde{\mathbf{M}}}'_{\alpha,K}$ follows similar arguments, using the fact that an operator and its adjoint have the same (operator) norm. □

Lemma 2.5.5. *Under Assumptions 2.3.1(ii) and 2.3.4(iii), if $\bar{\eta}_{n,K} = o(1)$ then*

$$\begin{aligned}
\sup_{\alpha\in\mathcal{A}} \left\| (\widehat{\widetilde{\mathbf{G}}}_K^{-1} \widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K})\widetilde{c}_{\alpha,K} \right\|_2 &= O_p(\eta_{n,K}) \\
\sup_{\alpha\in\mathcal{A}} \left\| (\widehat{\widetilde{\mathbf{G}}}_K^{-1} \widehat{\widetilde{\mathbf{M}}}'_{\alpha,K} - \widetilde{\mathbf{M}}'_{\alpha,K})\widetilde{c}^\star_{\alpha,K}/\|\widetilde{c}^\star_{\alpha,K}\|_2 \right\|_2 &= O_p(\eta_{n,K})\,.
\end{aligned}$$

Proof of Lemma 2.5.5. The same arguments as the proof of Lemma 2.5.4 give

$$\sup_{\alpha\in\mathcal{A}} \left\| (\widehat{\widetilde{\mathbf{G}}}_K^{-1} \widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K})\widetilde{c}_{\alpha,K} \right\|_2 = O_p(\eta_{2,n,K}) + O_p(\eta_{1,n,K}) + O_p(\bar{\eta}_{1,n,K}) \times O_p(\eta_{2,n,K})\,.$$

The result follows by definition of $\eta_{n,K}$ and $\bar{\eta}_{n,K}$ and the condition $\bar{\eta}_{n,K} = o(1)$. The proof with $\widehat{\widetilde{\mathbf{M}}}'_{\alpha,K}$ is the same. □

Proof of Theorem 2.3.1. Apply of Lemma 2.5.1 with $\mathbb{M}_\alpha = T_\alpha$, $\Pi_K^b \mathbb{M}_\alpha = T_{\alpha,\epsilon}$ on the

Hilbert space $L^2(Q)$. Set $\Gamma(\delta, \lambda_\alpha) = \Gamma(\frac{1}{2}\bar{\epsilon}, \rho_\alpha)$. The resolvent bound in Assumption 2.3.3(i) shows that for each $\alpha \in \mathcal{A}$ and $z \in \Gamma(\frac{1}{2}\bar{\epsilon}, \rho_\alpha)$

$$2\bar{\epsilon}^{-1} = \frac{1}{d(z, \rho_\alpha)} \leq \|R(\mathbb{M}_\alpha, z)\| \leq r(d(z, \rho_\alpha)) \leq r(\tfrac{1}{2}\bar{\epsilon})$$

which implies

$$0 < r(\tfrac{1}{2}\bar{\epsilon})^{-1} \leq \bar{r} \leq \tfrac{1}{2}\bar{\epsilon} < \infty\,.$$

Assumption 2.3.2(i) implies that $\sup_{\alpha \in \mathcal{A}} \|\Pi_K^b \mathbb{M} - \mathbb{M}_\alpha\| = o(1)$ thus $\sup_{\alpha \in \mathcal{A}} \|\Pi_K^b \mathbb{M} - \mathbb{M}_\alpha\| \leq \frac{1}{2}\bar{r}$ holds for all K sufficiently large. □

Proof of Theorem 2.3.2. Apply Lemma 2.5.1 with $\widetilde{\mathbf{M}}_{\alpha,K} = T_\alpha$, $\widehat{\widehat{\mathbf{G}}}_K^{-1} \widehat{\widetilde{\mathbf{M}}}_{\alpha,K} = T_{\alpha,\epsilon}$ on the Hilbert space $\mathbb{R}^K$ (with the Euclidian inner (dot) product). Set $\Gamma(\delta, \lambda_\alpha) = \Gamma(\frac{1}{4}\bar{\epsilon}, \rho_{\alpha,K})$. By Lemma 2.5.3 there is a $\bar{K}$ sufficiently large that $\inf_{\alpha \in \mathcal{A}} \inf_{z \in \sigma(\Pi_K^b \mathbb{M}_\alpha): z \neq \rho_{\alpha,K}} |z - \rho_{\alpha,K}| \geq \frac{1}{2}\bar{\epsilon}$ for all $K \geq \bar{K}$. Take $K \geq \bar{K}$. The fact that $\widetilde{\mathbf{M}}_{\alpha,K}$ and $\Pi_K^b \mathbb{M}_\alpha|_{B_K}$ are isomorphic and the resolvent bound in Assumption 2.3.3(ii) shows that for each $\alpha \in \mathcal{A}$ and $z \in \Gamma(\frac{1}{4}\bar{\epsilon}, \rho_{\alpha,K})$

$$4\bar{\epsilon}^{-1} = \frac{1}{d(z, \rho_{\alpha,K})} \leq \|R(\Pi_K^b \mathbb{M}_\alpha|_{B_K}, z)\| \leq r(d(z, \rho_\alpha)) \leq r(\tfrac{1}{4}\bar{\epsilon})$$

which implies

$$0 < r(\tfrac{1}{4}\bar{\epsilon})^{-1} \leq \bar{r} \leq \tfrac{1}{4}\bar{\epsilon} < \infty\,.$$

Lemma 2.5.4 provides that $\sup_{\alpha \in \mathcal{A}} \|\widehat{\widehat{\mathbf{G}}}_K^{-1} \widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K}\|_2 = O_p(\bar{\eta}_{n,K})$. This and the condition $\bar{\eta}_{n,K} = o(1)$ implies the condition

$$\sup_{\alpha \in \mathcal{A}} \left\| \widehat{\widehat{\mathbf{G}}}_K^{-1} \widehat{\widetilde{\mathbf{M}}}_{\alpha,K} - \widetilde{\mathbf{M}}_{\alpha,K} \right\|_2 \leq \frac{1}{2}\bar{r}$$

holds on a set whose probability is approaching one. Application of Lemma 2.5.1 on this set proves parts (i) and (ii), with $\eta_{n,K}$ given by Lemma 2.5.5.

Part (iii) follows by repeating this argument with $\widehat{\widetilde{\mathbf{M}}}'_{\alpha,K}$and $\widetilde{\mathbf{M}}'_{\alpha,K}$ in place of $\widehat{\widetilde{\mathbf{M}}}_{\alpha,K}$

and $\widetilde{\mathbf{M}}_{\alpha,K}$.

For part (iv), write

$$\begin{aligned}\sup_{\alpha\in\mathcal{A}}\|\widehat{\phi}^\star_\alpha-\phi^\star_{\alpha,K}\| &\leq \sup_{\alpha\in\mathcal{A}}\|\phi^\star_{\alpha,K}\|\left\|\frac{\widehat{\phi}^\star_\alpha}{\|\widehat{\phi}^\star_\alpha\|}-\frac{\phi^\star_{\alpha,K}}{\|\phi^\star_{\alpha,K}\|}\right\|+\sup_{\alpha\in\mathcal{A}}|\|\widehat{\phi}^\star_{\alpha,K}\|-\|\phi^\star_{\alpha,K}\||\\ &\leq \sup_{\alpha\in\mathcal{A}}\|\phi^\star_{\alpha,K}\|\times O_p(\eta_{n,K})+\sup_{\alpha\in\mathcal{A}}|\|\widehat{\phi}^\star_{\alpha,K}\|-\|\phi^\star_{\alpha,K}\||\end{aligned}$$

where the second line is by part (iii). Theorem 2.3.1 gives $\sup_{\alpha\in\mathcal{A}}\|\phi^\star_{\alpha,K}\|\leq\sup_{\alpha\in\mathcal{A}}\|\phi^*_\alpha\|+o(1)$ and it is easy to see that $\sup_{\alpha\in\mathcal{A}}\|\phi^*_\alpha\|<\infty$. The remaining term $\sup_{\alpha\in\mathcal{A}}|\|\widehat{\phi}^\star_{\alpha,K}\|-\|\phi^\star_{\alpha,K}\||$ can be shown to be $O_p(\bar{\eta}_{n,K})$ using a similar argument to the proof of part (iv) of Lemma 2.5.1. $\square$

Proof of Theorem 2.3.3. First apply Lemma 2.5.1 with $T_\alpha=\widetilde{\mathbf{M}}_{\alpha,K}$ and $T_{\alpha,\epsilon}=\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K}$. By Lemma 2.5.3 there is a $\bar{K}$ sufficiently large that $\inf_{\alpha\in\mathcal{A}}\inf_{z\in\sigma(\Pi^b_K\mathbb{M}_\alpha):z\neq\rho_{\alpha,K}}|z-\rho_{\alpha,K}|\geq\frac{1}{2}\bar{\epsilon}$ for all $K\geq\bar{K}$. Take $K\geq\bar{K}$. Then $\|R(\widetilde{\mathbf{M}}_{\alpha,K},z)\|\leq r(\frac{1}{4}\bar{\epsilon})$ for all $z\in\Gamma(\frac{1}{4}\bar{\epsilon},\rho_{\alpha,K})$ by Assumption 2.3.3(ii). The condition

$$r(\tfrac{1}{4}\bar{\epsilon})>\|\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K}-\widetilde{\mathbf{M}}_{\alpha,K}\|_2$$

holds on a set whose probability is approaching one, since $\|\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K}-\widetilde{\mathbf{M}}_{\alpha,K}\|_2=o_p(1)$ by Lemma 2.5.4 and the condition $\bar{\eta}_{n,K}=o(1)$. Therefore, the condition

$$r_{n,K}:=\inf_{\alpha\in\mathcal{A}}\inf_{z\in\Gamma(\frac{1}{4}\bar{\epsilon},\rho_{\alpha,K})}(\|R(\widetilde{\mathbf{M}}_{\alpha,K},z)\|_2\|\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K}-\widetilde{\mathbf{M}}_{\alpha,K}\|_2)^{-1}>1$$

holds on a set whose probability is approaching one, on which Lemma 2.5.1 provides that

$$\sup_{\alpha\in\mathcal{A}}\left|\widehat{\rho}_{\alpha,K}-\rho_\alpha-\widetilde{c}^{\star\prime}_{\alpha,K}(\widehat{\widetilde{\mathbf{G}}}_K^{-1}\widehat{\widetilde{\mathbf{M}}}_{\alpha,K}-\widetilde{\mathbf{M}}_{\alpha,K})\widetilde{c}_{\alpha,K}\right|\leq\frac{\bar{\epsilon}}{4r_{n,K}(r_{n,K}-1)}$$

uniformly for $\alpha \in \mathcal{A}$. The result follows by noticing that

$$\frac{1}{r_{n,K}(r_{n,K}-1)} = O_p(\bar{\eta}_{n,K}^2)$$

by definition of $r_{n,K}$ and Lemma 2.5.4. □

Proof of Lemma 2.4.1. Let $\bar{M}$ be a finite positive constant such that $m(x_0, x_1) \leq \bar{M}$. Let $m_{t,t+1} = m(X_t, X_{t+1})$. Consider the $K \times K$ random matrix

$$\Xi_{t,n} = n^{-1}\left(\widetilde{b}^K(X_t)m_{t,t+1}\widetilde{b}^K(X_{t+1}) - E[\widetilde{b}^K(X_0)m_{0,1}\widetilde{b}^K(X_1)]\right)$$

where clearly $E[\Xi_{t,n}] = 0$. Assumption 2.4.1 and definition of $\zeta_0(K)$ imply that

$$\|\Xi_{t,n}\|_2 \leq \frac{2\zeta_0(K)^2\bar{M}}{\underline{\lambda}n}\,. \tag{2.7}$$

By the triangle and Cauchy-Schwarz inequalities, for any $u, v \in \mathbb{R}^K$ with $u'u = 1$ and $v'v = 1$,

$$\begin{aligned} n^2E[u'\Xi_{t,n}\Xi'_{s,n}v] &\leq |E[u'\widetilde{b}^K(X_t)m_t\widetilde{b}^K(X_{t+1})'\widetilde{b}^K(X_{s+1})m_s\widetilde{b}^K(X_s)'v]| \\ &\quad +|E[u'\widetilde{b}^K(X_0)m_0\widetilde{b}^K(X_1)']E[\widetilde{b}^K(X_1)m_0\widetilde{b}^K(X_0)'v]| \\ &\leq \bar{M}^2E[|u'\widetilde{b}^K(X_t)||\widetilde{b}^K(X_{t+1})'\widetilde{b}^K(X_{s+1})||\widetilde{b}^K(X_s)'v|] \\ &\quad +\bar{M}^2\underline{\lambda}^{-1}\zeta_0(K)^2E[(u'\widetilde{b}^K(X_0))^2]^{1/2}E[(\widetilde{b}^K(X_0)'v)^2]^{1/2} \\ &\leq 2\underline{\lambda}^{-1}\bar{M}^2\zeta_0(K)^2E[(u'\widetilde{b}^K(X_0))^2]^{1/2}E[(\widetilde{b}^K(X_0)'v)^2]^{1/2} \\ &\leq 2\underline{\lambda}^{-1}\bar{M}^2\zeta_0(K)^2 \end{aligned}$$

where the final line is because $E[(u'\widetilde{b}^K(X_0))^2] = u'E[\widetilde{b}_K(X_0)\widetilde{b}_K(X_0)']u = u'u = 1$ for any $u \in \mathbb{R}^K$ with $u'u = 1$. Since $\|A\|_2 = \sup_{u,v\in\mathbb{R}^K:u'u=1,v'v=1} u'Av$ for any $K \times K$ matrix A,

$$\left\|E\left[\Xi_{t,n}\Xi'_{s,n}\right]\right\|_2 \leq \frac{2\bar{M}^2\zeta_0(K)^2}{\underline{\lambda}n^2}$$

and similarly for $\|E[\Xi'_{t,n}\Xi_{s,n}]\|_2$ by Corollary 4.4.2. $\square$

Proof of Lemma 2.4.2. By geometric rho-mixing there exists a finite positive C such that

$$\mathrm{Var}\left[\sum_{t=0}^{n-1} b(X_t, X_{t+1})\right] \leq CnE[b(X_0, X_1)^2]$$

uniformly for all measurable $b : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that $E[b(X_0, X_1)^2] < \infty$ (see Lemma 1.10.2). By the relation between the spectral and Frobenius norms,

$$\begin{aligned}
E[\|\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K\|_2^2] &\leq \frac{1}{n^2}\sum_{k=1}^{K}\sum_{l=1}^{K} \mathrm{Var}\left[\sum_{t=0}^{n-1} \tilde{b}_k^K(X_t)\tilde{b}_l^K(X_{t+1})m(X_t, X_{t+1})\right] \\
&\leq \frac{C}{n}\sum_{k=1}^{K}\sum_{l=1}^{K} E\left[(\tilde{b}_k^K(X_0)\tilde{b}_l^K(X_1)m(X_0, X_1))^2\right] \\
&\leq \frac{C\zeta_0(K)^2}{\underline{\lambda} n}\sum_{k=1}^{K} E\left[(\tilde{b}_k^K(X_0)m(X_0, X_1))^2\right]
\end{aligned}$$

where

$$\sum_{k=1}^{K} E\left[(\tilde{b}_k^K(X_0)m(X_0, X_1))^2\right] \leq \begin{cases} \underline{\lambda}^{-1}\zeta_0(K)^2 E[m(X_0, X_1)^2] & \\ K \sup_{x_0,x_1} |m(x_0, x_1)|^2 & \text{if } m \text{ is bounded.} \end{cases}$$

The result follows by Markov's inequality. $\square$

Proof of Lemma 2.4.3. By the arguments in the Proof of Lemma 2.4.2

$$\begin{aligned}
E[\|(\widehat{\widetilde{\mathbf{M}}}_K - \widetilde{\mathbf{M}}_K)v_K\|_2^2] &\leq \frac{1}{n^2}\sum_{k=1}^{K} \mathrm{Var}\left[\sum_{t=0}^{n-1} \tilde{b}_k^K(X_t)(\tilde{b}^K(X_{t+1})'v_K)m(X_t, X_{t+1})\right] \\
&\leq \frac{C}{n}\sum_{k=1}^{K} E\left[(\tilde{b}_k^K(X_0)(\tilde{b}^K(X_1)'v_K)m(X_0, X_1))^2\right] \\
&\leq \frac{C\zeta_0(K)^2}{\underline{\lambda} n} E\left[((\tilde{b}^K(X_1)'v_K)^2 m(X_0, X_1)^2\right]
\end{aligned}$$

for some finite positive constant C, where

$$E\left[((\widetilde{b}^K(X_1)'v_K)^2 m(X_0,X_1)^2\right] \leq \begin{cases} \underline{\lambda}^{-1}\zeta_0(K)^2\|v_K\|_2^2 E[m(X_0,X_1)^2] & \\ \\ \|v_K\|_2^2 \sup_{x_0,x_1}|m(x_0,x_1)|^2 & \text{if } m \text{ is bounded.} \end{cases}$$

The result follows by Markov's inequality. The proof with $\widehat{\widetilde{\mathbf{M}}}_K'$ is identical. $\square$

Chapter 3

Nonparametric identification of positive eigenfunctions

Recent work in economics has shown that important features of some models may be expressed as positive eigenfunctions of positive linear operators.[1] The eigenvalues associated with these positive eigenfunctions also have specific economic interpretations. As discussed in Chapter 1, Hansen and Scheinkman (2009) and Hansen (2012) show that information about the pricing of long-horizon assets can be extracted from a dynamic asset pricing model by studying the positive eigenfunction and eigenvalues of a collection of pricing operators. Moreover, the positive eigenfunction of this collection of pricing operators may, together with its eigenvalue, be used to decompose the stochastic discount factor into its permanent and transitory components (see also Alvarez and Jermann (2005) and Backus, Chernov, and Zin (2012)). In a related application, Hansen, Heaton, and Li (2008) and Hansen and Scheinkman (2012a,b) use positive eigenfunctions to analyze valuation in economies with recursive preferences and stochastic growth. Chen, Chernozhukov, Lee, and Newey (2014a) study identification in a semiparametric consumption capital asset pricing

[1] More generally the quantities can be expressed as a positive eigenvector of a linear operator on a vector space. The term eigenfunction is used when the quantity of interest is assumed to lie in a function space. As most spaces of interest in these economic applications are function spaces, the term eigenfunction is used henceforth.

model (C-CAPM) by constructing a positive operator whose positive eigenfunction is the external habit formation function to be identified. In their application, the eigenvalue associated with the positive eigenfunction is the reciprocal of the time preference parameter. Ross (2013) shows that the pricing kernel may be nonparametrically recovered from option prices by solving a positive eigenfunction problem.

The purpose of this chapter is to provide a set of nonparametric identification conditions for positive eigenfunctions, which are tractable and have general application in economics.[2] Existence conditions are also discussed. The genesis of this analysis is the classical Perron-Frobenius Theorem, which asserts that a positive matrix has a unique positive eigenvector. Identification conditions for positive operators on $L^p(\mu)$ spaces have been provided in Chen et al. (2014a) and Chapters 1 and 2 in the context of a semiparametric C-CAPM with habit formation and the Hansen and Scheinkman (2009) long-run analysis in discrete-time models, respectively. Identification is achieved in these studies by placing primitive conditions on the integrability and positivity of the integral kernel of the operator, from which identification follows by application of an integral-operator analogue of the Perron-Frobenius Theorem.

Taking a somewhat different tack, Hansen and Scheinkman (2009) use Markov process theory to derive identification conditions for the positive eigenfunction of pricing operators in continuous-time environments. They show that the positive eigenfunction is identified if the SDF process is strictly positive, and that the state process when discretely-sampled is irreducible (in the sense of Markov processes), Harris recurrent under the measure induced by the permanent component of the SDF, and that there exists a stationary distribution for the conditional expectations induced by the change of measure.

Identification of the positive eigenfunction of positive linear operators on a Banach lattice has been well studied in abstract functional analysis in the mathematics literature (see, e.g., Kreĭn and Rutman (1950); Schaefer (1960)). However, such function-analytic identi-

[2]The conditions identify the positive eigenfunction up to scale (any positive multiple of a positive eigenfunction is a positive eigenfunction). If the positive eigenfunction is normalized to have unit norm, then the conditions are sufficient for the point identification of the normalized positive eigenfunction.

fication conditions are typically very high level, relating to irreducibility of the operator and other properties of the resolvent of the operator.

This chapter aims to help bridge the gap between the high-level conditions in functional analysis and the primitive conditions for integral operators by providing a reasonably tractable set of identification conditions for the positive eigenfunction of a positive operator on a Banach lattice. When applied to the special case of operators on $L^2(\mu)$ spaces, the identification conditions herein are weaker than those previously provided in Chen et al. (2014a) and Chapter 1.[3] The identification conditions in Chen et al. (2014a) and Chapters 1 and 2 are positivity and integrability restrictions on the kernel representing the operator. In the applications considered in Chen et al. (2014a) and Chapter 1, the integral kernel is formulated in terms of, inter alia, stationary and conditional probability densities. Such identification conditions are effectively restrictions on probability densities. By contrast, the identification conditions in this chapter require a weaker form of compactness that that in Chen et al. (2014a) and Chapter 1 and are not formulated in terms of the integral kernel. This distinction is relevant for several applications in economics including, but not limited to, nonparametric identification of external habit formation models and asset pricing in higher-order Markov environments.

The general results are applied to obtain identification conditions for the positive eigenfunctions of pricing operators in dynamic asset pricing models. Identification conditions are presented for both stationary discrete-time environments and stationary continuous-time environments. The identification conditions complement those that Hansen and Scheinkman (2009) provide for general continuous-time environments. Nonparametric identification is shown to hold under a no-arbitrage type condition, a primitive condition on the time-series behavior of the variables in the model, and power compactness of an operator.

[3]The material presented in this chapter was first written in October 2012 presented in November 2012. In June 2013, the author of this dissertation and the authors of Escanciano and Hoderlein (2012) became aware of each other's independent unpublished works. Both works make use of similar tools from functional analysis, but focus on different applications. The present chapter considers application to stochastic discount factor decomposition. By contrast, Escanciano and Hoderlein (2012) consider nonparametric identification of marginal utilities in consumption-based asset pricing models and rely heavily on the representation of their operator as an integral operator with positive kernel.

Existence and nonparametric identification of the positive eigenfunction in discrete-time environments is established under an additional condition on the yield on zero-coupon bonds.

The identification results presented in this chapter establish several further properties of the positive eigenfunction. These properties are useful for estimation of the positive eigenfunction and its eigenvalue. In particular, under the identification conditions, the eigenvalue with which the positive eigenfunction is associated is the largest eigenvalue of the operator, and the largest eigenvalue is real, positive, and has a unique eigenfunction. This implies (under an additional mild regularity condition) that the positive eigenfunction and its eigenvalue are continuous with respect to perturbations of the underlying operator. Intuitively, if one can construct an estimator of the operator that is "close" to the true operator in an appropriate sense, then the maximum eigenvalue of the estimator, and its eigenfunction, should be close to the true eigenvalue/eigenfunction. This intuition is borne out in the asymptotic theory presented in Chapters 1 and 2.

The remainder of this chapter is organized as follows. Section 3.1 provides identification conditions for positive operators on $L^p(\mu)$ spaces. Section 3.2 then applies these conditions to study identification of the positive eigenfunction of pricing operators. Section 3.3 presents identification conditions for positive operators on Banach lattices. All proofs are presented in Section 3.5.

Definitions of relevant concepts from spectral theory and the theory of Banach lattices and positive operators are as in Section 2.1.

3.1 Positive operators on $L^p(\mu)$ spaces

This section provides sufficient conditions for the existence and nonparametric identification of positive eigenfunctions of positive linear operators on $L^p(\mu)$ spaces. Let $(\mathcal{X}, \mathscr{X}, \mu)$ be a σ-finite measure space and let $L^p(\mu)$ denote the space $L^p(\mathcal{X}, \mathscr{X}, \mu)$.[4] Let $T : L^p(\mu) \to L^p(\mu)$

[4]Recall that the space $L^p(\mu) := L^p(\mathcal{X}, \mathscr{X}, \mu)$ with $1 \leq p < \infty$ consists of all (equivalence classes of) measurable functions $f : \mathcal{X} \to \mathbb{R}$ such that $\int |f|^p \,\mathrm{d}\mu < \infty$. The space $L^\infty(\mu) := L^\infty(\mathcal{X}, \mathscr{X}, \mu)$ consists of

be a bounded linear operator. The operator T is said to be positive if $Tf \geq 0$ a.e.-$[\mu]$ whenever $f \geq 0$ a.e.-$[\mu]$. The identification conditions presented below are weaker than those in Chen et al. (2014a) and Chapter 1 (see Section 3.1.3). The cases $1 \leq p < \infty$ and $p = \infty$ are dealt with separately because of their different topologies.

3.1.1 Case 1: $1 \leq p < \infty$

The following conditions are sufficient for existence and nonparametric identification of the positive eigenfunction of a bounded linear operator $T : L^p(\mu) \to L^p(\mu)$ when $1 \leq p < \infty$.

Assumption 3.1.1. *(i) T is positive, (ii) T^n is compact for some $n \geq 1$, (iii)* $\mathrm{spr}(T) > 0$*, and (iv) for each $f \in L^p(\mu)$ such that $f \geq 0$ a.e.-$[\mu]$ and $f \neq 0$ there exists $n \geq 1$ such that $T^n f > 0$ a.e.-$[\mu]$.*

These conditions will be referred to as *positivity*, *power compactness*, *non-degeneracy*, and *eventual strong positivity*. The first positivity condition is typically an easy condition to motivate, either from the economic context of the problem or the structure of the operator. The power compactness condition is trivially satisfied if T is compact, though this more general condition suffices. Importantly, the non-degeneracy condition is trivially satisfied if T has a nonzero eigenvalue. The non-degeneracy condition is also satisfied if there exists a nonzero $f \in L^p(\mu)$ with $f \geq 0$ a.e.-$[\mu]$ for which $T^n f \geq \delta f$ a.e.-$[\mu]$ for some $n \geq 1$ and $\delta > 0$ (Schaefer, 1960, Proposition 3). Both the non-degeneracy and eventual strong positivity conditions have an economic interpretation in the asset pricing context discussed subsequently.

The following Theorem shows that Assumption 3.1.1 is sufficient for nonparametric identification of the positive eigenfunction of both T and its adjoint T^*. The Theorem also shows that the positive eigenfunctions of T and T^* correspond to the same eigenvalue, namely $\mathrm{spr}(T)$. Let $q = \infty$ if $p = 1$, otherwise let q be such that $q^{-1} + p^{-1} = 1$. The space $L^q(\mu)$ can be identified as the dual space of $L^p(\mu)$. The adjoint $T^* : L^q(\mu) \to L^q(\mu)$ of T

all (equivalence classes of) measurable functions $f : \mathcal{X} \to \mathbb{R}$ such that $\operatorname{ess\,sup} |f| < \infty$.

is a bounded linear operator defined by the equality

$$\int g(Tf)\,\mathrm{d}\mu = \int (T^* g) f\,\mathrm{d}\mu \tag{3.1}$$

for all $f \in L^p(\mu)$ and $g \in L^q(\mu)$.

Theorem 3.1.1. *Under Assumption 3.1.1, there exist eigenfunctions $\bar{f} \in L^p(\mu)$ and $\bar{f}^* \in L^q(\mu)$ of T and T^* with eigenvalue $\mathrm{spr}(T)$, such that $\bar{f} > 0$ a.e.-$[\mu]$ and $\bar{f}^* > 0$ a.e.-$[\mu]$. The eigenfunctions $\bar{f}$ and $\bar{f}^*$ are the unique eigenfunctions of T and T^* that are non-negative a.e.-$[\mu]$. Moreover, $\mathrm{spr}(T)$ is an isolated point of $\sigma(T)$ and has multiplicity one.*

Note that Theorem 3.1.1 shows not just that there is a unique *positive* eigenfunction of T, but that it is also the unique *non-negative* eigenfunction of T.

3.1.2 Case 2: $p = \infty$

Sufficient conditions for existence and nonparametric identification of the positive eigenfunction of a bounded linear operator $T : L^\infty(\mu) \to L^\infty(\mu)$ are now provided. There are two important differences from the $1 \leq p < \infty$ case. First, the eventual strong positivity condition needs to be strengthened. Second, identification of the adjoint eigenfunction is not considered, as the dual space of $L^\infty(\mu)$ is typically identified with a space of signed measures (rather than a function space).

Assumption 3.1.2. *(i) T is positive, (ii) T^n is compact for some $n \geq 1$, (iii) $\mathrm{spr}(T) > 0$, and (iv) for each $f \in L^\infty(\mu)$ such that $f \geq 0$ a.e.-$[\mu]$ and $f \neq 0$ there exists $n \geq 1$ such that $\operatorname{ess\,inf} T^n f > 0$.*

As with the preceding case, positivity may be motivated by the economic context of the problem or the structure of the operator. Assumption 3.1.2(iv) appears stronger than Assumption 3.1.1(iv). This modification is required because the topological properties of the space $L^\infty(\mu)$ are different from those of the other $L^p(\mu)$ spaces.

Theorem 3.1.2. *Under Assumption 3.1.2, there exists an eigenfunction $\bar{f} \in L^\infty(\mu)$ of T with eigenvalue* $\mathrm{spr}(T)$*, such that* $\operatorname{ess\,inf} \bar{f} > 0$ *and $\bar{f}$ is the unique eigenfunction of T that is non-negative a.e.-$[\mu]$. Moreover,* $\mathrm{spr}(T)$ *is an isolated point of $\sigma(T)$ and has multiplicity one.*

3.1.3 Related literature

The identification conditions in Chen et al. (2014a) and Chapters 1 and 2 are cast in terms of primitive positivity and integrability restrictions on the integral kernel representing the operator T. For the space $L^2(\mu)$, the conditions in Chen et al. (2014a) and Chapter 1 imply that $T : L^2(\mu) \to L^2(\mu)$ is given by

$$Tf(x) = \int k(x,y) f(y) \,\mathrm{d}\mu(y)$$

where the integral kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ satisfies:

(a) $k(x,y) > 0$ a.e.-$[\mu \otimes \mu]$, and

(b) $\int k^2 \,\mathrm{d}(\mu \otimes \mu) < \infty$.

Condition (a) guarantees positivity and eventual strong positivity (Assumptions 3.1.1(i) and (iv)). Condition (b) implies that T is Hilbert-Schmidt, and therefore compact, and therefore power compact (Assumption 3.1.1(ii)). Moreover, these two conditions jointly guarantee that $\mathrm{spr}(T) > 0$ (Assumption 3.1.1(iii); see Theorem V.6.6 of Schaefer (1974)). Therefore, Conditions (a) and (b) are sufficient for both existence and identification of the positive eigenfunction in the space $L^2(\mu)$.[5]

In the applications considered in Chen et al. (2014a) and Chapter 1, the kernel k is formulated in terms of, inter alia, stationary and conditional probability densities. Therefore, in these applications, Conditions (a) and (b) are effectively positivity and integrability restrictions on the probability densities.

[5]The identification conditions in Chapter 2 replace Condition (b) with a power compactness condition, but still impose primitive positivity restrictions on the kernel k (see Assumption 2.2.3).

In a number of economic applications the conditional density may fail to exist, making conditions on k less interpretable. More critically, in these applications the operator may be noncompact and therefore violate Condition (b). These applications include, but are not limited to, nonparametric marginal utility with external habit formation and asset pricing in higher-order Markov environments. The weaker identification conditions presented in Theorem 3.1.1 may prove useful in establishing identification in these applications, as outlined in the next example and the example in Section 3.2.1.

Example: external habit formation

Following Chen and Ludvigson (2009), assume that the representative agent's marginal utility of consumption at time t is given by

$$MU_t = C_t^{-\gamma}(1 - g(C_{t-1}/C_t, \ldots, C_{t-L}/C_t))^{-\gamma}$$

where C_t denotes consumption of the agent at time t and γ is a risk-aversion coefficient. When current and lagged consumption belong to the agent's information set at time t, the Euler equation can be re-written as

$$\begin{aligned} \beta^{-1} h(G_{t-1}, \ldots, G_{t-L}) &= E[G_{t+1}^{-\gamma} R_{i,t+1} h(G_t, \ldots, G_{t-L+1}) | G_{t-1}, \ldots, G_{t-L}] \\ &=: Th(G_{t-1}, \ldots, G_{t-L}) \end{aligned} \tag{3.2}$$

where $G_t = C_t/C_{t-1}$ is consumption growth from time $t-1$ to time t, $R_{i,t+1}$ is the (gross) return on asset i from time t to time $t+1$, β is the agent's time preference parameter, and

$$h(G_{t-1}, \ldots, G_{t-L}) = (1 - g(C_{t-1}/C_t, \ldots, C_{t-L}/C_t))^{-\gamma}.$$

Chen et al. (2014a) use the eigenfunction representation (3.2) to study nonparametric identification of h (given γ) when $L = 1$. In particular, Chen et al. (2014a) assume $\{G_t\}$ is a stationary process and place positivity and integrability restrictions on the density of G_{t+1} conditional on G_t as in Conditions (a) and (b) above. When $L > 1$ the density of $X_t := (G_t, \ldots, G_{t-L+1})$ given $X_{t-1} := (G_{t-1}, \ldots, G_{t-L})$ is degenerate because elements $G_{t-1}, \ldots, G_{t-L+1}$ of X_t are known when X_{t-1} is known. Moreover, when $L > 1$ the operator T may be noncompact and therefore violate Condition (b) (cf. Lemma 3.2.3). Theorem 3.1.1 may be used to study nonparametric identification of h with $L > 1$, although formal verification of Assumption 3.1.1 for this model is beyond the scope of this chapter.

3.2 Application: dynamic asset pricing models and the long run

Central to the literature following Hansen and Scheinkman (2009), and indeed any subsequent econometric implementation of this analysis, is the existence and identification of the positive eigenfunction of this family of operators.

New identification conditions are now presented for both discrete- and continuous-time environments by applying Theorem 3.1.1. The conditions relate to the inherent positivity of prices assigned to claims to non-negative payoffs. The identification conditions for continuous-time environments used here are stronger than those of Hansen and Scheinkman (2009) (for example, here a power compactness condition is imposed on the pricing operator and the environment is assumed to be stationary).[6] However, the stronger conditions both guarantee uniqueness of the positive eigenfunction and show how its eigenvalue is related to the spectrum of the pricing operators. Existence of the positive eigenfunction in discrete-time environments is also established here under a condition on the yield on zero-coupon bonds.

[6]The assumption of stationarity is not necessarily restrictive in an asset pricing context, and may sometimes be achieved by a change of variables. For example, consumption-based asset pricing models are often written in terms of consumption growth to avoid potential nonstationarity in aggregate consumption (Hansen and Singleton, 1982; Gallant and Tauchen, 1989)).

3.2.1 Identification in discrete-time environments

The identification conditions are presented for the more general environment described in Section 1.5.2 and Hansen and Scheinkman (2012b, 2013). As in Section 1.5.2, let the economy be characterized by a strictly stationary, discrete-time, first-order Markov state process $\{(X_t, Y_t)\}_{t=0}^{\infty}$ with support $\mathcal{X} \times \mathcal{Y}$ where $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ and $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$ are Borel sets, and where the joint distribution of (X_{t+1}, Y_{t+1}) conditioned on (X_t, Y_t) depends only on X_t. The single-period pricing operator $\mathbb{M}$ is given by

$$\mathbb{M}\psi(x) = E\left[m(X_t, X_{t+1}, Y_{t+1})\psi(X_{t+1})|X_t = x\right] \tag{3.3}$$

where $m : \mathcal{X} \times \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ is the single-period SDF function. The n-period pricing operator $\mathbb{M}_n$ is given by

$$\mathbb{M}_n\psi(x) = E\left[\left(\prod_{s=0}^{n-1} m(X_{t+s}, X_{t+s+1}, Y_{t+s+1})\right)\psi(X_{t+n})\middle| X_t = x\right]. \tag{3.4}$$

Each $\mathbb{M}_n$ factorizes as $\mathbb{M}_n = \mathbb{M}^n$ as a consequence of Markovianity of the state process. This general framework clearly the setting in which the state process is $\{X_t\}$ and $m(X_t, X_{t+1}, Y_{t+1}) = m(X_t, X_{t+1})$ as discussed in Chapter 1. However, unlike the identification conditions presented in Chapter 1, the transition density of X_{t+1} conditional on X_t is not required to exist. Thus the identification conditions presented in this chapter apply to environments characterized by higher-order Markov processes which would be precluded from the analysis of Chapter 1 (see Section 3.2.1).

The following identification results assume the operator $\mathbb{M}$ is observed. That is, the researcher knows both the distribution of (X_{t+1}, Y_{t+1}) given X_t, and the function m. Extension to the case in which a component of the state vector is latent is deferred to future research.

Let Q denote the stationary distribution of X on $\mathcal{X}$ and let $\mathscr{X}$ denote the Borel σ-algebra on $\mathcal{X}$. Identification conditions are now presented for the space $L^p(Q) :=$

$L^p(\mathcal{X}, \mathscr{X}, Q)$ for some $1 \leq p < \infty$. Extension to $L^\infty(Q)$ is straightforward.

Definition 3.2.1. *$\phi \in L^p(Q)$ is a* principal eigenfunction *of the family of pricing operators $\{\mathbb{M}_n : n \geq 1\}$ if $\phi > 0$ a.e.-$[Q]$ and $\mathbb{M}_n \phi = \rho^n \phi$ for each $n \geq 1$ and some $\rho \in \mathbb{R}$.*

Remark 3.2.1. *Let $\mathbb{M}\phi = \rho\phi$. The factorization $\mathbb{M}_n = \mathbb{M}^n$ implies that $\mathbb{M}_n \phi = \rho^n \phi$. Therefore, ϕ is a principal eigenfunction of $\{\mathbb{M}_n : n \geq 1\}$ if and only if ϕ is a positive eigenfunction of $\mathbb{M}$.*

Theorem 3.1.1 may therefore be applied to provide identification conditions for the principal eigenfunction in this discrete-time setting. A boundedness/compactness condition is required.

Assumption 3.2.1. *(i) $\mathbb{M} : L^p(Q) \to L^p(Q)$ is a bounded linear operator, and (ii) $\mathbb{M}_n$ is compact for some $n \geq 1$.*

It remains to verify the remaining conditions of Theorem 3.1.1 in this context. It is clear that $\mathbb{M} : L^p(Q) \to L^p(Q)$ is positive: the conditional expectation operator is a positive operator and m takes values in $[0, \infty)$.

The eventual strong positivity condition can be motivated in terms of a condition on the time-series behavior of the state process and a no-arbitrage condition.

Assumption 3.2.2. *For any $S \in \mathscr{X}$ with $Q(S) > 0$ there exists a $n_S \geq 1$ such that*

$$\Pr(X_{t+n_S} \in S | X_t = x) > 0$$

a.e.-$[Q]$.

Assumption 3.2.2 is similar to the irreducibility condition (in the sense of Markov processes) used in Hansen and Scheinkman (2009), though they require irreducibility under a change of measure induced by the permanent component. The identification conditions in Chapters 1 and 2 assume the stationary density of $\{X_t\}$ exists and is strictly positive, and the transition density of X_{t+1} given X_t exists and is strictly positive, which implies

Assumption 3.2.2. However, the transition density will fail to exist if the state process X_t is formed by rewriting a second- or higher-order Markov process as a first-order Markov process, in which case the extra generality of Assumption 3.2.2 is useful (see Section 3.2.1).

The principle of no-arbitrage asserts that any claim to a non-negative payoff which is positive with positive conditional probability must command a positive price (see, e.g., Hansen and Renault (2010) and references therein). It is helpful to think of the collection

$$\Psi = \{\psi \in L^p(Q) : \psi \geq 0 \text{ a.e.-}[Q], \psi \neq 0\} \tag{3.5}$$

as a menu of claims to future state-contingent payoffs. That is, at date t each pair $(\psi, n) \in \Psi \times \mathbb{N}$ represents a claim to $\psi(X_{t+n})$ units of the numeraire at time $t + n$.

Assumption 3.2.3. *The pricing system implied by (3.3) and (3.4) satisfies* no-arbitrage*, i.e., for each* $(\psi, n) \in \Psi \times \mathbb{N}$*,*

$$E[\psi(X_{t+n})|X_t = x] > 0 \quad \textit{implies} \quad \mathbb{M}_n\psi(x) > 0\,.$$

a.e.-$[Q]$.

Assumption 3.2.3 is similar to strict positivity of the continuous-time SDF process assumed by Hansen and Scheinkman (2009).

Lemma 3.2.1. $\mathbb{M}$ *satisfies eventual strong positivity under Assumptions 3.2.2 and 3.2.3.*

An identification result is presented first. An existence and identification result then follows under an additional condition on the yield on zero-coupon bonds.

Theorem 3.2.1. *Under Assumptions 3.2.1, 3.2.2 and 3.2.3, if* $\phi \in L^p(Q)$ *is a principal eigenfunction of* $\{\mathbb{M}_n : n \geq 1\}$ *then* ϕ *is the unique eigenfunction of* $\{\mathbb{M}_n : n \geq 1\}$ *in* $L^p(Q)$ *that is non-negative a.e.-*$[Q]$*. Moreover, its eigenvalue* ρ *is positive, has multiplicity one, and is the largest element in* $\sigma(\mathbb{M})$*.*

Theorem 3.2.1 shows that if one can be guaranteed of the existence of a principal eigenfunction of the family of operators $\{\mathbb{M}_n : n \geq 1\}$ then this principal eigenfunction must be unique. An extra condition on bond yields is sufficient for existence, in addition to uniqueness, of the principal eigenfunction.

The non-degeneracy condition may be formulated in terms of the yield on zero-coupon bonds. Let $1(\cdot) : \mathcal{X} \to \mathbb{R}$ denote the constant function, i.e. $1(x) = 1$ for all $x \in \mathcal{X}$. The price of a n-period zero-coupon bond is $\mathbb{M}_n 1(x)$. The yield-to-maturity $y_n(x)$ is defined implicitly by the relation

$$\mathbb{M}_n 1(x) = \exp(-n y_n(x))\,.$$

Assumption 3.2.4. *There exists $n \geq 1$ such that $y_n(x)$ is bounded uniformly a.e.-$[Q]$.*

For the C-CAPM with $m(X_t, X_{t+1}, Y_{t+1}) = \beta G_{t+1}^{-\gamma}$, where $G_{t+1} = g(X_t, X_{t+1}, Y_{t+1})$ and $\gamma \geq 1$, Assumption 3.2.4 is satisfied with $n = 1$ if expected consumption growth $E[g(X_t, X_{t+1}, Y_{t+1})|X_t = x]$ is bounded away from infinity a.e.-$[Q]$. Assumption 3.2.4 is also trivially satisfied (with $n = 1$) in models with a constant risk free rate.

Lemma 3.2.2. *$r(\mathbb{M}) > 0$ under Assumption 3.2.4.*

The following existence and identification result follows from Theorem 3.1.1.

Theorem 3.2.2. *Under Assumptions 3.2.1, 3.2.2, 3.2.3, and 3.2.4, there exists a $\phi \in L^p(Q)$ such that ϕ is a principal eigenfunction of $\{\mathbb{M}_n : n \geq 1\}$, and ϕ is the unique eigenfunction of $\{\mathbb{M}_n : n \geq 1\}$ in $L^p(Q)$ that is non-negative a.e.-$[Q]$. Moreover, its eigenvalue ρ is positive, has multiplicity one, and is the largest element in $\sigma(\mathbb{M})$.*

Remark 3.2.2. *Let $q = \infty$ if $p = 1$, otherwise let q be such that $q^{-1} + p^{-1} = 1$. Let $\mathbb{M}^* : L^q(Q) \to L^q(Q)$ denote the adjoint of $\mathbb{M}$ defined in (3.1). Under the conditions of Theorem 3.2.1 or 3.2.2, there exists an eigenfunction $\phi^* \in L^q(Q)$ of $\mathbb{M}^*$ corresponding to the eigenvalue ρ such that $\phi^* > 0$ a.e.-$[Q]$ and ϕ^* is the unique eigenfunction of $\mathbb{M}^*$ that is non-negative a.e.-$[Q]$.*

Remark 3.2.3. *Assumption 3.2.4 can be weakened to require only that there exists* $\psi \in \Psi$ *and* $n \in \mathbb{N}$ *such that the price of a claim to* $\psi(X_{t+n})$ *units of the numeraire at time* t *is bounded below by a constant multiple* $\delta > 0$ *of* $\psi(X_t)$*, a.e.-*$[Q]$*.*

Example: higher-order Markov processes

Higher-order Markov processes are popular modeling devices in the asset pricing literature. Examples include the ubiquitous vector autoregressive processes, as well as nonlinear processes such as autoregressive Gamma processes (Gourieroux and Jasiak, 2006) and compound autoregressive processes (Darolles, Gourieroux, and Jasiak, 2006). The fact that the Markov property is maintained while allowing for richer dynamics make these processes popular choices as state processes in dynamic asset pricing models. See, for example, Monfort and Pegoraro (2007), Bertholon, Monfort, and Pegoraro (2008) and Eraker (2008) for the use of these processes in modeling the term structure of interest rates.

Models with higher-order Markov state processes fall within the scope of the analysis of Hansen and Scheinkman (2009) by appropriately augmenting the state vector. Let $\{W_t\}_{t=-\infty}^{\infty}$ be a strictly stationary and ergodic Markov process of finite order $d > 1$. Define the augmented state vector $X_t = (W_t', W_{t-1}', \ldots, W_{t-d+1}')'$. Then $\{X_t\}$ is a strictly stationary and ergodic first-order Markov process.

In such higher-order Markov environments there are two issues which complicate identification of the principal eigenfunction using the identification conditions in Chen et al. (2014a) and Chapters 1 and 2. First, the transition density of X_{t+1} given X_t does not exist, since the $d-1$ elements $(W_t, \ldots, W_{t-d+2})$ of X_{t+1} are known when X_t is known. Second, in this setting the operator $\mathbb{M}$ may fail to be compact.

Lemma 3.2.3. *Let* W_t *have continuous support (so as to rule out a finite statespace). If there exists* $c > 0$ *such that* $E[m(X_t, X_{t+1}, Y_{t+1}) | X_t = x] \geq c$ *a.e.-*$[Q]$ *then* $\mathbb{M} : L^2(Q) \to L^2(Q)$ *is noncompact.*

When $\mathbb{M}$ is noncompact the integrability restriction in the identification conditions in

Chen et al. (2014a) and Chapter 1 will be violated (cf. Condition (b) in Section 3.1.3). Moreover, the conditions in the integral kernel representing the operator are less interpretable when the conditional density fails to exist.

Nevertheless, Theorems 3.2.1 and 3.2.2 may be used to study identification in higher-order environments. For example, if $\{W_t\}_{t=-\infty}^{\infty}$ is a VAR(d) process with i.i.d. $N(0,\Sigma)$ innovations (where Σ is finite and positive definite) then Assumption 3.2.2 will hold with $n = d$.

3.2.2 Identification in continuous-time environments

The conditions for nonparametric identification established for discrete-time environments have a natural extension to continuous-time environments, thereby providing an alternative set of identification conditions to those in Hansen and Scheinkman (2009).

Consider an environment characterized by a strictly stationary, continuous-time Markov process $\{Z_t\}_{t\in[0,\infty)}$ with support $\mathcal{Z} \subset \mathbb{R}^{d_z}$ where $\mathcal{Z}$ is a Borel set, and whose sample paths are right continuous with left limits. Following Hansen and Scheinkman (2009), consider a class of model in which for each $\tau \geq 0$ the price $\mathbb{M}_\tau\psi(Z_t)$ of a claim to $\psi(Z_{t+\tau})$ units of the numeraire at time $t+\tau$ is, at time t, given by

$$\mathbb{M}_\tau\psi(Z_t) = E\left[M_\tau(Z_t)\psi(Z_{t+\tau})|Z_t\right] \tag{3.6}$$

for each $\tau \in [0,\infty)$, where the SDF $M_\tau(Z_t)$ is a multiplicative functional of the sample path $\{Z_s : t \leq s \leq s+\tau\}$, i.e. $M_0(Z_t) = 1$ and $M_{\tau+\upsilon}(Z_t) = M_\tau(Z_{t+\upsilon})M_\upsilon(Z_t)$ for each $\tau,\upsilon \in [0,\infty)$. For example, if $\{Z_t\}$ is a diffusion process the functional

$$M_\tau(Z_t) = \exp\left(-\int_t^{t+\tau}\beta(Z_s)\,\mathrm{d}s\right)$$

is a multiplicative functional. This restriction on the SDF implies that the family of pricing operators $\{\mathbb{M}_\tau : \tau \in [0,\infty)\}$ factorizes as $\mathbb{M}_{\tau+\upsilon} = \mathbb{M}_\tau\mathbb{M}_\upsilon$ for $\tau,\upsilon \geq 0$ and that $\mathbb{M}_0 = I$, the identity operator.

Let Q denote the stationary distribution of Z on $\mathcal{Z}$ and let $\mathscr{Z}$ denote the Borel σ-algebra on $\mathcal{Z}$. The following conditions are sufficient for identification in $L^p(Q) := L^p(\mathcal{Z}, \mathscr{Z}, Q)$ for some $1 \leq p < \infty$. Extension of the following analysis to the space $L^\infty(Q)$ is straightforward.

Definition 3.2.2. *$\phi \in L^p(Q)$ is a* principal eigenfunction *of the family of pricing operators $\{\mathbb{M}_\tau : \tau \in [0, \infty)\}$ if $\phi > 0$ a.e.-$[Q]$ and $\mathbb{M}_\tau \phi = \rho^\tau \phi$ for each $\tau \geq 0$ and some $\rho \in \mathbb{R}$.*

Remark 3.2.4. *Taking $\tau = \tau^*$ in Definition 3.2.2 shows that ϕ is a principal eigenfunction of $\{\mathbb{M}_\tau : \tau \in [0, \infty)\}$ only if ϕ is a positive eigenfunction of $\mathbb{M}_{\tau^*}$ for each $\tau^* > 0$.*

Alternative identification conditions can therefore be provided by applying the discrete-time results to the "skeleton" of operators $\{\mathbb{M}_{n\tau^*} : n \geq 1\}$ for some fixed positive τ^*. Analysis of continuous-time Markov processes by means of their discretely-sampled skeleton is common practice. For instance, Hansen and Scheinkman (2009) cast their continuous-time identification conditions in terms of the skeleton of $\{Z_t\}_{t \in [0,\infty)}$ under a change of measure induced by the permanent component.

Assumption 3.2.5. *(i) $\mathbb{M}_{\tau^*} : L^p(Q) \to L^p(Q)$ is a bounded operator, and (ii) there exists $n \geq 1$ such that $\mathbb{M}_{n\tau^*}$ is compact.*

Assumption 3.2.6. *Assumptions 3.2.2 and 3.2.3 are satisfied with $\{Z_{n\tau^*}\}_{n=0}^\infty$ in place of $\{X_t\}_{t=0}^\infty$ and $\mathbb{M}_{\tau^*}$ in place of $\mathbb{M}$.*

Assumption 3.2.6 is similar to the assumptions of strong positivity of the SDF process and irreducibility of the state process in Hansen and Scheinkman (2009), but they do not assume compactness of the operator or stationarity of the state process. Instead, they assume the discretely sampled process $\{Z_{n\tau^*}\}$ satisfies additional stochastic stability conditions under a change of measure induced by the permanent component. In stationary environments, Assumptions 3.2.5 and 3.2.6 are sufficient to guarantee uniqueness of the positive eigenfunction and also show how ρ is related to the spectrum of the pricing operators.

Theorem 3.2.3. *Under Assumptions 3.2.5 and 3.2.6, if* $\phi \in L^p(Q)$ *is a principal eigenfunction of* $\{\mathbb{M}_\tau : \tau \in [0,\infty)\}$ *then* ϕ *is the unique eigenfunction of* $\{\mathbb{M}_\tau : \tau \in [0,\infty)\}$ *in* $L^p(Q)$ *that is non-negative a.e.-*$[Q]$*. Moreover, the eigenvalue* ρ *is positive, has multiplicity one, and* ρ^{τ^*} *is the largest element in* $\sigma(\mathbb{M}_{\tau^*})$*.*

Unlike the discrete-time case, proving existence of a positive eigenfunction of $\mathbb{M}_{\tau^*}$ is not enough to show existence of an eigenfunction of the whole family $\{\mathbb{M}_\tau : \tau \in [0,\infty)\}$ because this family is not characterized fully by $\mathbb{M}_{\tau^*}$. See Hansen and Scheinkman (2009) for existence conditions for continuous-time environments.

3.3 Positive operators on Banach lattices

The preceding identification theorems are based on an extension by Schaefer (1960) of the famous theorem of Kreĭn and Rutman (1950) to positive linear operators on Banach lattices. Identification of positive eigenfunctions of positive linear operators on Banach lattices is well studied in mathematics, as evidenced by the large body of work following Kreĭn and Rutman (1950). However, identification conditions are typically cast in terms of very high-level conditions, such as irreducibility of the operator and other properties of the resolvent of the operator. This section provides a particularization of these results to a set of sufficient conditions that are arguably more applicable in economics.

Let E be a Banach lattice and $T : E \to E$ be a bounded linear operator. Examples of Banach lattices of functions include the $L^p(\mu)$ spaces as dealt with in Section 3.1, the Banach spaces $C_b(\mathcal{X})$ of bounded continuous functions $f : \mathcal{X} \to \mathbb{R}$ on a (completely regular) Hausdorff space $\mathcal{X}$, and, if $\mathcal{X}$ is also compact, the space $C(\mathcal{X})$ of continuous functions $f : \mathcal{X} \to \mathbb{R}$.[7] Let E_+ and E_{++} denote the positive cone in E and its quasi-interior. Let T^* denote the adjoint of T, E^* denote the dual space of E, and E^*_+ and E^*_{++} denote the positive cone in E^* and its quasi-interior. Recall that T is said to be *positive* if $TE_+ \subseteq E_+$

[7]The terminology of a function space is used because the related identification problems in economics are typically in the context of operators on function spaces. More generally, there exist other Banach lattices that are not function spaces. The subsequent results apply equally to such spaces.

and *irreducible* if $TR(T,z) : (E_+ \setminus \{0\}) \to E_{++}$ for each $z \in (\text{spr}(T), \infty)$ where $R(T,z)$ is the resolvent of T (see Section 2.1).

Assumption 3.3.1. *(i) T is positive, (ii) $\text{spr}(T)$ is a pole of the resolvent of T, and (iii) T is irreducible.*

Theorem 3.3.1 (Schaefer (1960), Theorem 2; see also Schaefer (1999), p. 318)**.** *Under Assumption 3.3.1, $\text{spr}(T) > 0$, there exist eigenfunctions $\bar{f} \in E_{++}$ and $\bar{f}^* \in E_{++}^*$ of T and T^* corresponding to eigenvalue $\text{spr}(T)$, and $\text{spr}(T)$ has multiplicity one and is a first-order pole of the resolvent of T.*

Corollary 3.3.1. *Under Assumption 3.3.1, the eigenfunctions $\bar{f}$ and $\bar{f}^*$ are the unique eigenfunctions of T and T^* belonging to E_+ and E_+^*.*

The high-level Assumptions 3.3.1(ii) and 3.3.1(iii) may be difficult to motivate in an economic context. However it is possible to provide more tractable sufficient conditions for these assumptions. These sufficient conditions also yield a stronger result than Theorem 3.3.1 and Corollary 3.3.1 with regards to the separation of $\text{spr}(T)$ in the spectrum of T, which is useful in the econometric analysis of positive eigenfunctions of T.

Assumption 3.3.2. *(i) T is positive, (ii) T^n is compact for some $n \geq 1$, (iii) $\text{spr}(T) > 0$, and (iv) for each $f \in (E_+ \setminus \{0\})$ there exists $n \geq 1$ such that $T^n f \in E_{++}$.*

Assumption 3.3.2 is sufficient for existence and identification of the positive eigenfunctions of T and T^*, and shows that the eigenvalue with which these are associated is the largest element of $\sigma(T)$.

Theorem 3.3.2. *Under Assumption 3.3.2, there exist eigenfunctions $\bar{f} \in E_{++}$ and $\bar{f}^* \in E_{++}^*$ of T and T^* with eigenvalue $\text{spr}(T)$, and $\bar{f}$ and $\bar{f}^*$ are the unique eigenfunctions of T and T^* belonging to E_+ and E_+^*. Moreover, $\text{spr}(T)$ is an isolated point of $\sigma(T)$, has multiplicity one, and is the unique element in $\sigma(T)$ belonging to the circle $\{z \in \mathbb{C} : |z| = \text{spr}(T)\}$.*

3.4 Conclusion

This chapter provides sufficient conditions for the nonparametric identification and existence of the positive eigenfunction of a positive linear operator. Application to the pricing of long-horizon assets in discrete- and continuous-time dynamic asset pricing models is discussed.

3.5 Proofs

Proof of Theorem 3.1.1. Immediate by application of Theorem 3.3.2. □

Proof of Theorem 3.1.2. Immediate by application of Theorem 3.3.2. □

Proof of Lemma 3.2.1. Take any $\psi \in \Psi$. Let $N_\psi = \{x \in \mathcal{X} : \psi(x) > 0\}$ denote the support of ψ. Assumption 3.2.2 shows that $\Pr(X_{t+n} \in N_\psi | X_t = x) > 0$ a.e.-$[Q]$ for some $n \geq 1$, and so $E[\psi(X_{t+n})|X_t = x] > 0$ a.e.-$[Q]$. Assumption 3.2.3 then implies that $\mathbb{M}_n\psi(x) > 0$ a.e.-$[Q]$. □

Proof of Theorem 3.2.1. The proof is by application of Theorem 3.1.1. Assumption 3.2.1, 3.2.2 and 3.2.3 satisfy Assumptions 3.1.1(i), 3.1.1(ii) and 3.1.1(iv). For Assumption 3.1.1(iii), $\text{spr}(T) \geq \rho$ by definition and $\rho > 0$ since $\mathbb{M}_n\phi = \rho^n\phi > 0$ a.e.-$[Q]$ for some $n \in \mathbb{N}$ by Lemma 3.2.1. □

Proof of Lemma 3.2.2. Let $1(\cdot) : \mathcal{X} \to \mathbb{R}$ denote the constant function, i.e. $1(x) = 1$ for all $x \in \mathcal{X}$. Under Assumption 3.2.4, there exists $\epsilon > 0$ such that

$$\mathbb{M}^n 1(x) = \mathbb{M}_n 1(x) = \exp(-ny_n(x)) \geq \epsilon \times 1(x) \tag{3.7}$$

a.e.-$[Q]$. The result follows by applying Proposition 3 of Schaefer (1960). □

Proof of Theorem 3.2.2. Immediate by application of Theorem 3.1.1. □

Proof of Lemma 3.2.3. The proof is by contradiction. Let $\{\psi_i\}_{i\in\mathbb{N}}$ be a sequence of functions belonging to the ball $\{\psi \in L^2(Q) : \psi(x_t) = \psi(w_{t-1}, \dots, w_{t-d+1}), \|\psi\| = 1\}$ such that $\{\psi_i\}_{i\in\mathbb{N}}$ has no convergent subsequence (we can always choose such a sequence because this ball is not compact when W_t has continuous support).

Observe that $\mathbb{M}\psi_i(x_t) = \psi(w_t, \dots, w_{t-d+2})E[m(X_t, X_{t+1}, Y_{t+1})|X_t = x_t]$ for each ψ_i. Assume $\mathbb{M}$ is compact. Then $\{\mathbb{M}\psi_i\}_{i\in\mathbb{N}}$ has a convergent subsequence, say $\{\mathbb{M}\psi_{i_k}\}_{k\in\mathbb{N}}$. Moreover,

$$
\begin{aligned}
& \|\mathbb{M}\psi_{i_k} - \mathbb{M}\psi_{i_j}\| \\
= \; & \|(\psi_{i_k}(w_t, \dots, w_{t-d+2}) - \psi_{i_j}(w_t, \dots, w_{t-d+2}))E[m(X_t, X_{t+1}, Y_{t+1})|X_t = x_t]\| \\
\geq \; & c\|\psi_{i_k} - \psi_{i_j}\|.
\end{aligned}
$$

The subsequence $\{\mathbb{M}\psi_{i_k}\}_{k\in\mathbb{N}}$ is convergent and therefore Cauchy, and so the subsequence $\{\psi_{i_k}\}_{k\in\mathbb{N}}$ is Cauchy by the above inequality. Therefore the subsequence $\{\psi_{i_k}\}_{k\in\mathbb{N}}$ must converge (by completeness of $L^2(Q)$), which contradicts the fact that $\{\psi_i\}_{i\in\mathbb{N}}$ has no convergent subsequence. □

Proof of Theorem 3.2.3. Apply Theorem 3.2.1 with $\mathbb{M}_{\tau^*}$ in place of $\mathbb{M}$. □

Proof of Corollary 3.3.1. The eigenvalue $\operatorname{spr}(T)$ has multiplicity one by Theorem 3.3.1, so any other eigenfunction (i.e. one that is not a constant multiple of $\bar{f}$) must correspond to an eigenvalue different from $\operatorname{spr}(T)$. Suppose $\bar{g} \in (E_+ \setminus \{0\})$ be an eigenfunction of T corresponding to $\lambda \neq \operatorname{spr}(T)$. Then, with $\langle f, f^* \rangle := f^*(f)$ for $f^* \in E^*$, $f \in E$,

$$
\lambda\langle \bar{g}, \bar{f}^* \rangle = \langle T\bar{g}, \bar{f}^* \rangle = (\bar{f}^* \circ T)(\bar{g}) = T^*(\bar{f}^*)(\bar{g}) = \operatorname{spr}(T)\bar{f}^*(\bar{g}) = \operatorname{spr}(T)\langle \bar{g}, \bar{f}^* \rangle \qquad (3.8)
$$

which is a contradiction, since $\bar{f}^* \in E^*_{++}$ and $g \in (E_+ \setminus \{0\})$ implies $\langle \bar{g}, \bar{f}^* \rangle > 0$.

As the multiplicities of eigenvalues of finite multiplicity of T are preserved under adjoints (Dunford and Schwartz, 1958, Exercise VII.5.35), $\operatorname{spr}(T)$ is an eigenvalue of T^* with multiplicity one. That $\bar{f}^* \in E^*_{++}$ is the unique eigenfunction of T^* in E^*_+ follows by a similar argument. $\square$

Proof of Theorem 3.3.2. Assumption 3.3.1(ii) is satisfied under Assumptions 3.3.2(i)–(iii) because $\mathrm{spr}(T) \in \sigma(T)$ since T is positive (Schaefer, 1999, p. 312) and each nonzero element of $\sigma(T)$ is a pole of the resolvent of T since T^n is compact (Dunford and Schwartz, 1958, p. 579). Assumption 3.3.2(iv) implies that for each $f \in (E_+ \setminus \{0\})$ and $f^* \in E^*_+$ there exists $n \geq 1$ such that $\langle T^n f, f^* \rangle > 0$, so T is irreducible (Schaefer, 1974, Proposition III.8.3). The first part then follows immediately from Theorem 3.3.1 and Corollary 3.3.1.

For the second part, compactness of T^n for some $n \geq 1$ and the fact that $\sigma(T)^n = \sigma(T^n)$ implies that the only possible limit point of elements of $\sigma(T)$ is zero (Dunford and Schwartz, 1958, p. 579). Therefore $\mathrm{spr}(T)$ is an isolated point of $\sigma(T)$. That $\mathrm{spr}(T)$ has multiplicity one is immediate from Theorem 3.3.1. Finally, that $\mathrm{spr}(T)$ is the unique point in $\sigma(T)$ belonging to the circle $\{z \in \mathbb{C} : |z| = \mathrm{spr}(T)\}$ follows under Assumption 3.3.2(iv) by application of Proposition V.5.6 of Schaefer (1974). $\square$

Chapter 4

Optimal Uniform Convergence Rates for Sieve Nonparametric Instrumental Variables Regression[1]

In economics and other social sciences one frequently encounters the relation

$$Y_{1i} = h_0(Y_{2i}) + \epsilon_i \tag{4.1}$$

where Y_{1i} is a response variable, Y_{2i} is a predictor variable, h_0 is an unknown structural function of interest, and ϵ_i is an error term. However, a latent external mechanism may "determine" or "cause" Y_{1i} and Y_{2i} simultaneously, in which case the conditional mean restriction $E[\epsilon_i|Y_{2i}] = 0$ fails and Y_{2i} is said to be *endogenous*.[2] When the regressor Y_{2i} is endogenous one cannot use standard nonparametric regression techniques to consistently estimate h_0. In this instance one typically assumes that there exists a vector of *instrumental variables* X_i such that $E[\epsilon_i|X_i] = 0$ and for which there is a nondegenerate relationship

[1]This chapter is joint work with Xiaohong Chen.

[2]In a canonical example of this relation, Y_{1i} may be the hourly wage of person i and Y_{2i} may include the education level of person i. The latent ability of person i affects both Y_{1i} and Y_{2i}. See Blundell and Powell (2003) for other examples and discussions of endogeneity in semi/nonparametric regression models.

between X_i and Y_{2i}. Such a setting permits estimation of h_0 using nonparametric instrumental variables (NPIV) techniques based on a sample $\{(X_i, Y_{1i}, Y_{2i})\}_{i=1}^n$. In this chapter the data are strictly stationary in that (X_i, Y_{1i}, Y_{2i}) has the same (unknown) distribution F_{X,Y_1,Y_2} as that of (X, Y_1, Y_2) for all i.[3]

NPIV estimation has been the subject of much research in recent years, both because of its practical importance to applied economics and its prominent role in the literature on linear ill-posed inverse problems with unknown operators. In many economic applications the joint distribution F_{X,Y_2} of X_i and Y_{2i} is unknown but is assumed to have a continuous density. Therefore the conditional expectation operator $Th(\cdot) = E[h(Y_{2i})|X_i = \cdot]$ is typically unknown but compact. Model (4.1) with $E[\epsilon_i|X_i] = 0$ can be equivalently written as

$$\begin{aligned} Y_{1i} &= Th_0(X_i) + u_i \\ E[u_i|X_i] &= 0 \end{aligned} \tag{4.2}$$

where $u_i = h_0(Y_{2i}) - Th_0(X_i) + \epsilon_i$. Model (4.2) is called the reduced-form NPIV model if T is assumed to be unknown and the nonparametric indirect regression (NPIR) model if T is assumed to be known. Let $\widehat{E}[Y_1|X = \cdot]$ be a consistent estimator of $E[Y_1|X = \cdot]$. Regardless of whether the compact operator T is unknown or known, nonparametric recovery of h_0 by inversion of the conditional expectation operator T on the left-hand side of the Fredholm equation of the first kind

$$Th(\cdot) = \widehat{E}[Y_1|X = \cdot] \tag{4.3}$$

leads to an ill-posed inverse problem (see, e.g., Kress (1999)). Consequently, some form of regularization is required for consistent nonparametric estimation of h_0. In the literature there are several popular methods of NPIV estimation, including but not limited to (1) finite-dimensional sieve minimum distance estimators (Newey and Powell, 2003; Ai and Chen, 2003; Blundell, Chen, and Kristensen, 2007); (2) kernel-based Tikhonov regularization estimators (Hall and Horowtiz, 2005; Darolles, Fan, Florens, and Renault, 2011;

[3]The subscript i denotes either the individual i in a cross-sectional sample or the time period i in a time-series sample. Since the sample is strictly stationary the subscript i is sometimes omitted.

Gagliardini and Scaillet, 2012) and their Bayesian version (Florens and Simoni, 2012); (3) orthogonal series Tikhonov regularization estimators (Hall and Horowtiz, 2005); (4) orthogonal series Galerkin-type estimators (Horowitz, 2011); (5) general penalized sieve minimum distance estimators (Chen and Pouzo, 2012) and their Bayesian version (Liao and Jiang, 2011). See Horowitz (2011) for a recent review and additional references.

Existing work on convergence rates for various NPIV estimators has only studied L^2-norm convergence rates. In particular, Hall and Horowtiz (2005) are the first to establish the minimax risk lower bound in L^2-norm loss for a class of mildly ill-posed NPIV models, and show that their estimators attain the lower bound. Chen and Reiss (2011) derive the minimax risk lower bound in L^2-norm loss for a large class of NPIV models that could be mildly or severely ill-posed, and show that the sieve minimum distance estimator of Blundell, Chen, and Kristensen (2007) achieves the lower bound. Subsequently, some other NPIV estimators listed above have also been shown to achieve the optimal L^2-norm convergence rates. As yet there are no published results on sup-norm (uniform) convergence rates for any NPIV estimators, nor results on what are the minimax risk lower bounds in sup-norm loss for any class of NPIV models.

Sup-norm convergence rates for any estimators of h_0 are important for constructing uniform confidence bands for the unknown h_0 in NPIV models and for conducting inference on nonlinear functionals of h_0, but are currently missing. This chapter examines the uniform convergence properties of the sieve minimum distance estimator of h_0 for the NPIV model, which is a nonparametric series two-stage least squares regression estimator (Newey and Powell, 2003; Ai and Chen, 2003; Blundell, Chen, and Kristensen, 2007). Sieve NPIV estimators are easy to compute and have been used in empirical work in demand analysis (Blundell, Chen, and Kristensen, 2007; Chen and Pouzo, 2009), asset pricing (Chen and Ludvigson, 2009), and other applied fields in economics. Also, this class of estimators is known to achieve the optimal L^2-norm convergence rates for both mildly and severely ill-posed NPIV models.

We first establish a general upper bound (Theorem 4.1.1) on the uniform convergence rate of a sieve estimator, allowing for endogenous regressors and weakly dependent data. To provide sharp bounds on the sieve approximation error or "bias term" the proof strategy of Huang (2003) is extended from sieve nonparametric least squares (LS) regression to the sieve NPIV estimator. Together, these tools yield sup-norm convergence rates for the spline and wavelet sieve NPIV estimators under i.i.d. data. Under conditions similar to those for the L^2-norm convergence rates for the sieve NPIV estimators, the sup-norm convergence rates obtained in this chapter coincide with the known optimal L^2-norm rates for severely ill-posed problems, and are power of $\log(n)$ slower than the optimal L^2-norm rates for mildly ill-posed problems. The minimax risk lower bound in sup-norm loss for h_0 in a NPIR model (i.e., (4.2) with a known compact T) uniformly over Hölder balls is established, which in turn provides a lower bound in sup-norm loss for h_0 in a NPIV model uniformly over Hölder balls. The lower bound is shown to coincide with the sup-norm convergence rates for the spline and wavelet sieve NPIV estimators obtained.

To establish the general upper bound, a new exponential inequality for sums of weakly dependent random matrices is derived in Section 4.4. This allows for a weakening of the conditions under which the optimal uniform convergence rates can be obtained. As an indication of the sharpness of the general upper bound result, it is shown that it recovers the optimal uniform convergence rates for spline and wavelet LS regression estimators with weakly dependent data and heavy-tailed error terms. Precisely, for beta-mixing dependent data and finite $(2+\delta)$-th moment error term (for $\delta \in (0,2)$), spline and wavelet nonparametric LS regression estimators are shown to attain the minimax risk lower bound in sup-norm loss of Stone (1982). This result should be very useful to the literature on nonparametric estimation with financial time series.

The NPIV model falls within the class of statistical linear ill-posed inverse problems with *unknown* operators and additive noise. There is a vast literature on statistical linear ill-posed inverse problems with known operators and additive noise. Some recent references include but are not limited to Cavalier, Golubev, Picard, and Tsybakov (2002), Cohen, Hoffmann, and Reiss (2004) and Cavalier (2008), of which density deconvolution is an important and extensively-studied problem (see, e.g., Carroll and Hall (1988); Zhang (1990); Fan (1991); Hall and Meister (2007); Lounici and Nickl (2011)). There are also papers on statistical linear ill-posed inverse problems with pseudo-unknown operators (i.e., known eigenfunctions but unknown singular values) (see, e.g., Cavalier and Hengartner (2005), Loubes and Marteau (2012)). Related papers that allow for an unknown linear operator but assume the existence of an estimator of the operator (with rate) include Efromovich and Koltchinskii (2001), Hoffmann and Reiss (2008) and others. Most of the published work in the statistical literature on linear ill-posed inverse problems also focuses on the rate optimality in L^2-norm loss, except that of Lounici and Nickl (2011) which recently establishes the optimal sup-norm convergence rate for a wavelet density deconvolution estimator. Therefore, the results in this chapter on minimax risk lower bounds in sup-norm loss for the NPIR and NPIV models also contribute to the large literature on statistical ill-posed inverse problems.

The rest of the chapter is organized as follows. Section 4.1 outlines the model and presents a general upper bound on the uniform convergence rates for a sieve estimator. Section 4.2 establishes the optimal uniform convergence rates for the sieve NPIV estimators, allowing for both mildly and severely ill-posed inverse problems. Section 4.3 derives the optimal uniform convergence rates for the sieve nonparametric (least squares) regression, allowing for dependent data. Section 4.4 provides useful exponential inequalities for sums of random matrices, and the reinterpretation of equivalence of the theoretical and empirical L^2 norms as a criterion regarding convergence of a random matrix. 4.5 contains a brief review of the spline and wavelet sieve spaces.

Notation: $\|\cdot\|_2$ denotes the Euclidean norm when applied to vectors and the matrix spectral norm (largest singular value) when applied to matrices. If A is a square matrix, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote its smallest and largest eigenvalues, respectively, and A^- denotes its Moore-Penrose generalized inverse. If $\{a_n : n \geq 1\}$ and $\{b_n : n \geq 1\}$ are two sequences of non-negative numbers, $a_n \lesssim b_n$ means there exists a finite positive C such that $a_n \leq Cb_n$ for all n sufficiently large, and $a_n \asymp b_n$ means $a_n \lesssim b_n$ and $b_n \lesssim a_n$. $\#(\mathcal{S})$ denotes the cardinality of a set $\mathcal{S}$ of finitely many elements. Let $\mathrm{BSpl}(K, [0,1]^d, \gamma)$ and $\mathrm{Wav}(K, [0,1]^d, \gamma)$ denote tensor-product B-spline (with smoothness γ) and wavelet (with regularity γ) sieve spaces of dimension K on $[0,1]^d$ (see Section 4.5 for details on construction of these spaces).

The notation used for function spaces is different from that used in the preceding chapters. Here spaces of functions of the endogenous and exogenous variables are of interest. The L^p spaces in this chapter are labelled according to the variables with which the spaces are associated. That is, for a random variable Z let $L^q(Z)$ denote the spaces of (equivalence classes of) measurable functions of z with finite q-th moment if $1 \leq q < \infty$ and let $\| \cdot \|_{L^q(Z)}$ denote the $L^q(Z)$ norm. With some abuse of notation, let $L^\infty(Z)$ denote the space of measurable functions of z with finite sup norm $\| \cdot \|_\infty$.

4.1 Uniform convergence rates for sieve NPIV estimators

We begin by considering the NPIV model

$$\begin{aligned} Y_{1i} &= h_0(Y_{2i}) + \epsilon_i \\ E[\epsilon_i | X_i] &= 0 \end{aligned} \tag{4.4}$$

where $Y_1 \in \mathbb{R}$ is a response variable, Y_2 is an endogenous regressor with support $\mathcal{Y}_2 \subset \mathbb{R}^d$ and X is a vector of conditioning variables (also called instruments) with support $\mathcal{X} \subset \mathbb{R}^{d_x}$. The object of interest is the unknown structural function $h_0 : \mathcal{Y}_2 \to \mathbb{R}$ which belongs to some infinite-dimensional parameter space $\mathcal{H} \subset L^2(Y_2)$. It is assumed hereafter that h_0 is identified uniquely by the conditional moment restriction (4.4). See Newey and Powell (2003), Blundell, Chen, and Kristensen (2007), Darolles, Fan, Florens, and Renault (2011), Andrews (2011), D'Haultfoeuille (2011), Chen, Chernozhukov, Lee, and Newey (2014a) and references therein for sufficient conditions for identification.

4.1.1 Sieve NPIV estimators

The sieve NPIV estimator due to Newey and Powell (2003), Ai and Chen (2003), and Blundell, Chen, and Kristensen (2007) is a nonparametric series two-stage least squares estimator. Let the sieve spaces $\{\Psi_J : J \geq 1\} \subseteq L^2(Y_2)$ and $\{B_K : K \geq 1\} \subset L^2(X)$ be sequences of subspaces of dimension J and K spanned by sieve basis functions such that Ψ_J and B_K become dense in $\mathcal{H} \subset L^2(Y_2)$ and $L^2(X)$ as $J, K \to \infty$. For given J and K, let $\{\psi_{J1}, \ldots, \psi_{JJ}\}$ and $\{b_{K1}, \ldots, b_{KK}\}$ be sets of sieve basis functions whose closed linear span generates Ψ_J and B_K respectively.

In the first stage, the conditional moment function $m(x, h) : \mathcal{X} \times \mathcal{H} \to \mathbb{R}$ given by

$$m(x,h) = E[Y_1 - h(Y_2)|X = x]$$

is estimated using the series (least squares) regression estimator

$$\widehat{m}(x,h) = \sum_{i=1}^{n} b^K(x)'(B'B)^- b^K(X_i)(Y_{1i} - h(Y_{2i})) \tag{4.5}$$

where

$$\begin{aligned} b^K(x) &= (b_{K1}(x), \ldots, b_{KK}(x))' \\ B &= (b^K(X_1), \ldots, b^K(X_n))'. \end{aligned}$$

The sieve NPIV estimator $\widehat{h}$ is then defined as the solution to the second-stage minimization problem

$$\widehat{h} = \arg\min_{h \in \Psi_J} \frac{1}{n} \sum_{i=1}^{n} \widehat{m}(X_i, h)^2$$

which may be solved in closed form to give

$$\widehat{h}(y_2) = \psi^J(y_2)'[\Psi' B(B'B)^- B'\Psi]^- \Psi' B(B'B)^- B'Y$$

where

$$\begin{aligned}
\psi^J(y_2) &= (\psi_{J1}(y_2), \ldots, \psi_{JJ}(y_2))' \\
\Psi &= (\psi^J(Y_{21}), \ldots, \psi^J(Y_{2n}))' \\
Y &= (Y_{11}, \ldots, Y_{1n})' .
\end{aligned}$$

Under mild regularity conditions (see Newey and Powell (2003), Blundell, Chen, and Kristensen (2007) and Chen and Pouzo (2012)), $\widehat{h}$ is a consistent estimator of h_0 (in both $\|\cdot\|_{L^2(Y_2)}$ and $\|\cdot\|_\infty$ norms) as $n, J, K \to \infty$, provided $J \leq K$ and J increases appropriately slowly so as to regularize the ill-posed inverse problem.[4] The modified sieve estimator (or orthogonal series Galerkin-type estimator) of Horowitz (2011) corresponds to the sieve NPIV estimator with $J = K$ and $\psi^J(\cdot) = b^K(\cdot)$ being orthonormal basis in $L^2(\mathit{Lebesgue})$.

4.1.2 A general upper bound on uniform convergence rates

We first present a general calculation for sup-norm convergence which will be used to obtain uniform convergence rates for both the sieve NPIV and the sieve LS estimators below.

The sieve estimators are invariant to an invertible transformation of the sieve basis functions. Therefore the sieve spaces B_K and Ψ_J are renormalized so that $\{\widetilde{b}_{K1}, \ldots, \widetilde{b}_{KK}\}$ and $\{\widetilde{\psi}_{J1}, \ldots, \widetilde{\psi}_{JJ}\}$ form orthonormal bases for B_K and Ψ_J. This is achieved by setting $\widetilde{b}^K(x) = E[b^K(X)b^K(X)']^{-1/2}b^K(x)$ where $^{-1/2}$ denotes the inverse of the positive-definite matrix square root (which exists under Assumption 4.1.4(ii) below), with $\widetilde{\psi}^J$ similarly defined. Let

$$\begin{aligned}
\widetilde{B} &= (\widetilde{b}^K(X_1), \ldots, \widetilde{b}^K(X_n))' \\
\widetilde{\Psi} &= (\widetilde{\psi}^J(Y_{21}), \ldots, \widetilde{\psi}^J(Y_{2n}))'
\end{aligned}$$

[4] Here K denotes the "smoothing parameter" (i.e. the dimension of the sieve space used to estimate the conditional moments in (4.5)) and J to denote the "regularization parameter" (i.e. the dimension of the sieve space used to approximate the unknown h_0). Note that Chen and Reiss (2011) use J and m, Blundell, Chen, and Kristensen (2007) and Chen and Pouzo (2012) use J and k to denote the smoothing and regularization parameters, respectively.

and define the $J \times K$ matrices

$$\begin{aligned} S &= E[\widetilde{\psi}^J(Y_2)\widetilde{b}^K(X)'] \\ \widehat{S} &= \widetilde{\Psi}'\widetilde{B}/n\,. \end{aligned} \tag{4.6}$$

Let $\sigma_{JK}^2 = \lambda_{\min}(SS')$. For each $h \in \Psi_J$ define

$$\Pi_K Th(\cdot) = \widetilde{b}^K(x)'E[\widetilde{b}^K(X)(Th)(X)] = \widetilde{b}^K(\cdot)'E[\widetilde{b}^K(X)h(Y_2)] \tag{4.7}$$

which is the $L^2(X)$ orthogonal projection of $Th(\cdot)$ onto B_K. The variational characterization of singular values gives

$$\sigma_{JK} = \inf_{h \in \Psi_J : \|h\|_{L^2(Y_2)}=1} \|\Pi_K Th\|_{L^2(X)} \leq 1\,.$$

Finally, define P_n as the second-stage empirical projection operator onto the sieve space Ψ_J after projecting onto the instrument space B_K, viz.

$$P_n h_0(y_2) = \widetilde{\psi}^J(y_2)[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-}\widehat{S}']^{-}\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-}\widetilde{B}'H_0/n \tag{4.8}$$

where $H_0 = (h_0(Y_{21}), \ldots, h_0(Y_{2n}))'$.

We first decompose the sup-norm error as

$$\|h_0 - \widehat{h}\|_\infty \leq \|h_0 - P_n h_0\|_\infty + \|P_n h_0 - \widehat{h}\|_\infty$$

and calculate the uniform convergence rate for the "variance term" $\|\widehat{h} - P_n h_0\|_\infty$ in this section. Control of the "bias term" $\|h_0 - P_n h_0\|_\infty$ is left to the subsequent sections, which will be dealt with under additional regularity conditions for the NPIV model and the LS regression model separately.

Let $Z_i = (X_i, Y_{1i}, Y_{2i})$ and $\mathcal{F}_{i-1} = \sigma(X_i, X_{i-1}, \epsilon_{i-1}, X_{i-2}, \epsilon_{i-2}, \ldots)$.

Assumption 4.1.1. *(i) $\{Z_i\}_{i=-\infty}^{\infty}$ is strictly stationary and ergodic, (ii) X has support*

$\mathcal{X} = [0,1]^d$ *and* Y_2 *has support* $\mathcal{Y}_2 = [0,1]^d$, *(iii) the distributions of* X *and* Y_2 *have density (with respect to Lebesgue measure) which is uniformly bounded away from zero and infinity over* $\mathcal{X}$ *and* $\mathcal{Y}_2$ *respectively.*

The results stated in this section do not actually require that $\dim(X) = \dim(Y_2)$. However, most published papers on NPIV models assume $\dim(X) = \dim(Y_2) = d$. This convention is followed in Assumption 1(ii) and the remainder of the chapter.

Assumption 4.1.2. *(i)* $(\epsilon_i, \mathcal{F}_{i-1})_{i=-\infty}^{\infty}$ *is a strictly stationary martingale difference sequence, (ii) the conditional second moment* $E[\epsilon_i^2|\mathcal{F}_{i-1}]$ *is uniformly bounded away from zero and infinity, (iii)* $E[|\epsilon_i|^{2+\delta}] < \infty$ *for some* $\delta > 0$.

Assumption 4.1.3. *(i) Sieve basis* $\psi^J(\cdot)$ *is Hölder continuous with smoothness* $\gamma > p$ *and* $\sup_{y_2 \in \mathcal{Y}_2} \|\psi^J(y_2)\|_2 \lesssim \sqrt{J}$, *(ii)* $\lambda_{\min}(E[\psi^J(Y_2)\psi^J(Y_2)']) \geq \underline{\lambda} > 0$ *for all* $J \geq 1$.

In what follows, $p > 0$ indicates the smoothness of the function $h_0(\cdot)$ (see Assumption 4.2.1 in Section 4.2).

Assumption 4.1.4. *(i) Sieve basis* $b^K(\cdot)$ *is Hölder continuous with smoothness* $\gamma_x \geq \gamma > p$ *and* $\sup_{x \in \mathcal{X}} \|_2 b^K(x)\|_2 \lesssim \sqrt{K}$, *(ii)* $\lambda_{\min}(E[b^K(X)b^K(X)']) \geq \underline{\lambda} > 0$ *for all* $K \geq 1$.

The preceding assumptions on the data generating process allow for quite general weakly-dependent data but also nest i.i.d. sequences. In an i.i.d. setting, Assumption 4.1.2(i) is satisfied under an arbitrary ordering of the data because the ith innovation is orthogonal to the jth instrument and jth innovation for all $j \neq i$ by virtue of independence of (X_i, ϵ_i) and (X_j, ϵ_j) for all $j \neq i$; moreover, Assumption 4.1.2(ii) reduces to requiring that $E[\epsilon_i^2|X_i = x]$ be bounded uniformly from zero and infinity which is standard (see, e.g., Newey (1997); Hall and Horowtiz (2005)). The value of δ in Assumption 4.1.2(iii) depends on the context. For example, $\delta \geq d/p$ will be shown to be sufficient to attain the optimal sup-norm convergence rates for series LS regression in Section 4.3, whereas lower values of δ suffice to attain the optimal sup-norm convergence rates for the sieve NPIV estimator in Section 4.2. Rectangular support and bounded densities of the endogenous regressor and instrument are assumed in Hall and Horowtiz (2005). Assumptions 4.1.3(i) and 4.1.4(i) are satisfied by many widely used sieve bases such as spline, wavelet and cosine sieves, but they rule out polynomial and power series sieves (see, e.g., Newey (1997); Huang (1998)). The instruments sieve basis $b^K(\cdot)$ is used to approximate the conditional expectation operator $Th = E[h(Y_2)|X = \cdot)$, which is a smoothing operator. Thus Assumption 4.1.4(i) assumes that the sieve basis $b^K(\cdot)$ (for Th) is smoother than that of the sieve basis $\psi^J(\cdot)$ (for h).

In the next theorem, the upper bound on the "variance term" $\|\widehat{h} - P_n h_0\|_\infty$ holds under general weak dependence as captured by Condition (ii) on the convergence of the random matrices $\widetilde{B}'\widetilde{B}/n - I_K$ and $\widehat{S} - S$.

Theorem 4.1.1. *Let Assumptions 4.1.1, 4.1.2, 4.1.3 and 4.1.4 hold. If* $\sigma_{JK} > 0$ *then:*

$$\|h_0 - \widehat{h}\|_\infty \leq \|h_0 - P_n h_0\|_\infty + O_p\left(\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right)$$

provided $n, J, K \to \infty$ *and*

(i) $J \leq K$, $K \lesssim (n/\log n)^{\delta/(2+\delta)}$, *and* $\sigma_{JK}^{-1}\sqrt{K(\log n)/n} \lesssim 1$

(ii) $\sigma_{JK}^{-1}\left(\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 + \|\widehat{S} - S\|\right) = O_p(\sqrt{(\log n)/K}) = o_p(1)$.

The restrictions on J, K and n in Conditions (i) and (ii) merit a brief explanation. The restriction $J \leq K$ merely ensures that the sieve NPIV estimator is well defined. The restriction $K \lesssim (n/\log n)^{\delta/(2+\delta)}$ is used to perform a truncation argument using the existence of $(2+\delta)$-th moment of the error terms (see Assumption 4.1.2). Condition (ii) ensures that J increases sufficiently slowly that with probability approaching one the minimum eigenvalue of the "denominator" matrix $\Psi' B(B'B)^- B'\Psi/n$ is positive and bounded below by a multiple of σ_{JK}^2, thereby regularizing the ill-posed inverse problem. It also ensures the error in estimating the matrices $(\widetilde{B}'\widetilde{B}/n)$ and $\widehat{S}$ vanishes sufficiently quickly that it doesn't affect the convergence rate of the estimator.

Remark 4.1.1. *Section 4.4 provides very mild low-level sufficient conditions for Condition (ii) to hold under weakly dependent data. In particular, when specializing Corollary 4.4.1* *to i.i.d. data* $\{(X_i, Y_{2i})\}_{i=1}^n$ *(also see Lemma 4.4.2), under Assumptions 4.1.3 and 4.1.4 and* $J \leq K$*:*

$$\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 = O_p(\sqrt{K(\log K)/n}), \quad \|\widehat{S} - S\|_2 = O_p(\sqrt{K(\log K)/n}).$$

4.2 Optimal uniform convergence rates for sieve NPIV estimators

4.2.1 Upper bounds on uniform convergence rates

We now exploit the specific linear structure of the sieve NPIV estimator to derive uniform convergence rates for the mildly and severely ill-posed cases. Some additional assumptions are required so as to control the "bias term" $\|h_0 - P_n h_0\|_\infty$ and to relate the estimator to the measure of ill-posedness.

We first impose a standard smoothness condition on the unknown structural function h_0 to facilitate comparison with Stone (1982)'s minimax risk lower bound in sup-norm loss for a nonparametric regression function. Recall that $\mathcal{Y}_2 = [0,1]^d$. Deferring definitions to Triebel (2006, 2008), let $B^p_{q,q}([0,1]^d)$ denote the Besov space of smoothness p on the domain $[0,1]^d$ and $\|\cdot\|_{B^p_{q,q}}$ denote the usual Besov norm on this space. Special cases include the Sobolev class of smoothness p, namely $B^p_{2,2}([0,1]^d)$, and the Hölder-Zygmund class of smoothness p, namely $B^p_{\infty,\infty}([0,1]^d)$. Let $B(p,L)$ denote a Hölder ball of smoothness p and radius $0 < L < \infty$, i.e. $B(p,L) = \{h \in B^p_{\infty,\infty}([0,1]^d) : \|h\|_{B^p_{\infty,\infty}} \leq L\}$.

Assumption 4.2.1. $h_0 \in \mathcal{H} = B^p_{\infty,\infty}([0,1]^d)$ *for some* $p \geq d/2$*.*

Assumptions 4.1.3 and 4.2.1 imply that there is $\pi_J h_0 \in \Psi_J$ such that $\|h_0 - \pi_J h_0\|_\infty = O(J^{-p/d})$.

Let $T : L^q(Y_2) \to L^q(X)$ denote the conditional expectation operator for $1 \leq q \leq \infty$:

$$Th(x) = E[h(Y_{2i})|X_i = x]\,.$$

When Y_2 is endogenous, T is compact under mild conditions on the conditional density of Y_2 given X. For $q' \geq q \geq 1$, define the measure of ill-posedness (over a sieve space Ψ_J) as

$$\tau_{q,q',J} = \sup_{h \in \Psi_J : \|Th\|_{L^q(X)} \neq 0} \frac{\|h\|_{L^{q'}(Y_2)}}{\|Th\|_{L^q(X)}}.$$

The $\tau_{2,2,J}$ measure of ill-posedness is clearly related to the earlier definition of σ_{JK}. Indeed,

$$\sigma_{JK} = \inf_{h \in \Psi_J : \|h\|_{L^2(Y_2)}=1} \|\Pi_K Th\|_{L^2(X)} \leq \inf_{h \in \Psi_J : \|h\|_{L^2(Y_2)}=1} \|Th\|_{L^2(X)} = (\tau_{2,2,J})^{-1}$$

when $J \leq K$. The sieve measures of ill-posedness, $\tau_{2,2,J}$ and σ_{JK}^{-1}, are clearly non-decreasing in J. In Blundell, Chen, and Kristensen (2007), Horowitz (2011) and Chen and Pouzo (2012), the NPIV model is said to be

- *mildly ill-posed* if $\tau_{2,2,J} = O(J^{\varsigma/d})$ for some $\varsigma > 0$;
- *severely ill-posed* if $\tau_{2,2,J} = O(\exp(\frac{1}{2}J^{\varsigma/d}))$ for some $\varsigma > 0$.

These measures of ill-posedness are not exactly the same as (but are related to) the measure of ill-posedness used in Hall and Horowtiz (2005) and Cavalier (2008). In the latter papers, it is assumed that the compact operator $T : L^2(Y_2) \to L^2(X)$ admits a singular value decomposition $\{\mu_k; \phi_{1k}, \phi_{0k}\}_{k=1}^{\infty}$, where $\{\mu_k\}_{k=1}^{\infty}$ are the singular numbers arranged in non-increasing order ($\mu_k \geq \mu_{k+1} \searrow 0$), $\{\phi_{1k}(y_2)\}_{k=1}^{\infty}$ and $\{\phi_{0k}(x)\}_{k=1}^{\infty}$ are eigenfunction (orthonormal) bases for $L^2(Y_2)$ and $L^2(X)$ respectively, and ill-posedness is measured in terms of the rate of decay of the singular values towards zero. Denote T^* as the adjoint operator of T: $\{T^* g\}(Y_2) \equiv E[g(X)|Y_2]$, which maps $L^2(X)$ into $L^2(Y_2)$. Then a compact T implies that T^*, T^*T and TT^* are also compact, and that $T\phi_{1k} = \mu_k \phi_{0k}$ and $T^*\phi_{0k} = \mu_k \phi_{1k}$ for all k. Note that $\|Th\|_{L^2(X)} = \|(T^*T)^{1/2}h\|_{L^2(Y_2)}$ for all $h \in Dom(T)$. The following lemma provides some relations between these different measures of ill-posedness.

Lemma 4.2.1. *Let the conditional expectation operator $T : L^2(Y_2) \to L^2(X)$ be compact and injective. Then: (1) $\sigma_{JK}^{-1} \geq \tau_{2,2,J} \geq 1/\mu_J$; (2) If the sieve space Ψ_J spans the closed* *linear subspace (in $L^2(Y_2)$) generated by $\{\phi_{1k} : k = 1, ..., J\}$, then: $\tau_{2,2,J} \leq 1/\mu_J$; (3) If, in addition, $J \leq K$ and the sieve space B_K contains the closed linear subspace (in $L^2(X)$) generated by $\{\phi_{0k} : k = 1, ..., J\}$, then: $\sigma_{JK}^{-1} \leq 1/\mu_J$ and hence $\sigma_{JK}^{-1} = \tau_{2,2,J} = 1/\mu_J$.*

Lemma 4.2.1 parts (1) and (2) is Lemma 1 of Blundell, Chen, and Kristensen (2007). A sufficient condition to bound the sieve measures of ill-posedness σ_{JK}^{-1} and $\tau_{2,2,J}$ is the following.

Assumption 4.2.2. *(**sieve reverse link condition**) There is a continuous increasing function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that: (a) $\|Th\|^2_{L^2(X)} \gtrsim \sum_{j=1}^{J} \varphi(j^{-2/d})|E[h(Y_2)\widetilde{\psi}_{Jj}(Y_2)]|^2$ for all $h \in \Psi_J$; or (b) $\|\Pi_K Th\|^2_{L^2(X)} \gtrsim \sum_{j=1}^{J} \varphi(j^{-2/d})|E[h(Y_2)\widetilde{\psi}_{Jj}(Y_2)]|^2$ for all $h \in \Psi_J$*

It is clear that Assumption 4.2.2(b) implies Assumption 4.2.2(a). Assumption 4.2.2(a) is the so-called "sieve reverse link condition" used in Chen and Pouzo (2012), which is weaker than the "reverse link condition" imposed in Chen and Reiss (2011) and others in the ill-posed inverse literature: $\|Th\|^2_{L^2(X)} \gtrsim \sum_{j=1}^{\infty} \varphi(j^{-2/d})|E[h(Y_2)\widetilde{\psi}_{Jj}(Y_2)]|^2$ for all $h \in B(p, L)$.

Remark 4.2.1. *(1) Assumption 4.2.2(a) implies that $\tau_{2,2,J} \lesssim \left(\varphi(J^{-2/d})\right)^{-1/2}$. (2) Assumption 4.2.2(b) implies that $\tau_{2,2,J} \leq \sigma_{JK}^{-1} \lesssim \left(\varphi(J^{-2/d})\right)^{-1/2}$.*

Given Remark 4.2.1, the NPIV model is called

- *mildly ill-posed* if $\sigma_{JK}^{-1} = O(J^{\varsigma/d})$ or $\varphi(t) = t^{\varsigma}$ for some $\varsigma > 0$;
- *severely ill-posed* if $\sigma_{JK}^{-1} = O(\exp(\frac{1}{2}J^{\varsigma/d}))$ or $\varphi(t) = \exp(-t^{-\varsigma/2})$ for some $\varsigma > 0$.

Define

$$\sigma_{\infty,JK} = \inf_{h \in \Psi_J : \|h\|_\infty = 1} \|\Pi_K Th\|_\infty \leq (\tau_{\infty,\infty,J})^{-1}.$$

Assumption 4.2.3. *(i) $T : L^q(Y_2) \to L^q(X)$ is compact and injective for $q = 2$ and $q = \infty$, (ii) $\sigma_{\infty,JK}^{-1}\|\Pi_K T(h_0 - \pi_J h_0)\|_\infty \lesssim \|h_0 - \pi_J h_0\|_\infty$.*

Injectivity of the conditional expectation operator T in Assumption 4.2.3(i) may be interpreted as a nonparametric version of the usual relevance condition in linear instrumental variables models. Assumption 4.2.3(ii) is a sup-norm analogue of the so-called "stability condition" imposed in the ill-posed inverse regression literature, such as Assumption 6 of Blundell, Chen, and Kristensen (2007) and Assumption 5.2(ii) of Chen and Pouzo (2012).

Spline or wavelet sieves together with sharp bounds on the approximation error due to Huang (2003) are used to control the "bias term" $\|P_n h_0 - h_0\|_\infty$.[5] Control of the "bias term" $\|P_n h_0 - h_0\|_\infty$ is more involved in the sieve NPIV context than the sieve nonparametric LS regression context. In particular, control of this term makes use of an additional argument using exponential inequalities. To simplify presentation, the next theorem just presents the uniform convergence rate for sieve NPIV estimators under i.i.d. data.

Theorem 4.2.1. *Let Assumptions 4.1.1, 4.1.2, 4.1.3 (with $\Psi_J = BSpl(J, [0,1]^d, \gamma)$ or $Wav(J, [0,1]^d, \gamma)$), 4.1.4 (with $\Psi_J = BSpl(J, [0,1]^d, \gamma_x)$ or $Wav(J, [0,1]^d, \gamma_x)$), 4.2.1 and 4.2.3 hold. If $\{(X_i, Y_{2i})\}_{i=1}^n$ is i.i.d. then:*

$$\|h_0 - \widehat{h}\|_\infty = O_p(J^{-p/d} + \sigma_{JK}^{-1}\sqrt{K(\log n)/n})$$

provided $J \leq K$, $K \lesssim (n/\log n)^{\delta/(2+\delta)}$, and $\sigma_{JK}^{-1} K \sqrt{(\log n)/n} \lesssim 1$ as $n, J, K \to \infty$.

(1) Mildly ill-posed case ($\sigma_{JK}^{-1} = O(J^{\varsigma/d})$ or $\varphi(t) = t^\varsigma$). If Assumption 4.1.2 holds with $\delta \geq d/(\varsigma + p)$, and $J \asymp K \asymp (n/\log n)^{d/(2(p+\varsigma)+d)}$ with $K/J \to c_0 \geq 1$, then:

$$\|h_0 - \widehat{h}\|_\infty = O_p((n/\log n)^{-p/(2(p+\varsigma)+d)}).$$

(2) Severely ill-posed case ($\sigma_{JK}^{-1} = O(\exp(\frac{1}{2}J^{\varsigma/d}))$ or $\varphi(t) = \exp(-t^{-\varsigma/2})$). If Assumption 4.1.2 holds with $\delta > 0$, and $J = c_0'(\log n)^{d/\varsigma}$ for any $c_0' \in (0,1)$ with $K = c_0 J$ for some finite $c_0 \geq 1$, then:

$$\|h_0 - \widehat{h}\|_\infty = O_p((\log n)^{-p/\varsigma}).$$

[5] The key property of spline and wavelet sieve spaces that permits this sharp bound is their local support (see the appendix to Huang (2003)). Other sieve bases such as orthogonal polynomial bases do not have this property and are therefore unable to attain the optimal sup-norm convergence rates for NPIV or nonparametric series LS regression.

Remark 4.2.2. *Under conditions similar to those for Theorem 4.2.1, Blundell, Chen, and Kristensen (2007), Chen and Reiss (2011) and Chen and Pouzo (2012) previously obtained the following $L^2(Y_2)$-norm convergence rate for the sieve NPIV estimator:*

$$\|h_0 - \widehat{h}\|_{L^2(Y_2)} = O_p(J^{-p/d} + \tau_{2,2,J}\sqrt{K/n}).$$

(1) Mildly ill-posed case ($\tau_{2,2,J} = O(J^{\varsigma/d})$ or $\varphi(t) = t^{\varsigma}$),

$$\|h_0 - \widehat{h}\|_{L^2(Y_2)} = O_p(n^{-p/(2(p+\varsigma)+d)})\,.$$

(2) Severely ill-posed case ($\tau_{2,2,J} = O(\exp(\frac{1}{2}J^{\varsigma/d}))$ or $\varphi(t) = \exp(-t^{-\varsigma/2})$),

$$\|h_0 - \widehat{h}\|_{L^2(Y_2)} = O_p((\log n)^{-p/\varsigma})\,.$$

Chen and Reiss (2011) show that these $L^2(Y_2)$-norm rates are optimal in the sense that they coincide with the minimax risk lower bound in $L^2(Y_2)$ loss. It is interesting to see that the sup-norm convergence rate is the same as the known optimal $L^2(Y_2)$-norm rate for the severely ill-posed case, and is only power of $\log(n)$ slower than the known optimal $L^2(Y_2)$-norm rate for the mildly ill-posed case. The next subsection shows that these sup-norm convergence rates are optimal.

4.2.2 Lower bounds on uniform convergence rates

For severely ill-posed NPIV models, Chen and Reiss (2011) already showed that $(\log n)^{-p/\varsigma}$ is the minimax lower bound in $L^2(Y_2)$-norm loss uniformly over a class of functions that include the Hölder ball $B(p,L)$ as a subset. Therefore, for a severely ill-posed NPIV model with $\delta_n = (\log n)^{-p/\varsigma}$,

$$\inf_{\widetilde{h}_n}\sup_{h\in B(p,L)} \mathbb{P}_h\left(\|h-\widetilde{h}_n\|_\infty \geq c\delta_n\right) \geq \inf_{\widetilde{h}_n}\sup_{h\in B(p,L)} \mathbb{P}_h\left(\|h-\widetilde{h}_n\|_{L^2(Y_2)} \geq c\delta_n\right) \geq c'$$

where $\inf_{\widetilde{h}_n}$ denotes the infimum over all estimators based on a random sample of size n drawn from the NPIV model, and the finite positive constants c, c' do not depend on sample size n. This and Remark 4.2.2(2) together imply that the sieve NPIV estimator attains the optimal uniform convergence rate in the severely ill-posed case.

We next show that the sup-norm rate for the sieve NPIV estimator obtained in the mildly ill-posed case is also optimal. A primitive smoothness condition on the conditional expectation operator $T : L^2(Y_2) \to L^2(X)$ is first imposed.

Assumption 4.2.4. *There is a $\varsigma > 0$ such that $\|Th\|_{L^2(X)} \lesssim \|h\|_{B^{-\varsigma}_{2,2}}$ for all $h \in B(p, L)$.*

Assumption 4.2.4 is a special case of the so-called "link condition" in Chen and Reiss (2011) for the mildly ill-posed case. It can be equivalently stated as: $\|Th\|^2_{L^2(X)} \lesssim \sum_{j=1}^{\infty} \varphi(j^{-2/d})|E[h(Y_2)\widetilde{\psi}_{Jj}(Y_2)]|^2$ for all $h \in B(p, L)$, with $\varphi(t) = t^{\varsigma}$ for the mildly ill-posed case. Under this assumption, $n^{-p/(2(p+\varsigma)+d)}$ is the minimax risk lower bound uniformly over the Hölder ball $B(p, L)$ in $L^2(Y_2)$-norm loss for the mildly ill-posed NPIR and NPIV models (see Chen and Reiss (2011)).

Theorem 4.2.2. *Let Assumption 4.2.4 hold for the NPIV model with a random sample $\{(Y_{1i}, Y_{2i}, X_i)\}_{i=1}^n$. Then:*

$$\liminf_{n\to\infty} \inf_{\widetilde{h}_n} \sup_{h\in B(p,L)} \mathbb{P}_h \left(\|h - \widetilde{h}_n\|_\infty \geq c(n/\log n)^{-p/(2(p+\varsigma)+d)} \right) \geq c' > 0,$$

where $\inf_{\widetilde{h}_n}$ denotes the infimum over all estimators based on the sample of size n, and the finite positive constants c, c' do not depend on n.

As in Chen and Reiss (2011), Theorem 4.2.2 is proved by (i) noting that the risk (in sup-norm loss) for the NPIV model is at least as large as the risk (in sup-norm loss) for the NPIR model, and (ii) calculating a lower bound (in sup-norm loss) for the NPIR model. Consider a Gaussian reduced-form NPIR model with known operator T, given by

$$\begin{aligned} Y_{1i} &= Th_0(X_i) + u_i, \quad i = 1, ..., n, \\ u_i|X_i &\sim N(0, \sigma^2(X_i)) \quad \text{with} \quad \inf_x \sigma^2(x) \geq \sigma_0^2 > 0\,. \end{aligned} \tag{4.9}$$

Theorem 4.2.2 therefore follows from a sup-norm analogue of Lemma 1 of Chen and Reiss (2011) and the following theorem, which establishes a lower bound on minimax risk over Hölder classes under sup-norm loss for the NPIR model.

Theorem 4.2.3. *Let Assumption 4.2.4 hold for the NPIR model (4.9) with a random sample $\{(Y_{1i}, X_i)\}_{i=1}^n$. Then:*

$$\liminf_{n\to\infty} \inf_{\widetilde{h}_n} \sup_{h \in B(p,L)} \mathbb{P}_h \left(\|h - \widetilde{h}_n\|_\infty \geq c(n/\log n)^{-p/(2(p+\varsigma)+d)} \right) \geq c' > 0,$$

where $\inf_{\widetilde{h}_n}$ denotes the infimum over all estimators based on the sample of size n, and the finite positive constants c, c' depend only on p, L, d, ς and σ_0.

4.3 Optimal uniform convergence rates for sieve LS estimators

The standard nonparametric regression model can be recovered as a special case of (4.4) in which there is no endogeneity, i.e. $Y_2 = X$ and

$$\begin{aligned} Y_{1i} &= h_0(X_i) + \epsilon_i \\ E[\epsilon_i | X_i] &= 0 \end{aligned} \tag{4.10}$$

in which case $h_0(x) = E[Y_{1i}|X_i = x]$.

Stone (1982) (also see Tsybakov (2009)) establishes that $(n/\log n)^{-p/(2p+d)}$ is the minimax risk lower bound in sup-norm loss for the nonparametric LS regression model (4.10) with $h_0 \in B(p, L)$. In this section the general upper bound (Theorem 4.1.1) is applied to show that spline and wavelet sieve LS estimators attain this minimax lower bound for weakly dependent data allowing for heavy-tailed error terms ϵ_i.

Our proof proceeds by noticing that the sieve LS regression estimator

$$\widehat{h}(x) = b^K(x)(B'B)^- B'Y$$

obtains as a special case of the NPIV estimator by setting $Y_2 = X$, $\psi^J = b^K$, $J = K$ and $\gamma = \gamma_x$. In this setting, the quantity $P_n h_0(x)$ just reduces to the orthogonal projection of h_0 onto the sieve space B_K under the inner product induced by the empirical distribution, viz.

$$P_n h_0(x) = \widetilde{b}^K(x)(\widetilde{B}'\widetilde{B}/n)^- \widetilde{B}' H_0/n\,.$$

Moreover, in this case the $J \times K$ matrix S defined in (4.6) reduces to the $K \times K$ identity matrix I_K and its smallest singular value is unity (whence $\sigma_{JK} = 1$). Therefore, the general calculation presented in Theorem 4.1.1 can be used to control the "variance term" $\|\widehat{h} - P_n h_0\|_\infty$. The "bias term" $\|P_n h_0 - h_0\|_\infty$ is controlled as in Huang (2003). It is worth emphasizing that no explicit weak dependence condition is placed on the regressors $\{X_i\}_{i=-\infty}^{\infty}$. Instead, weak dependence is implicitly captured by Condition (ii) on convergence of $\widetilde{B}'\widetilde{B}/n - I_K$.

Theorem 4.3.1. *Let Assumptions 4.1.1, 4.1.2, 4.1.4 (with $\Psi_J = \mathit{BSpl}(J, [0,1]^d, \gamma)$ or $\mathit{Wav}(J, [0,1]^d, \gamma)$) and 4.2.1 hold for Model (4.10). Then:*

$$\|\widehat{h} - h_0\|_\infty = O_p(K^{-p/d} + \sqrt{K(\log n)/n})$$

provided $n, K \to \infty$, and

(i) $K \lesssim (n/\log n)^{\delta/(2+\delta)}$ and $\sqrt{K(\log n)/n} \lesssim 1$

(ii) $\|(\widetilde{B}'\widetilde{B}/n) - I_K\| = O_p(\sqrt{(\log n)/K}) = o_p(1)$.

Condition (ii) is satisfied by applying Lemma 4.4.2 for i.i.d. data and Lemma 4.4.3 for weakly dependent data. Theorem 4.3.1 shows that spline and wavelet sieve LS estimators can achieve this minimax lower bound for weakly dependent data.

Corollary 4.3.1. *Let Assumptions 4.1.1, 4.1.2 (with $\delta \geq d/p$), 4.1.4 (with $\Psi_J = \mathit{BSpl}(J, [0,1]^d, \gamma)$ or $\mathit{Wav}(J, [0,1]^d, \gamma)$) and 4.2.1 hold for Model (4.10). If $K \asymp (n/\log n)^{d/(2p+d)}$ then:*

$$\|\widehat{h} - h_0\|_\infty = O_p((n/\log n)^{-p/(2p+d)})$$

provided that one of the followings is satisfied

(1) the regressors are i.i.d.;

(2) the regressors are exponentially beta-mixing and $d < 2p$*;*

(3) the regressors are algebraically beta-mixing at rate γ *and* $(2+\gamma)d < 2\gamma p$*.*

Corollary 4.3.1 states that for i.i.d. data, Stone's optimal sup-norm convergence rate is achieved by spline and wavelet LS estimators whenever $\delta \geq d/p$ and $d \leq 2p$ (Assumption 4.2.1). If the regressors are exponentially beta-mixing the optimal rate of convergence is achieved with $\delta \geq d/p$ and $d < 2p$. The restrictions $\delta \geq d/p$ and $(2+\gamma)d < 2\gamma p$ for algebraically beta-mixing (at a rate γ) reduces naturally towards the exponentially mixing conditions as the dependence becomes weaker (i.e. γ becomes larger). In all cases, a smoother function (i.e., bigger p) means a lower value of δ, and therefore heaver-tailed error terms ϵ_i, are permitted while still obtaining the optimal sup-norm convergence rate. In particular this is achieved with $\delta = d/p \leq 2$ for i.i.d. data. Recently, Belloni et al. (2012) require that the conditional $(2+\eta)$th moment (for some $\eta > 0$) of ϵ_i be uniformly bounded for spline LS regression estimators to achieve the optimal sup-norm rate for i.i.d. data.[6] Uniform convergence rates of series LS estimators have also been studied by Newey (1997), de Jong (2002), Song (2008), Lee and Robinson (2013) and others, but the sup-norm rates obtained in these papers are slower than the minimax risk lower bound in sup-norm loss of Stone (1982).Our result is the first such optimal sup-norm rate result for a sieve nonparametric LS estimator allowing for weakly-dependent data with heavy-tailed error terms. It should be very useful for nonparametric estimation of financial time-series models that have heavy-tailed error terms.

[6]Chen would like to thank Jianhua Huang for working together on an earlier draft that does achieve the optimal sup-norm rate for a polynomial spline LS estimator with i.i.d. data, but under a stronger condition that $E[\epsilon_i^4|X_i = x]$ is uniformly bounded in x.

4.4 Useful results on random matrices

4.4.1 Convergence rates for sums of dependent random matrices

In this subsection a Bernstein inequality for sums of independent random matrices due to Tropp (2012) is adapted to obtain convergence rates for sums of random matrices formed from beta-mixing (absolutely regular) sequences, where the dimension, norm, and variance measure of the random matrices are allowed to grow with the sample size. These inequalities are particularly useful for establishing convergence rates for semi/nonparametric sieve estimators with weakly-dependent data. The inequality derived in this chapter extends the following result due to Tropp (2012) for independent random matrices.

Theorem 4.4.1 (Tropp (2012))**.** *Let $\{\Xi_i\}_{i=1}^n$ be a finite sequence of independent random matrices with dimensions $d_1 \times d_2$. Assume $E[\Xi_i] = 0$ for each i and $\max_{1\leq i\leq n} \|\Xi_i\|_2 \leq R_n$, and define*

$$\sigma_n^2 = \max\left\{\left\|\sum_{i=1}^n E[\Xi_i\Xi_i']\right\|_2, \left\|\sum_{i=1}^n E[\Xi_i'\Xi_i]\right\|_2\right\}.$$

Then for all $t \geq 0$,

$$\mathbb{P}\left(\left\|\sum_{i=1}^n \Xi_i\right\|_2 \geq t\right) \leq (d_1+d_2)\exp\left(\frac{-t^2/2}{\sigma_n^2 + R_n t/3}\right).$$

Corollary 4.4.1. *Under the conditions of Theorem 4.4.1, if $R_n\sqrt{\log(d_1+d_2)} = o(\sigma_n)$ then*

$$\left\|_2 \sum_{i=1}^n \Xi_{i,n}\right\|_2 = O_p(\sigma_n\sqrt{\log(d_1+d_2)}).$$

We now provide a version of Theorem 4.4.1 and Corollary 4.4.1 for matrix-valued functions of beta-mixing sequences (see Section 2.4 for definitions of forms of beta-mixing). The following extension of Theorem 4.4.1 is made using a Berbee's lemma and a coupling argument (see, e.g., Doukhan et al. (1995)).

Theorem 4.4.2. *Let $\{X_i\}_{i=-\infty}^{\infty}$ be a strictly stationary beta-mixing sequence and let $\Xi_{i,n} = \Xi_n(X_i)$ for each i where $\Xi_n : \mathcal{X} \to \mathbb{R}^{d_1\times d_2}$ is a sequence of measurable $d_1 \times d_2$ matrix-*

valued functions. Assume $E[\Xi_{i,n}] = 0$ *and* $\|\Xi_{i,n}\|_2 \leq R_n$ *for each* i *and define* $s_n^2 = \max_{1\leq i,j\leq n} \max\{\|E[\Xi_{i,n}\Xi_{j,n}']\|_2, \|E[\Xi_{i,n}'\Xi_{j,n}]\|_2\}$. *Let* q *be an integer between* 1 *and* $n/2$ *and let* $I_r = q[n/q]+1,\ldots,n$ *when* $q[n/q] < n$ *and* $I_r = \emptyset$ *when* $q[n/q] = n$. *Then for all* $t \geq 0$,

$$\mathbb{P}\left(\left\|\sum_{i=1}^n \Xi_{i,n}\right\|_2 \geq 6t\right) \leq \frac{n}{q}\beta(q) + \mathbb{P}\left(\left\|\sum_{i\in I_r} \Xi_{i,n}\right\|_2 \geq t\right) + 2(d_1+d_2)\exp\left(\frac{-t^2/2}{nqs_n^2 + qR_nt/3}\right)$$

(where $\|\sum_{i\in I_r}\Xi_{i,n}\| := 0$ *whenever* $I_r = \emptyset$*).*

Corollary 4.4.2. *Under the conditions of Theorem 4.4.2, if* $q = q(n)$ *is chosen such that* $\frac{n}{q}\beta(q) = o(1)$ *and* $R_n\sqrt{q\log(d_1+d_2)} = o(s_n\sqrt{n})$ *then*

$$\left\|\sum_{i=1}^n \Xi_{i,n}\right\|_2 = O_p(s_n\sqrt{nq\log(d_1+d_2)})\,.$$

4.4.2 Empirical identifiability

This subsection provides a readily verifiable condition under which, with probability approaching one (wpa1), the theoretical and empirical L^2 norms are equivalent over a linear sieve space. This equivalence, referred to by Huang (2003) as *empirical identifiability*, has several applications in nonparametric sieve estimation. In the context of nonparametric series regression, empirical identifiability ensures the estimator is the orthogonal projection of Y onto the sieve space under the empirical inner product and is uniquely defined (Huang, 2003). Empirical identifiability is also used to establish the large-sample properties of sieve conditional moment estimators (Chen and Pouzo, 2012). A sufficient condition for empirical identifiability is now cast in terms of convergence of a random matrix, which is verified for i.i.d. and beta-mixing sequences.

A subspace $\mathcal{A} \subseteq L^2(X)$ is said to be *empirically identifiable* if $\frac{1}{n}\sum_{i=1}^n b(X_i)^2 = 0$ implies $b = 0$ a.e.-$[F_X]$ where F_X dentoes the distribution of X. A sequence of spaces

$\{\mathcal{A}_K : K \geq 1\} \subseteq L^2(X)$ is empirically identifiable wpa1 as $K = K(n) \to \infty$ with n if

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{a \in A_K} \left| \frac{\frac{1}{n}\sum_{i=1}^n a(X_i)^2 - E[a(X)^2]}{E[a(X)^2]} \right| > t \right) = 0 \tag{4.11}$$

for any $t > 0$. Huang (1998) uses a chaining argument to provide sufficient conditions for (4.11) over the linear space B_K under i.i.d. sampling. Chen and Pouzo (2012) use this argument to establish convergence of sieve conditional moment estimators. Although easy to establish for i.i.d. sequences, it may be difficult to verify (4.11) via chaining arguments for certain types of weakly dependent sequences. To this end, the following is a readily verifiable sufficient condition for empirical identifiability for linear sieve spaces. Let $B_K = clsp\{b_{K1}, \ldots, b_{KK}\}$ denote a general linear sieve space and let $\widetilde{B} = (\widetilde{b}^K(X_1), \ldots, \widetilde{b}^K(X_n))'$ where $\widetilde{b}^K(x)$ is the orthonormalized vector of basis functions.

Condition 4.4.1. $\lambda_{\min}(E[b^K(X)b^K(X)']) > 0$ *for each* $K \geq 1$ *and* $\|\widetilde{B}'\widetilde{B}/n - I_K\|_2 = o_p(1)$.

Lemma 4.4.1. *If* $\lambda_{\min}(E[b^K(X)b^K(X)']) > 0$ *for each* $K \geq 1$ *then*

$$\sup_{b \in B_K} \left| \frac{\frac{1}{n}\sum_{i=1}^n b(X_i)^2 - E[b(X)^2]}{E[b(X)^2]} \right| = \|\widetilde{B}'\widetilde{B}/n - I_K\|_2^2 .$$

Corollary 4.4.3. *Under Condition 4.4.1,* B_K *is empirically identifiable wpa1.*

Condition 4.4.1 is a sufficient condition for (4.11) with a linear sieve space B_K. It should be noted that convergence is only required in the spectral norm. In the i.i.d. case this allows for K to increase more quickly with n than is achievable under the chaining argument of Huang (1998). Recall $\zeta_0(K) = \sup_{x\in\mathcal{X}} \|b^K(x)\|_2$, where $\zeta_0(K) = O(\sqrt{K})$ for tensor products of splines, trigonometric polynomials or wavelets and $\zeta_0(K) = O(K)$ for tensor products of power series or polynomials (Newey, 1997; Huang, 1998). Under the chaining argument of Huang (1998), (4.11) is achieved under the restriction $\zeta_0(K)^2K/n = o(1)$. Huang (2003) relaxes this restriction to $K(\log n)/n = o(1)$ for a polynomial spline sieve. This result is generalized using Lemma 4.4.1 and Corollary 4.4.1

Lemma 4.4.2. *If* $\{X_i\}_{i=1}^n$ *is i.i.d. and* $\lambda_{\min}(E[b^K(X)b^K(X)']) \geq \underline{\lambda} > 0$ *for each* $K \geq 1$, *then*

$$\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 = O_p(\zeta_0(K)\sqrt{(\log K)/n})$$

provided $\zeta_0(K)^2(\log K)/n = o(1)$.

Remark 4.4.1. *If* $\{X_i\}_{i=1}^n$ *is i.i.d.,* $K(\log K)/n = o(1)$ *is sufficient for sieve bases that are tensor products of splines, trigonometric polynomials or wavelets, and* $K^2(\log K)/n = o(1)$ *is sufficient for sieve bases that are tensor products of power series or polynomials.*

The following lemma is useful to provide sufficient conditions for empirical identifiability for beta-mixing sequences, which uses Theorem 4.4.2.

Lemma 4.4.3. *If* $\{X_i\}_{i=-\infty}^{\infty}$ *is strictly stationary and beta-mixing with mixing coefficients such that one can choose an integer sequence* $q = q(n) \leq n/2$ *with* $\beta(q)n/q = o(1)$ *and* $\lambda_{\min}(E[b^K(X)b^K(X)']) \geq \underline{\lambda} > 0$ *for each* $K \geq 1$, *then*

$$\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 = O_p(\zeta_0(K)\sqrt{q(\log K)/n})$$

provided $\zeta_0(K)^2 q \log K/n = o(1)$.

Remark 4.4.2. *If* $\{X_i\}_{i=-\infty}^{\infty}$ *is algebraically beta-mixing at rate* γ, $Kn^{1/(1+\gamma)}(\log K)/n = o(1)$ *is sufficient for sieve bases that are tensor products of splines, trigonometric polynomials or wavelets, and* $K^2 n^{1/(1+\gamma)}(\log K)/n = o(1)$ *is sufficient for sieve bases that are tensor products of power series or polynomials.*

Remark 4.4.3. *If* $\{X_i\}_{i=-\infty}^{\infty}$ *is geometrically beta-mixing,* $K(\log n)^2/n = o(1)$ *is sufficient for sieve bases that are tensor products of splines, trigonometric polynomials or wavelets, and* $K^2(\log n)^2/n = o(1)$ *is sufficient for sieve bases that are tensor products of power series or polynomials.*

4.5 Brief review of B-spline and wavelet sieve spaces

We first outline univariate B-spline and wavelet sieve spaces on $[0,1]$, then deal with the multivariate case by constructing a tensor-product sieve basis.

B-splines B-splines are defined by their order $m \geq 1$ and number of interior knots $N \geq 0$. Define the knot set

$$t_{-(m-1)} = \ldots = t_0 \leq t_1 \leq \ldots \leq t_N \leq t_{N+1} = \ldots = t_{N+m}$$

where $t_0 = 0$ and $t_{N+1} = 1$. The B-spline basis is then defined recursively via the De Boor relation. This results in a total of $K = N + m$ splines which together form a partition of unity. Each spline is a polynomial of degree $m-1$ on each interior interval $I_1 = [t_0, t_1), \ldots, I_n = [t_N, t_{N+1}]$ and is $(m-2)$-times continuously differentiable on $[0,1]$ whenever $m \geq 2$. The mesh ratio is defined as

$$\text{mesh}(K) = \frac{\max_{0 \leq n \leq N}(t_{n+1} - t_n)}{\min_{0 \leq n \leq N}(t_{n+1} - t_n)} .$$

We let the space $\text{BSpl}(K, [0,1])$ be the closed linear span of these $K = N + m$ splines. The space $\text{BSpl}(K, [0,1])$ has *uniformly bounded mesh ratio* if $\text{mesh}(K) \leq \kappa$ for all $N \geq 0$ and some $\kappa \in (0, \infty)$. The space $\text{BSpl}(K, [0,1])$ has *smoothness* $\gamma = m - 2$, which is denoted as $\text{BSpl}(K, [0,1], \gamma)$ for simplicity. See De Boor (2001) and Schumaker (2007) for further details.

Wavelets We follow the construction of Cohen et al. (1993a,b) for building a wavelet basis for $[0,1]$. Let (ϕ, ψ) be a father and mother wavelet pair that has N vanishing moments and $\text{support}(\phi) = \text{support}(\psi) = [0, 2N-1]$. For given j, the approximation space V_j wavelet space W_j each consist of 2^j functions $\{\phi_{jk}\}_{1 \leq k \leq 2^j}$ and $\{\psi_{jk}\}_{1 \leq k \leq 2^j}$ respectively, such that $\{\phi_{jk}\}_{1 \leq k \leq 2^j - 2N}$ and $\{\psi_{jk}\}_{1 \leq k \leq 2^j - 2N}$ are interior wavelets for which $\phi_{jk}(\cdot) = 2^{j/2}\phi(2^j(\cdot) - k)$ and $\psi_{jk}(\cdot) = 2^{j/2}\psi(2^j(\cdot) - k)$, complemented with another N left-edge functions and N right-edge functions. Choosing $L \geq 1$ such that $2^L \geq 2N$, let the space $\mathrm{Wav}(K, [0,1])$ be the closed linear span of the set of functions

$$W_{LJ} = \{\phi_{Lk} : 1 \leq k \leq 2^L\} \cup \{\psi_{jk} : k = 1, \ldots, 2^j \text{ and } j = L, \ldots, J-1\}$$

for integer $J > L$, and let $K = \#(W_{LJ})$. Let $\mathrm{Wav}(K, [0,1])$ have *regularity* γ if ϕ and ψ are both γ times continuously differentiable, which is denoted as $\mathrm{Wav}(K, [0,1], \gamma)$ for simplicity.

Tensor products To construct a tensor-product B-spline basis of smoothness γ for $[0,1]^d$ with $d > 1$, first construct d univariate B-spline bases for $[0,1]$, say G_i with $G_i = \mathrm{BSpl}(k, [0,1])$ and smoothness γ for each $1 \leq i \leq d$. Then set $K = k^d$ and let $\mathrm{BSpl}(K, [0,1]^d)$ be spanned by the unique k^d functions given by $\prod_{i=1}^d g_i$ with $g_i \in G_i$ for $1 \leq i \leq d$. The tensor-product wavelet basis $\mathrm{Wav}(K, [0,1]^d)$ of regularity γ for $[0,1]^d$ is formed similarly as the tensor product of d univariate Wavelet bases of regularity γ (see Triebel (2006, 2008)).

Wavelet characterization of Besov norms Let $f \in B^\alpha_{p,q}([0,1]^d)$ have wavelet expansion

$$f = \sum_{k=-\infty}^{\infty} \mathsf{a}_k(f)\phi_{Lk} + \sum_{j=L}^{\infty}\sum_{k=-\infty}^{\infty} \mathsf{b}_{jk}(f)\psi_{jk}$$

where $\{\phi_{Lk}, \psi_{jk}\}_{j,k}$ are a Wavelet basis with regularity $\gamma > \alpha$. Equivalent norms to the $B^\alpha_{\infty,\infty}$ and $B^\alpha_{2,2}$ norms may be formulated equivalently in terms of the wavelet coefficient sequences $\{\mathsf{a}_k\}_k$ and $\{\mathsf{b}_{jk}\}_{j,k}$, namely $\|\cdot\|_{b^\alpha_{\infty,\infty}}$ and $\|\cdot\|_{b^\alpha_{2,2}}$, given by

$$\begin{aligned}
\|f\|_{b^\alpha_{\infty,\infty}} &= \sup_k |\mathsf{a}_k(f)| + \sup_{j,k} 2^{j(\alpha+d/2)}|\mathsf{b}_{jk}(f)| \\
\|f\|_{b^\alpha_{2,2}} &= \|\mathsf{a}_{(\cdot)}(f)\| + \left(\textstyle\sum_{j=0}^{\infty}(2^{j\alpha}\|\mathsf{b}_{j(\cdot)}(f)\|)^2\right)^{1/2}
\end{aligned}$$

where $\|\mathsf{a}_{(\cdot)}(f)\|$ and $\|\mathsf{b}_{j(\cdot)}(f)\|$ denote the infinite-dimensional Euclidean norm for the sequences $\{\mathsf{a}_k(f)\}_k$ and $\{\mathsf{b}_{jk}(f)\}_k$ (see, e.g., Johnstone (2013) and Triebel (2006, 2008)).

4.6 Proofs of main results

Proof of Theorem 4.1.1. It is enough to show that $\|\widehat{h} - P_n h_0\|_\infty = O_p(\sigma_{JK}^{-1}\sqrt{K(\log n)/n})$. First write

$$\widehat{h}(y_2) - P_n h_0(y_2) = \widetilde{\psi}^J(y_2)'[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widehat{S}']^- \widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widetilde{B}'e/n$$

where $e = (\epsilon_1, \ldots, \epsilon_n)$. Convexity of $\mathcal{Y}_2$ (Assumption 4.1.1(ii)), smoothness of $\widetilde{\psi}^J$ (Assumption 4.1.3(i)) and the mean value theorem provide that, for any $(y, y^*) \in \mathcal{Y}_2^2$,

$$\begin{aligned}
& |\widehat{h}(y) - P_n h_0(y) - (\widehat{h}(y^*) - P_n h_0(y^*))| \\
= \; & |(\widetilde{\psi}^J(y) - \widetilde{\psi}^J(y^*))'[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widehat{S}']^- \widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widetilde{B}'e/n| \\
= \; & |(y - y^*)'\nabla\widetilde{\psi}^J(y^{**})'[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widehat{S}']^- \widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widetilde{B}'e/n| \\
\leq \; & J^\alpha \|y - y^*\|_2 \|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widehat{S}']^- \widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1} \widetilde{B}'e/n\|_2
\end{aligned}$$

for some y^{**} in the segment between y and y^*, and some $\alpha > 0$ (and independent of y and y^*).

We first show that $T_1 := \|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widehat{S}']^- \widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widetilde{B}'e/n\|_2 = o_p(1)$. By the triangle inequality and properties of the matrix spectral norm,

$$T_1 \;\leq\; (\|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widehat{S}']^- \widehat{S}(\widetilde{B}'\widetilde{B}/n)^- - [SS']^{-1}S\|_2 + \|[SS']^{-1}S\|_2)\|\widetilde{B}'e/n\|_2$$

whence by Lemma 4.7.2 (under condition (ii) of the Theorem), wpa1

$$T_1 \;\lesssim\; \left\{\sigma_{JK}^{-1}\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 + \sigma_{JK}^{-2}\left(\|\widehat{S} - S\|_2 + \|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2\right) + \sigma_{JK}^{-1}\right\}\|\widetilde{B}'e/n\|_2\,.$$

Noting that $\|\widetilde{B}'e/n\|_2 = O_p(\sqrt{K/n})$ (by Markov's inequality under Assumptions 4.1.2 and 4.1.4), it follows by conditions (i) and (ii) of the Theorem that $T_1 = o_p(1)$. Therefore, for any fixed $\bar{M} > 0$,

$$\limsup_{n\to\infty} \mathbb{P}\left(\|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widehat{S}']^- \widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widetilde{B}' e/n\|_2 > \bar{M}\right) = 0\,.$$

Let $\mathscr{B}_n$ denote the event $\|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widehat{S}']^- \widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widetilde{B}' e/n\|_2 \le \bar{M}$ and observe that $\mathbb{P}(\mathscr{B}_n^c) = o(1)$. On $\mathscr{B}_n$, for any $C \ge 1$, a finite positive $\beta = \beta(C)$ and $\gamma = \gamma(C)$ can be chosen such that

$$J^\alpha \|y_0 - y_1\|_2 \|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widehat{S}']^- \widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widetilde{B}' e/n\|_2 \le C\sigma_{JK}^{-1}\sqrt{K(\log n)/n}$$

whenever $\|y_0 - y_1\|_2 \le \beta n^{-\gamma}$. Let $\mathcal{S}_n$ be the smallest subset of $\mathcal{Y}_2$ such that for each $y \in \mathcal{Y}_2$ there exists a $y_n \in \mathcal{S}_n$ with $\|y_n - y\|_2 \le \beta n^{-\gamma}$. For any $y \in \mathcal{Y}_2$ let $y_n(y)$ denote the $y_n \in \mathcal{S}_n$ nearest (in Euclidean distance) to y. Therefore,

$$|\widehat{h}(y) - P_n h_0(y) - (\widehat{h}(y_n(y)) - P_n h_0(y_n(y)))| \le C\sigma_{JK}^{-1}\sqrt{K(\log n)/n} \tag{4.12}$$

for any $y \in \mathcal{Y}_2$, on $\mathscr{B}_n$.

For any $C \ge 1$, straightforward arguments yield

$$\begin{aligned}
& \mathbb{P}\left(\|\widehat{h} - P_n h_0\|_\infty \ge 4C\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right) \\
\le\ & \mathbb{P}\left(\left\{\|\widehat{h} - P_n h_0\|_\infty \ge 4C\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right\} \cap \mathscr{B}_n\right) + \mathbb{P}(\mathscr{B}_n^c) \\
\le\ & \mathbb{P}\left(\left\{\sup_{y\in\mathcal{Y}_2} |\widehat{h}(y) - P_n h_0(y) - (\widehat{h}(y_n(y)) - P_n h_0(y_n(y)))| \ge 2C\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right\} \cap \mathscr{B}_n\right) \\
& + \mathbb{P}\left(\left\{\max_{y_n\in\mathcal{S}_n} |\widehat{h}(y_n) - P_n h_0(y_n)| \ge 2C\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right\} \cap \mathscr{B}_n\right) + \mathbb{P}(\mathscr{B}_n^c) \\
=\ & \mathbb{P}\left(\left\{\max_{y_n\in\mathcal{S}_n} |\widehat{h}(y_n) - P_n h_0(y_n)| \ge 2C\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right\} \cap \mathscr{B}_n\right) + o(1)
\end{aligned}$$

where the final line is by (4.12) and the fact that $\mathbb{P}(\mathscr{B}_n^c) = o(1)$. For the remaining term:

$$
\begin{aligned}
& \mathbb{P}\left(\left\{\max_{y_n \in \mathcal{S}_n} |\widehat{h}(y_n) - P_n h_0(y_n)| \geq 2C\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right\} \cap \mathscr{B}_n\right) \\
\leq \;& \mathbb{P}\left(\max_{y_n \in \mathcal{S}_n} |\widetilde{\psi}^J(y_n)'[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^-\widehat{S}']^-\widehat{S}(\widetilde{B}'\widetilde{B}/n)^-\widetilde{B}'e/n| \geq 2C\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right) \\
\leq \;& \mathbb{P}\left(\max_{y_n \in \mathcal{S}_n} |\widetilde{\psi}^J(y_n)'\{[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^-\widehat{S}']^-\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- - [SS']^{-1}S\}\widetilde{B}'e/n| \geq C\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right) \\
& + \mathbb{P}\left(\max_{y_n \in \mathcal{S}_n} |\widetilde{\psi}^J(y_n)'[SS']^{-1}S\widetilde{B}'e/n| \geq C\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right) \\
=: \;& P_1 + P_2 \,.
\end{aligned}
$$

It is now shown that a sufficiently large C can be chosen to make P_1 and P_2 arbitrarily small as $n, J, K \to \infty$. Observe that $\mathcal{S}_n$ has cardinality $\lesssim n^\nu$ for some $\nu = \nu(C) \in (0, \infty)$ under Assumption 4.1.1(ii).

Control of P_1: The Cauchy-Schwarz inequality and Assumption 4.1.3 yield

$$
\begin{aligned}
& |\widetilde{\psi}^J(y_n)'\{[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^-\widehat{S}']^-\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- - [SS']^{-1}S\}\widetilde{B}'e/n| \\
\lesssim \;& \sqrt{J}\|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^-\widehat{S}']^-\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- - [SS']^{-1}S\|_2 \times O_p(\sqrt{K/n})
\end{aligned}
$$

uniformly for $y_n \in \mathcal{S}_n$ (recalling that $\|\widetilde{B}'e/n\|_2 = O_p(\sqrt{K/n})$ under Assumptions 4.1.2 and 4.1.4). Therefore, P_1 will vanish asymptotically provided

$$
T_2 := \sigma_{JK}\sqrt{J}\|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^-\widehat{S}']^-\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- - [SS']^{-1}S\|_2/\sqrt{\log n} = o_p(1)\,.
$$

Under condition (ii), the bound

$$
T_2 \lesssim \sqrt{J}\left\{\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 + \sigma_{JK}^{-1}\left(\|\widehat{S} - S\|_2 + \|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2\right)\right\}/\sqrt{\log n}
$$

holds wpa1 by Lemma 4.7.2, and so $T_2 = o_p(1)$ by virtue of conditions (i) and (ii) of the Theorem.

Control of P_2: Let $\{M_n : n \geq 1\}$ be an increasing sequence diverging to $+\infty$ and let

$$\begin{aligned}
\epsilon_{1,i,n} &= \epsilon_i\{|\epsilon_i| \leq M_n\} \\
\epsilon_{2,i,n} &= \epsilon_i - \epsilon_{1,i,n} \\
g_{i,n} &= \psi^J(y_n)' \left[SS'\right]^{-1} S\widetilde{b}^K(X_i)\,.
\end{aligned}$$

Simple application of the triangle inequality yields

$$\begin{aligned}
P_2 &\leq \mathbb{P}(\mathscr{A}_n^c) \\
&\quad + (\#\mathcal{S}_n) \max_{y_n \in \mathcal{S}_n} \mathbb{P}\left(\left\{\left|\sum_{i=1}^n g_{i,n}(\epsilon_{1,i,n} - E[\epsilon_{1,i,n}|\mathcal{F}_{i-1}])\right| > \frac{C}{3}\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right\} \cap \mathscr{A}_n\right) \\
&\quad + \mathbb{P}\left(\max_{y_n \in \mathcal{S}_n}\left|\frac{1}{n}\sum_{i=1}^n g_{i,n}E[\epsilon_{1,i,n}|\mathcal{F}_{i-1}]\right| \geq \frac{C}{3}\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right) \\
&\quad + \mathbb{P}\left(\max_{y_n \in \mathcal{S}_n}\left|\frac{1}{n}\sum_{i=1}^n g_{i,n}\epsilon_{2,i,n}\right| \geq \frac{C}{3}\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right) \\
&=: \mathbb{P}(\mathscr{A}_n^c) + P_{11} + P_{12} + P_{13}\,.
\end{aligned}$$

where $\mathscr{A}_n$ is a measurable set to be defined for which $\mathbb{P}(\mathscr{A}_n^c) = o(1)$.

It will now be shown that P_{21}, P_{22} and P_{23} vanish asymptotically provided a sequence $\{M_n : n \geq 1\}$ may be chosen such that $\sqrt{nJ/\log n} = O(M_n^{1+\delta})$ and $M_n = O(\sqrt{n/(J\log n)})$ and $J \leq K$. Choosing $J \leq K$ and setting $M_n^{1+\delta} \asymp \sqrt{nK/\log n}$ trivially satisfies the condition $\sqrt{nK/\log n} = O(M_n^{1+\delta})$. The condition $M_n = O(\sqrt{n/(K\log n)})$ is satisfied for this choice of M_n provided $K \lesssim (n/\log n)^{\delta/(2+\delta)}$.

Control of P_{22} and P_{23}: For term P_{23}, first note that

$$|g_{i,n}| \lesssim \sigma_{JK}^{-1}\sqrt{JK}$$

whenever $\sigma_{JK} > 0$ by the Cauchy-Schwarz inequality, and Assumptions 4.1.3 and 4.1.4.

This, together with Markov's inequality and Assumption 4.1.2(iii) yields

$$\begin{aligned}\mathbb{P}\left(\max_{y_n\in\mathcal{S}_n}\left|\frac{1}{n}\sum_{i=1}^n g_{i,n}\epsilon_{2,i,n}\right| \geq \frac{C}{3}\sigma_{JK}^{-1}\sqrt{K(\log n)/n}\right) &\lesssim \frac{\sigma_{JK}^{-1}\sqrt{JK}E[|\epsilon_i|\{|\epsilon_i|>M_n\}]}{\sigma_{JK}^{-1}\sqrt{K(\log n)/n}} \\ &\leq \sqrt{\frac{nJ}{\log n}}\frac{E[|\epsilon_i|^{2+\delta}\{|\epsilon_i|>M_n\}]}{M_n^{1+\delta}}\end{aligned}$$

which is $o(1)$ provided $\sqrt{nJ/\log n} = O(M_n^{1+\delta})$. P_{22} is controlled by an identical argument, using the fact that $E[\epsilon_{1,i,n}|\mathcal{F}_{i-1}] = -E[\epsilon_{2,i,n}|\mathcal{F}_{i-1}]$ by Assumption 4.1.2(i).

Control of P_{21}: P_{21} is to be controlled using an exponential inequality for martingales due to van de Geer (1995). Let $\mathscr{A}_n$ denote the set on which $\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 \leq \frac{1}{2}$ and observe that $\mathbb{P}(\mathscr{A}_n^c) = o(1)$ under the condition $\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 = o_p(1)$. Under Assumptions 4.1.2(ii), 4.1.3, and 4.1.4, the predictable variation of the summands in P_{21} may be bounded using the fact that

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n E[(g_{i,n}(\epsilon_{1,i,n} - E[\epsilon_{1,i,n}|\mathcal{F}_{i-1}]))^2|\mathcal{F}_{i-1}] &\lesssim n^{-1}\widetilde{\psi}^J(y_n)'[SS']^{-1}S(\widetilde{B}'\widetilde{B}/n)S'[SS']^{-1}\widetilde{\psi}^J(y_n) \\ &\lesssim \sigma_{JK}^{-2}J/n \quad \text{on } \mathscr{A}_n\end{aligned}$$

uniformly for $y_n \in \mathcal{S}_n$. Moreover, under Assumption 4.1.4, each summand is bounded uniformly for $y_n \in \mathcal{S}_n$ by

$$|n^{-1}g_{i,n}(\epsilon_{1,i,n} - E[\epsilon_{1,i,n}|\mathcal{F}_{i-1}])| \lesssim \frac{\sigma_{JK}^{-1}\sqrt{JK}M_n}{n}.$$

Lemma 2.1 of van de Geer (1995) then provides that P_{21} may be bounded by

$$\begin{aligned}P_{21} &\lesssim n^{\nu}\exp\left\{-\frac{C\sigma_{JK}^{-2}K(\log n)/n}{c_1\sigma_{JK}^{-2}J/n + c_2 n^{-1}\sigma_{JK}^{-2}\sqrt{JK}M_n\sqrt{CK(\log n)/n}}\right\} \\ &\lesssim \exp\left\{\log n - \frac{CK(\log n)/n}{c_3 J/n}\right\} + \exp\left\{\log n - \frac{\sqrt{CK(\log n)/n}}{c_4 K M_n/n}\right\}\end{aligned}$$

for finite positive constants $c_1,\ldots,c_4$. Thus P_{21} is $o(1)$ for large enough C by virtue of the conditions $M_n = O(\sqrt{n/(J\log n)})$ and $J \leq K$. □

Proof of Theorem 4.2.1. Theorem 4.1.1 gives $\|\widehat{h} - P_n h_0\|_\infty = O_p(\sigma_{JK}^{-1}\sqrt{K(\log n)/n})$ provided the conditions of Theorem 4.1.1 are satisfied. The conditions $J \leq K$ and $K \lesssim (n/\log n)^{\delta/(2+\delta)}$ are satisfied by hypothesis. Corollary 4.4.1 (under Assumptions 4.1.3 and 4.1.4 and the fact that $\{(X_i, Y_{2i})\}_{i=1}^n$ are i.i.d. and $J \leq K$) yields

$$\begin{aligned} \|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 &= O_p(\sqrt{K(\log K)/n}) && (4.13) \\ \|\widehat{S} - S\|_2 &= O_p(\sqrt{K(\log K)/n})\,. && (4.14) \end{aligned}$$

Therefore, the conditions of Theorem 4.1.1 are satisfied by these rates and the conditions on J and K in Theorem 4.2.1.

It remains to control the approximation error $\|P_n h_0 - h_0\|_\infty$. Under Assumptions 4.1.1, 4.1.3 (with $\Psi_J = \mathrm{BSpl}(J, [0,1]^d, \gamma)$ *or* $\mathrm{Wav}(J, [0,1]^d, \gamma)$) and 4.2.1 there exists a $\pi_J h_0 = \widetilde{\psi}^{J\prime} c_J \in \Psi_J$ with $c_J \in \mathbb{R}^J$ such that

$$\|h_0 - \pi_J h_0\|_\infty = O(J^{-p/d}) \qquad (4.15)$$

(see, e.g., Huang (1998)) so it suffices to control $\|P_n h_0 - \pi_J h_0\|_\infty$.

Both $P_n h_0$ and $\pi_J h_0$ lie in Ψ_J, so $\|P_n h_0 - \pi_J h_0\|_\infty$ may be rewritten as

$$\begin{aligned} \|P_n h_0 - \pi_J h_0\|_\infty &= \frac{\|P_n h_0 - \pi_J h_0\|_\infty}{\|_2\Pi_K T(P_n h_0 - \pi_J h_0)\|_\infty} \times \|_2\Pi_K T(P_n h_0 - \pi_J h_0)\|_\infty \\ &\leq \sigma_{\infty,JK}^{-1} \times \|_2\Pi_K T(P_n h_0 - \pi_J h_0)\|_\infty && (4.16) \end{aligned}$$

where

$$\Pi_K T(P_n h_0 - \pi_J h_0)(x) = \widetilde{b}^K(x) S'[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widehat{S}]^- \widehat{S}(\widetilde{B}'\widetilde{B}/n)^- \widetilde{B}'(H_0 - \Psi c_J)/n\,.$$

Define the $K \times K$ matrices

$$\begin{aligned} D &= S'[SS']^{-1}S \\ \widehat{D} &= (\widetilde{B}'\widetilde{B}/n)^{-}\widehat{S}'[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-}\widehat{S}]^{-}\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-}\,. \end{aligned} \tag{4.17}$$

By the triangle inequality,

$$\begin{aligned} & \|_2\Pi_K T(P_n h_0 - \pi_J h_0)\|_\infty \\ \leq\ & \|\widetilde{b}^K(x)\widehat{D}\widetilde{B}'(H_0 - \Psi c_J)/n\|_\infty && (4.18) \\ & +\|\widetilde{b}^K(x)\{S' - (\widetilde{B}'\widetilde{B}/n)^{-}\widehat{S}'\}[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-}\widehat{S}]^{-}\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-}\widetilde{B}'(H_0 - \Psi c_J)/n\|_\infty && (4.19) \end{aligned}$$

The arguments below will show that

$$\|_2\Pi_K T(P_n h_0 - \pi_J h_0)\|_\infty = O_p(\sqrt{K(\log n)/n}) \times \|h_0 - \pi_J h_0\|_\infty + O_p(1) \times \|_2\Pi_K T(h_0 - \pi_J h_0)\|_\infty\,. \tag{4.20}$$

Substituting (4.20) into (4.16) and using Assumption 4.2.3(ii), the bound $\sigma_{\infty,JK}^{-1} \lesssim \sqrt{J} \times \sigma_{JK}^{-1}$ (under Assumption 4.1.3), equation (4.15), and the condition $p \geq d/2$ in Assumption 4.2.1 yields the desired result

$$\|P_n h_0 - \pi_J h_0\|_\infty = O_p(J^{-p/d} + \sigma_{JK}^{-1}\sqrt{K(\log n)/n})\,.$$

Control of (4.18): By the triangle and Cauchy-Schwarz inequalities and compatibility of the spectral norm under multiplication,

$$
\begin{aligned}
(4.18) \quad \leq\ & \|\widetilde{b}^K(x)(\widehat{D}-D)\widetilde{B}'(H_0-\Psi c_J)/n\|_\infty \\
& +\|\widetilde{b}^K(x)D\{\widetilde{B}'(H_0-\Psi c_J)/n - E[\widetilde{b}^K(X_i)(h_0(Y_{2i})-\pi_J h_0(Y_{2i}))]\}\|_\infty \\
& +\|\widetilde{b}^K(x)DE[\widetilde{b}^K(X_i)\Pi_K T(h_0-\pi_J h_0)(X_i)]\|_\infty \\
\lesssim\ & \sqrt{K}\|\widehat{D}-D\|_2\{O_p(\sqrt{K/n})\times\|h_0-\pi_J h_0\|_\infty + \|_2\Pi_K T(h_0-\pi_J h_0)\|_{L^2(X)}\} \\
& +O_p(\sqrt{K(\log n)/n})\times\|h_0-\pi_J h_0\|_\infty \\
& +\|\widetilde{b}^K(x)DE[\widetilde{b}^K(X_i)\Pi_K T(h_0-\pi_J h_0)(X_i)]\|_\infty
\end{aligned}
$$

by Lemma 4.7.3 and properties of the spectral norm. Lemma 4.7.2 (the conditions of Lemma 4.7.2 are satisfied by (4.13) and (4.14) and the condition $\sigma_{JK}^{-1}K\sqrt{(\log n)/n}\lesssim 1$) and the condition $\sigma_{JK}^{-1}K\sqrt{(\log n)/n}\lesssim 1$ yield $\sqrt{K}\|\widehat{D}-D\|_2 = O_p(1)$. Finally, Lemma 4.7.1 (under Assumptions 4.1.1 and 4.1.4) provides that

$$\|\widetilde{b}^K(x)DE[\widetilde{b}^K(X_i)\Pi_K T(h_0-\pi_J h_0)(X_i)]\|_\infty \lesssim \|\Pi_K T(h_0-\pi_J h_0)\|_\infty$$

and so

$$(4.18) = O_p(\sqrt{K(\log n)/n})\times\|h_0-\pi_J h_0\|_\infty + O_p(1)\times\|\Pi_K T(h_0-\pi_J h_0)\|_\infty$$

as required.

Control of (4.19): By the Cauchy-Schwarz inequality, compatibility of the spectral norm under multiplication, and Assumption 4.1.4,

$$
\begin{aligned}
(4.19) \quad \lesssim\ & \sqrt{K}\|S'-(\widetilde{B}'\widetilde{B}/n)^-\widehat{S}'\|_2\|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^-\widehat{S}]^-\widehat{S}(\widetilde{B}'\widetilde{B}/n)^-\|_2\|\widetilde{B}'(H_0-\Psi c_J)/n\|_2 \\
\lesssim\ & \sigma_{JK}^{-1}\sqrt{K}\{\|(\widetilde{B}'\widetilde{B}/n)-I_K\|_2+\|\widehat{S}-S\|_2\}\|\widetilde{B}'(H_0-\Psi c_J)/n\|_2
\end{aligned}
$$

where the second line holds wpa1, by Lemma 4.7.2 (the conditions of Lemma 4.7.2 are satisfied by virtue of (4.13) and (4.14) and the condition $\sigma_{JK}^{-1} K\sqrt{(\log n)/n} \lesssim 1$). Applying (4.13) and (4.14) and the condition $\sigma_{JK}^{-1} K\sqrt{(\log n)/n} \lesssim 1$ again yields

$$
\begin{aligned}
(4.19) \quad &= \quad O_p(1) \times \|\widetilde{B}'(H_0 - \Psi c_J)/n\|_2 \\
&= \quad O_p(\sqrt{K/n}) \times \|h_0 - \pi_J h_0\|_\infty + O_p(1) \times \|\Pi_K T(h_0 - \pi_J h_0)\|_\infty
\end{aligned}
$$

by Lemma 4.7.3, equation (4.15), and the relation between $\|\cdot\|_\infty$ and $\|\cdot\|_{L^2(X)}$. □

Proof of Lemma 4.2.1. As already mentioned, Assumption 4.2.3(i) implies that the operators T, T^*, TT^* and T^*T are all compact with the singular value system $\{\mu_k; \phi_{1k}, \phi_{0k}\}_{k=1}^\infty$ where $\mu_1 = 1 \geq \mu_2 \geq \mu_3 \geq ... \searrow 0$. For any $h \in B(p, L) \subset L^2(Y_2)$, $g \in L^2(X)$,

$$
(Th)(x) = \sum_{k=1}^\infty \mu_k \langle h, \phi_{1k}\rangle_{Y_2} \phi_{0k}(x), \quad (T^*g)(y_2) = \sum_{k=1}^\infty \mu_k \langle g, \phi_{0k}\rangle_X \phi_{1k}(y_2).
$$

Let $\mathcal{P}_J = clsp\{\phi_{0k} : k = 1, ..., J\}$, and note that $\mathcal{P}_J$ is a closed linear subspace of B_K under the conditions of part (3).

To prove part (3), for any $h \in \Psi_J$ with $J \leq K$,

$$
\begin{aligned}
\Pi_K Th(\cdot) \quad &= \quad \sum_{j=1}^K \langle Th, \widetilde{b}_{Kj}\rangle_X \widetilde{b}_{Kj}(\cdot) \\
&= \quad \sum_{j=1}^J \langle Th, \phi_{0j}\rangle_X \phi_{0j}(\cdot) + R(\cdot, h)
\end{aligned}
$$

for some $R(\cdot,h) \in B_K \backslash \mathcal{P}_J$. Therefore

$$\begin{aligned}
\sigma_{JK}^2 &= \inf_{h\in\Psi_J:\|h\|_{L^2(Y_2)}=1} \|\Pi_K Th(X)\|_{L^2(X)}^2 \\
&= \inf_{h\in\Psi_J:\|h\|_{L^2(Y_2)}=1} \left\| \sum_{j=1}^J \langle Th, \phi_{0j}\rangle_X \phi_{0j}(\cdot) + R(\cdot,h) \right\|_{L^2(X)}^2 \\
&= \inf_{h\in\Psi_J:\|h\|_{L^2(Y_2)}=1} \left(\sum_{j=1}^J \langle Th, \phi_{0j}\rangle_X^2 + \|R(\cdot,h)\|_{L^2(X)}^2 \right) \\
&\geq \inf_{h\in\Psi_J:\|h\|_{L^2(Y_2)}=1} \left(\sum_{j=1}^J \langle Th, \phi_{0j}\rangle_X^2 \right) \\
&= \inf_{h\in\Psi_J:\|h\|_{L^2(Y_2)}=1} \left(\sum_{j=1}^J \mu_j^2 \langle h, \phi_{1j}\rangle_{Y_2}^2 \right) \\
&\geq \mu_J^2 \inf_{h\in\Psi_J:\|h\|_{L^2(Y_2)}=1} \left(\sum_{j=1}^J \langle h, \phi_{1j}\rangle_{Y_2}^2 \right) \quad = \quad \mu_J^2 .
\end{aligned}$$

This, together with part (1), gives $1/\mu_J \geq \sigma_{JK}^{-1} \geq \tau_{2,2,J} \geq 1/\mu_J$. □

Proof of Theorem 4.2.3. Our proof proceeds by application of Theorem 2.5 of Tsybakov (2009) (page 99).

We first explain the scalar ($d = 1$) case in detail. Let $\{\phi_{jk}, \psi_{jk}\}_{j,k}$ be a wavelet basis for $L^2([0,1])$ as in the construction of Cohen et al. (1993a,b) with regularity $\gamma > \max\{p, \varsigma\}$ using a pair (ϕ,ψ) for which support(ϕ) = support(ψ) = $[0, 2N-1]$. The precise type of wavelet is not important, all that is required is that $\|\psi\|_\infty < \infty$. For given j, the wavelet space W_j consists of 2^j functions $\{\psi_{jk}\}_{1\leq k\leq 2^j}$, such that $\{\psi_{jk}\}_{1\leq k\leq 2^j-2N}$ are interior wavelets for which $\psi_{jk}(\cdot) = 2^{j/2}\psi(2^j(\cdot)-k)$. Choose j deterministically with n, such that

$$2^j \asymp (n/\log n)^{1/(2(p+\varsigma)+1)} .$$

By construction, the support of each interior wavelet is an interval of length $2^{-j}(2N-1)$. Thus for all j sufficiently large,[7] a set $M \subset \{1,\ldots,2^j - 2N\}$ of interior wavelets with

[7] Hence the lim inf in the statement of the Lemma.

$\#(M) \gtrsim 2^j$ may be chosen such that $\text{support}(\psi_{jm}) \cap \text{support}(\psi_{jm'}) = \emptyset$ for all $m, m' \in M$ with $m \neq m'$. Note also that, by construction, $\#(M) \leq 2^j$ (since there are $2^j - 2N$ interior wavelets).

We begin by defining a family of submodels. Let $h_0 \in B(p, L)$ be such that $\|h_0\|_{B^p_{\infty,\infty}(\mathcal{Y}_2)} \leq L/2$, and for each $m \in M$ let

$$h_m = h_0 + c_0 2^{-j(p+1/2)} \psi_{jm}$$

where c_0 is a positive constant to be defined subsequently. Noting that

$$\begin{aligned} c_0 2^{-j(p+1/2)} \|\psi_{jm}\|_{B^p_{\infty,\infty}} &\lesssim c_0 2^{-j(p+1/2)} \|\psi_{jm}\|_{b^p_{\infty,\infty}} \\ &\leq c_0 \end{aligned}$$

it follows by the triangle inequality that $\|h_m\|_{B^p_{\infty,\infty}} \leq L$ uniformly in m for all sufficiently small c_0. For $m \in \{0\} \cup M$ let P_m be the joint distribution of $\{(X_i, Y_{1i})\}_{i=1}^n$ with $Y_{1i} = Th_m(X_i) + u_i$ for the Gaussian NPIR model (4.9).

To apply Theorem 2.5 of Tsybakov (2009), first note that for any $m \in M$

$$\begin{aligned} \|h_0 - h_m\|_\infty &= c_0 2^{-j(p+1/2)} \|\psi_{jm}\|_\infty \\ &= c_0 2^{-jp} \|\psi\|_\infty \end{aligned}$$

and for any $m, m' \in M$ with $m \neq m'$

$$\begin{aligned} \|h_m - h_{m'}\|_\infty &= c_0 2^{-j(p+1/2)} \|\psi_{jm} - \psi_{jm'}\|_\infty \\ &= 2c_0 2^{-jp} \|\psi\|_\infty \end{aligned}$$

by virtue of the disjoint support of $\{\psi_{jm}\}_{m \in M}$. Using the KL divergence for the multivariate normal distribution (under the Gaussian NPIR model (4.9)), Assumption 4.2.4 and the equivalence between the Besov function-space and sequence-space norms, the KL distance

$K(P_m, P_0)$ is

$$\begin{aligned}
K(P_m, P_0) &\leq \frac{1}{2}\sum_{i=1}^{n}(c_0 2^{-j(p+1/2)})^2 E\left[\frac{(T\psi_{jm}(X_i))^2}{\sigma^2(X_i)}\right] \\
&\leq \frac{1}{2}\sum_{i=1}^{n}(c_0 2^{-j(p+1/2)})^2 \frac{E\left[(T\psi_{jm}(X_i))^2\right]}{\sigma_0^2} \\
&= \frac{n}{2\sigma_0^2}(c_0 2^{-j(p+1/2)})^2 \|(T^*T)^{1/2}\psi_{jm}(Y_2)\|^2_{L^2(Y_2)} \\
&\lesssim \frac{n}{2\sigma_0^2}(c_0 2^{-j(p+1/2)})^2 (2^{-j\varsigma})^2 \\
&= \frac{n}{2\sigma_0^2} c_0^2 2^{-j(2(p+\varsigma)+1)} \\
&\lesssim c_0^2 \log n
\end{aligned}$$

since $2^{-j} \asymp ((\log n)/n)^{1/(2(p+\varsigma)+1)}$. Moreover, since $\#(M) \asymp 2^j$,

$$\log(\#(M)) \lesssim \log n + \log\log n$$

Therefore, c_0 may be chosen sufficiently small that $\|h_m\|_{B^p_{\infty,\infty}} \leq L$ and $K(P_m, P_0) \leq \frac{1}{8}\log(\#(M))$ uniformly in m for all n sufficiently large. All conditions of Theorem 2.5 of Tsybakov (2009) are satisfied and the result follows.

The multivarite case uses similar arguments for a tensor-product wavelet basis (see Triebel (2006, 2008)). Choose the same j for each univariate space such that $2^j \asymp (n/\log n)^{1/(2(p+\varsigma)+d)}$. Therefore, the tensor-product wavelet space has dimension $2^{jd} \asymp (n/\log n)^{d/(2(p+\varsigma)+d)}$. Then construct the same family of submodels, setting $h_m = h_0 + c_0 2^{-j(p+d/2)}\psi_{jm}$ where ψ_{jm} is now the product of d interior univariate wavelets defined previously. Thus

$$\|h_m - h_{m'}\|_\infty \gtrsim c_0 2^{-jp}$$

for each $m, m' \in \{0\} \cup M$ with $m \neq m'$, and

$$\begin{aligned} K(P_m, P_0) &\lesssim \frac{n}{2\sigma_0^2}(c_0 2^{-j(p+d/2)})^2 (2^{-j\varsigma})^2 \\ &= \frac{n}{2\sigma_0^2} c_0^2 2^{-j(2(p+\varsigma)+d)} \\ &\lesssim c_0^2 \log n\,. \end{aligned}$$

The result follows as in the univariate case. $\square$

Proof of Theorem 4.3.1. The variance term is immediate from Theorem 4.1.1 with $\sigma_{JK} = 1$. The bias calculation follows from Huang (2003) under Assumptions 4.1.1(ii), 4.1.4 (with $B_K = \text{BSpl}(K, [0,1]^d, \gamma)$ *or* $\text{Wav}(K, [0,1]^d, \gamma)$), and 4.2.1 and the fact that the empirical and true $L^2(X)$ norms are equivalent over B_K wpa1 by virtue of the condition $\|\widetilde{B}'\widetilde{B}/n - I_K\|_2 = o_p(1)$, which is implied by Condition (ii). $\square$

Proof of Corollary 4.3.1. By Theorem 4.3.1, the rate $(n/\log n)^{-p/(2p+d)}$ is achieved by setting $K \asymp (n/\log n)^{d/(2p+d)})$, with $\delta \geq d/p$ for condition (i) to hold. (1) When the regressors are i.i.d., by Lemma 4.4.2, condition (ii) is satisfied provided that $d \leq 2p$ (which is assumed in Assumption 4.2.1). (2) When the regressors are exponentially beta-mixing, by Lemma 4.4.3, condition (ii) is satisfied provided that $d < 2p$. (3) When the regressors are algebraically beta-mixing at rate γ, by Lemma 4.4.3, condition (ii) is satisfied provided that $(2+\gamma)d < 2\gamma p$. $\square$

Proof of Corollary 4.4.1. Follows by Theorem 4.4.1 with $t = C\sigma_n\sqrt{\log(d_1+d_2)}$ for sufficiently large C, and applying the condition $R_n\sqrt{\log(d_1+d_2)} = o(\sigma_n)$. $\square$

Proof of Theorem 4.4.2. By Berbee's lemma (enlarging the probability space as necessary) the processs $\{X_i\}$ can be coupled with a process X_i^* such that $Y_k := \{X_{(k-1)q+1}, \ldots, X_{kq}\}$ and $Y_k^* := \{X^*_{(k-1)q+1}, \ldots, X^*_{kq}\}$ are identically distributed for each $k \geq 1$, $\mathbb{P}(Y_k \neq Y_k^*) \leq \beta(q)$ for each $k \geq 1$ and $\{Y_1^*, Y_3^*, \ldots\}$ are independent and $\{Y_2^*, Y_4^*, \ldots\}$ are independent (see, e.g., Doukhan et al. (1995)). Let I_e and I_o denote the indices of $\{1, \ldots, n\}$ corresponding to the odd- and even-numbered blocks, and I_r the indices in the remainder, so $I_r = q[n/q]+1,\dots,n$ when $q[n/q] < n$ and $I_r = \emptyset$ when $q[n/q] = n$.

Let $\Xi^*_{i,n} = \Xi(X^*_{i,n})$. By the triangle inequality,

$$\begin{aligned}
& \mathbb{P}\left(\|\textstyle\sum_{i=1}^n \Xi_{i,n}\|_2 \geq 6t\right) \\
\leq\; & \mathbb{P}(\|\textstyle\sum_{i=1}^{[n/q]q} \Xi^*_{i,n}\|_2 + \|\sum_{i\in I_r} \Xi_{i,n}\|_2 + \|\sum_{i=1}^{[n/q]q}(\Xi^*_{i,n} - \Xi_{i,n})\|_2 \geq 6t) \\
\leq\; & \tfrac{n}{q}\beta(q) + \mathbb{P}\left(\|\textstyle\sum_{i\in I_r} \Xi_{i,n}\|_2 \geq t\right) + \mathbb{P}\left(\|\textstyle\sum_{i\in I_e} \Xi^*_{i,n}\|_2 \geq t\right) + \mathbb{P}\left(\|\textstyle\sum_{i\in I_o} \Xi^*_{i,n}\|_2 \geq t\right)
\end{aligned}$$

Theorem 4.4.1 is applied to control these last two terms, recognizing that $\sum_{i\in I_e} \Xi^*_{i,n}$ and $\sum_{i\in I_o} \Xi^*_{i,n}$ are each the sum of fewer than $[n/q]$ independent $d_1 \times d_2$ matrices, namely $W^*_k = \sum_{i=(k-1)q+1}^{kq} \Xi^*_{i,n}$ where $\|W^*_k\|_2 \leq qR_n$ and $\max\{\|E[W^*_k W^{*\prime}_k]\|_2, \|E[W^{*\prime}_k W^*_k]\|_2\} \leq q^2 s_n$. Thus Theorem 4.4.1 yields

$$\mathbb{P}\left(\left\|\sum_{i\in I_e} \Xi^*_{i,n}\right\|_2 \geq t\right) \leq (d_1+d_2)\exp\left(\frac{-t^2/2}{nqs_n^2 + qR_n t/3}\right)$$

and similarly for I_o. □

Proof of Corollary 4.4.2. Follows by Theorem 4.4.2 with $t = Cs_n\sqrt{nq\log(d_1+d_2)}$ for sufficiently large C, and applying the conditions $\frac{n}{q}\beta(q) = o(1)$ and $R_n\sqrt{q\log(d_1+d_2)} = o(s_n\sqrt{n})$. □

Proof of Lemma 4.4.1. Let $G_K = E[b^K(X)b^K(X)']$. Since $B_K = clsp\{b_1,\dots,b_K\}$,

$$\begin{aligned}
& \sup\{|\tfrac{1}{n}\textstyle\sum_{i=1}^n b(X_i)^2 - 1| : b \in B_K, E[b(X)^2] = 1\} \\
=\; & \sup\{|c'(B'B/n - G_K)c| : c \in \mathbb{R}^K, \|G_K^{1/2}c\|_2 = 1\} \\
=\; & \sup\{|c'G_K^{1/2}(G_K^{-1/2}(B'B/n)G_K^{-1/2} - I_K)G_K^{1/2}c| : c \in \mathbb{R}^K, \|G_K^{1/2}c\|_2 = 1\} \\
=\; & \sup\{|c'(\widetilde{B}'\widetilde{B}/n - I_K)c| : c \in \mathbb{R}^K, \|c\|_2 = 1\} \\
=\; & \|\widetilde{B}'\widetilde{B}/n - I_K\|_2^2
\end{aligned}$$

as required. □

Proof of Lemma 4.4.2. Follows by Corollary 4.4.1 with $\Xi_{i,n} = n^{-1}(\tilde{b}^K(X_i)\tilde{b}^K(X_i)' - I_K)$, $R_n \lesssim n^{-1}(\zeta_0(K)^2 + 1)$, and $\sigma_n^2 \lesssim n^{-1}(\zeta_0(K)^2 + 1)$. □

Proof of Lemma 4.4.3. Follows by Corollary 4.4.2 with $\Xi_{i,n} = n^{-1}(\tilde{b}^K(X_i)\tilde{b}^K(X_i)' - I_K)$, $R_n \lesssim n^{-1}(\zeta_0(K)^2 + 1)$, and $s_n^2 \lesssim n^{-2}(\zeta_0(K)^2 + 1)$. □

4.7 Supplementary lemmas and their proofs

Huang (2003) provides conditions under which the operator norm of orthogonal projections onto sieve spaces are *stable* in sup norm as the dimension of the sieve space increases. The following Lemma shows the same is true for the operator $Q_K : L^\infty(X) \to L^\infty(X)$ given by

$$Q_K u(x) = \tilde{b}^K(x) D E[\tilde{b}^K(X) u(X)]$$

where $D = S'[SS']^{-1}S$, i.e.

$$\limsup_{K\to\infty} \sup_{u \in L^\infty(X)} \frac{\|Q_K u\|_\infty}{\|u\|_\infty} \le C$$

for some finite positive constant C. The proof follows by simple modification of the arguments in Theorem A.1 in Huang (2003) (see also Corollary A.1 of Huang (2003)).

Lemma 4.7.1. *Q_K is stable in sup norm under Assumption 4.1.1 and 4.1.4.*

Proof of Lemma 4.7.1. The assumptions of Theorem A.1 of Huang (2003) with ν and ν_n taken to be the distribution of X are satisfied under Assumption 4.1.1. Let P_K denote the orthogonal projection onto the sieve space, i.e.

$$P_K u(x) = b^K(x)' E[b^K(X) b^K(X)']^{-1} E[b^K(X) u(X)]$$

for any $u \in L^\infty(X)$. Let $\langle \cdot, \cdot \rangle$ denote the $L^2(X)$ inner product. Since D is an orthogonal projection matrix and P_K is a $L^2(X)$ orthogonal projection onto B_K, for any $u \in L^\infty(X)$

$$\begin{aligned}
\|Q_K u\|^2_{L^2(X)} &= E[u(X)\widetilde{b}^K(X)']D^2 E[\widetilde{b}^K(X)u(X)] \\
&\leq E[u(X)\widetilde{b}^K(X)']E[\widetilde{b}^K(X)u(X)] \\
&= \|P_K u\|^2_{L^2(X)} \\
&\leq \|u\|^2_{L^2(X)}\,.
\end{aligned} \tag{4.21}$$

As in Huang (2003), let Δ index a partition of $\mathcal{X}$ into finitely many polyhedra. Let $v \in L^\infty(X)$ be supported on δ_0 for some $\delta_0 \in \Delta$ (i.e. $v(x) = 0$ if $x \notin \delta_0$). For some coefficients $\alpha_1, \ldots, \alpha_K$,

$$Q_K v(x) = \sum_{i=1}^{K} \alpha_i b_{Ki}(x)\,.$$

Let $d(\cdot,\cdot)$ be the distance measure between elements of Δ defined in the Appendix of Huang (2003). Let l be a nonnegative integer and let $I_l \subset \{1, \ldots, K\}$ be the set of indices such that for any $i \in I_l$ the basis function b_{Ki} is active on a $\delta \in \Delta$ with $d(\delta, \delta_0) \leq l$. Finally, let

$$v_l(x) = \sum_{i \in I_l} \alpha_i b_{Ki}(x)\,.$$

For any $v \in L^\infty(X)$,

$$
\begin{aligned}
& \|Q_K v\|^2_{L^2(X)} \\
= \;& \|Q_K v - v\|^2_{L^2(X)} + \|v\|^2_{L^2(X)} + 2\langle Q_K v - v, v\rangle \\
= \;& \|P_K v - v\|^2_{L^2(X)} + \|Q_K v - P_K v\|^2_{L^2(X)} + 2\langle Q_K v - P_K v, P_K v - v\rangle + \|v\|^2_{L^2(X)} + 2\langle Q_K v - v, v\rangle \\
\leq \;& \|v_l - v\|^2_{L^2(X)} + \|Q_K v - P_K v\|^2_{L^2(X)} + 2\langle Q_K v - P_K v, P_K v - v\rangle + \|v\|^2_{L^2(X)} + 2\langle Q_K v - v, v\rangle \\
= \;& \|v_l - v\|^2_{L^2(X)} + \|v\|^2_{L^2(X)} + \langle Q_K v - P_K v, Q_K v + P_K v - 2v\rangle + 2\langle Q_K v - v, v\rangle \qquad (4.22) \\
= \;& \|v_l - v\|^2_{L^2(X)} + \|v\|^2_{L^2(X)} + \|Q_K v\|^2_{L^2(X)} - \|P_K v\|^2_{L^2(X)} - 2\langle Q_K v - P_K v, v\rangle + 2\langle Q_K v - v, v\rangle \\
= \;& \|v_l - v\|^2_{L^2(X)} + \|v\|^2_{L^2(X)} + \|Q_K v\|^2_{L^2(X)} - \|P_K v\|^2_{L^2(X)} + 2\langle P_K v - v, v\rangle \\
= \;& \|v_l - v\|^2_{L^2(X)} + \|v\|^2_{L^2(X)} + \|Q_K v\|^2_{L^2(X)} + \|P_K v\|^2_{L^2(X)} - 2\|v\|_{L^2(X)} \\
\leq \;& \|v_l - v\|^2_{L^2(X)} + \|v\|^2_{L^2(X)} \qquad (4.23)
\end{aligned}
$$

where (4.22) uses the fact that P_K is an orthogonal projection, and (4.23) follows from (4.21). The remainder of the proof of Theorem A.1 of Huang (2003) goes through under these modifications. $\square$

The next lemma provides useful bounds on the estimated matrices encountered in the body of the chapter. Recall the definitions of $\widehat{D}$ and D in expression (4.17).

Lemma 4.7.2. *Under Assumption 4.1.3(ii) and 4.1.4(ii), if $J \leq K$ and $\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 = o_p(1)$, then wpa1*

(i) $(\widetilde{B}'\widetilde{B}/n)$ *is invertible and* $\|(\widetilde{B}'\widetilde{B}/n)^{-1}\|_2 \leq 2$

(ii) $\|(\widetilde{B}'\widetilde{B}/n)^{-}\widehat{S}' - S'\|_2 \lesssim \|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 + \|\widehat{S} - S\|_2$

(iii) $\|(\widetilde{B}'\widetilde{B}/n)^{-1/2}\widehat{S}' - S'\|_2 \lesssim \|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 + \|\widehat{S} - S\|_2$.

If, in addition, $\sigma_{JK}^{-1}(\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 + \|\widehat{S} - S\|_2) = o_p(1)$, then wpa1

(iv) $(\widetilde{B}'\widetilde{B}/n)^{-1/2}\widehat{S}'$ *has full column rank and* $\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1}\widehat{S}$ *is invertible*

(v) $\|\widehat{D}-D\|_2 \lesssim \|(\widetilde{B}'\widetilde{B}/n)-I_K\|_2 + \sigma_{JK}^{-1}(\|\widehat{S}-S\|_2 + \|(\widetilde{B}'\widetilde{B}/n)-I_K\|_2)$

(vi) $\|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-}\widehat{S}']^{-}\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-} - [SS']^{-1}S\|_2 \lesssim \sigma_{JK}^{-1}\|(\widetilde{B}'\widetilde{B}/n)-I_K\|_2 + \sigma_{JK}^{-2}(\|\widehat{S}-S\|_2 + \|(\widetilde{B}'\widetilde{B}/n)-I_K\|_2)$.

Proof of Lemma 4.7.2. First note that under Assumption 4.1.3(ii) and 4.1.4(ii), $\|S\|_2 \le 1$ since S is isomorphic to the $L^2(X)$ orthogonal projection of T onto the space B_K, restricted to Ψ_J.

(i) Let $\mathscr{A}_n$ denote the event $\|(\widetilde{B}'\widetilde{B}/n)-I_K\|_2 \le \frac{1}{2}$. The condition $\|(\widetilde{B}'\widetilde{B}/n)-I_K\|_2 = o_p(1)$ implies that $\mathbb{P}(\mathscr{A}_n^c) = o(1)$. Clearly $\|(\widetilde{B}'\widetilde{B}/n)^{-1}\|_2 \le 2$ on $\mathscr{A}_n$.

(ii) Working on $\mathscr{A}_n$ (on which the generalized inverse may be replaced with an inverse), Assumption 4.1.4(ii), the triangle inequality, and compatibility of the spectral norm under multiplication yields

$$\begin{aligned}
\|(\widetilde{B}'\widetilde{B}/n)^{-1}\widehat{S}' - S'\|_2 &\le \|(\widetilde{B}'\widetilde{B}/n)^{-1}\widehat{S}' - (\widetilde{B}'\widetilde{B}/n)^{-1}S'\|_2 + \|(\widetilde{B}'\widetilde{B}/n)^{-1}S' - S'\|_2 \\
&\le \|(\widetilde{B}'\widetilde{B}/n)^{-1}\|_2\|\widehat{S}-S\|_2 + \|(\widetilde{B}'\widetilde{B}/n)^{-1} - I_K\|_2\|S'\|_2 \\
&\le 2\|\widehat{S}-S\|_2 + \|(\widetilde{B}'\widetilde{B}/n)^{-1} - I_K\|_2 \\
&= 2\|\widehat{S}-S\|_2 + \|(\widetilde{B}'\widetilde{B}/n)^{-1}[(\widetilde{B}'\widetilde{B}/n) - I_K]\|_2 \\
&\le 2\|\widehat{S}-S\|_2 + 2\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2
\end{aligned}$$

(iii) Follows the same arguments as (ii), noting additionally that $\lambda_{\min}((\widetilde{B}'\widetilde{B}/n)^{-1}) \le \frac{1}{2}$ on $\mathscr{A}_n$, in which case

$$\|(\widetilde{B}'\widetilde{B}/n)^{-1/2} - I_K\|_2 \le (1+2^{-1/2})^{-1}\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 \tag{4.24}$$

by Lemma 2.2 of Schmitt (1992).

(iv) Let $s_J(A)$ denote the Jth largest singular value of a $J \times K$ matrix A. Weyl's inequality yields

$$|s_J(\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1/2}) - \sigma_{JK}| \leq \|(\widetilde{B}'\widetilde{B}/n)^{-1/2}\widehat{S}' - S'\|_2 \,. \tag{4.25}$$

This and the condition $\sigma_{JK}^{-1}(\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 + \|\widehat{S} - S\|_2) = o_p(1)$ together imply that

$$|s_J(\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1/2}) - \sigma_{JK}| \leq \frac{1}{2}\sigma_{JK}$$

wpa1. Let $\mathscr{C}_n$ be the intersection of $\mathscr{A}_n$ with the set on which this bound obtains. Then $\mathbb{P}(\mathscr{C}_n^c) = o(1)$. Clearly $(\widetilde{B}'\widetilde{B}/n)^{-1/2}\widehat{S}'$ has full column rank J and $\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1}\widehat{S}$ is invertible on $\mathscr{C}_n$.

(v) On $\mathscr{C}_n \subseteq \mathscr{A}_n$, $\|(\widetilde{B}'\widetilde{B}/n)^{-1/2}\|_2 \leq \sqrt{2}$. Let

$$\begin{aligned} \widehat{H} &= (\widetilde{B}'\widetilde{B}/n)^{-1/2}\widehat{S}'[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1}\widehat{S}]^{-1}\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1/2} \\ H &= S'[SS']^{-1}S \end{aligned}$$

where $H = D$.

Working on $\mathscr{C}_n$, similar arguments to those used to prove parts (ii) and (iii) yield

$$\begin{aligned} & \|\widehat{D} - D\|_2 \\ \leq\; & \|(\widetilde{B}'\widetilde{B}/n)^{-1/2} - I_K\|_2(\|(\widetilde{B}'\widetilde{B}/n)^{-1/2}\|_2 + 1) + \|\widehat{H} - H\|_2\|(\widetilde{B}'\widetilde{B}/n)^{-1/2}\|_2 \\ \leq\; & (1+\sqrt{2})\|(\widetilde{B}'\widetilde{B}/n)^{-1/2} - I_K\|_2 + \sqrt{2}\|\widehat{H} - H\|_2 \,. \end{aligned} \tag{4.26}$$

Since $\widehat{H}$ and H are orthogonal projection matrices, part (1.5) of Theorem 1.1 of Li et al. (2013) implies

$$\|\widehat{H} - H\|_2 \leq \sigma_{JK}^{-1}\|(\widetilde{B}'\widetilde{B}/n)^{-1/2}\widehat{S}' - S'\|_2 \tag{4.27}$$

on $\mathscr{C}_n$. Part (v) is then proved by substituting (4.27) and (4.24) into (4.26).

(vi) Working on $\mathscr{C}_n$ (and replacing the generalized inverses with inverses), similar arguments used to prove part (v) yield

$$
\begin{aligned}
& \|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1}\widehat{S}']^{-1}\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1} - [SS']^{-1}S\|_2 \\
\leq\ & \|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1}\widehat{S}']^{-1}\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1/2}\|_2\|(\widetilde{B}'\widetilde{B}/n)^{-1/2} - I_K\|_2 \\
& + \|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1}\widehat{S}']^{-1}\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1/2} - [SS']^{-1}S\|_2 \\
\leq\ & 2\sigma_{JK}^{-1}\|(\widetilde{B}'\widetilde{B}/n)^{-1/2} - I_K\|_2 \\
& + \|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1}\widehat{S}']^{-1}\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1/2} - [SS']^{-1}S\|_2 \,. \qquad (4.28)
\end{aligned}
$$

Theorem 3.1 of Ding and Huang (1997) yields

$$
\|[\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1}\widehat{S}']^{-1}\widehat{S}(\widetilde{B}'\widetilde{B}/n)^{-1/2} - [SS']^{-1}S\|_2 \lesssim \sigma_{JK}^{-2}\|(\widetilde{B}'\widetilde{B}/n)^{-1/2}\widehat{S}' - S'\|_2 \quad (4.29)
$$

wpa1 by virtue of part (iii) and the condition $\sigma_{JK}^{-1}(\|(\widetilde{B}'\widetilde{B}/n) - I_K\|_2 + \|\widehat{S} - S\|_2) = o_p(1)$. Substituting (4.29) and (4.24) into (4.28) establishes (vi).

This completes the proof. □

Lemma 4.7.3. *Under Assumption 4.1.4, if $\{X_i, Y_{2i}\}_{i=1}^n$ are i.i.d. then*

(i) $\|\widetilde{B}'(H_0 - \Psi c_J)/n\|_2 \leq O_p(\sqrt{K/n}) \times \|h_0 - \pi_J h_0\|_\infty + \|\Pi_K T(h_0 - \pi_J h_0)\|_{L^2(X)}$

(ii) $\|\widetilde{b}^K(x)D\{\widetilde{B}'(H_0 - \Psi c_J)/n - E[\widetilde{b}^K(X_i)(h_0(Y_{2i}) - \pi_J h_0(Y_{2i}))]\}\|_\infty = \|h_0 - \pi_J h_0\|_\infty \times O_p(\sqrt{K(\log n)/n})$.

Proof of Lemma 4.7.3. We prove Lemma 4.7.3 by part.

(i) First write

$$
\begin{aligned}
\|\widetilde{B}'(H_0 - \Psi c_J)/n\|_2 \ \leq\ & \|\widetilde{B}'(H_0 - \Psi c_J)/n - E[\widetilde{b}^K(X_i)(h_0(Y_{2i}) - \pi_J h_0(Y_{2i}))]\|_2 \\
& + \|E[\widetilde{b}^K(X_i)(h_0(Y_{2i} - \pi_J h_0(Y_{2i}))]\|_2
\end{aligned}
$$

and note that

$$\|E[\widetilde{b}^K(X_i)(h_0(Y_{2i}) - \pi_J h_0(Y_{2i}))]\|_2^2 = \|\Pi_K T(h_0 - \pi_J h_0)\|_{L^2(X)}^2 .$$

Finally,

$$\|\widetilde{B}'(H_0 - \Psi c_J)/n - E[\widetilde{b}^K(X_i)(h_0(Y_{2i}) - \pi_J h_0(Y_{2i}))]\|_2 = O_p(\sqrt{K/n}) \times \|h_0 - \pi_J h_0\|_\infty .$$

by Markov's inequality under Assumption 4.1.4 and the fact that $\{X_i, Y_{2i}\}_{i=1}^n$ are i.i.d.

(ii) An argument similar to the proof of Theorem 4.1.1 converts the problem of controlling the supremum that of controlling the maximum evaluated at finitely many points, where the collection of points has cardinality increasing polynomially in n. Let $\mathcal{S}'_n$ be the set of points. Also define

$$\Delta_{i,J,K} = \widetilde{b}^K(X_i)(h_0(Y_{2i}) - \pi_J h_0(Y_{2i})) - E[\widetilde{b}^K(X_i)(h_0(Y_{2i}) - \pi_J h_0(Y_{2i}))]\} .$$

Then it suffices to show that sufficiently large C may be chosen that

$$(\#\mathcal{S}'_n) \max_{x_n \in \mathcal{S}'_n} \mathbb{P}\left(\left|\sum_{i=1}^n n^{-1}\widetilde{b}^K(x_n) D\Delta_{i,J,K}\right| > C\|h_0 - \pi_J h_0\|_\infty \sqrt{K(\log n)/n}\right) = o(1) . \tag{4.30}$$

The summands in (4.30) have mean zero (by the law of iterated expectations). Under Assumption 4.1.4 the summands in (4.30) are bounded uniformly for $x_n \in \mathcal{S}'_n$ by

$$|n^{-1}\widetilde{b}^K(x_n) D\Delta_{i,J,K}| \lesssim \frac{K}{n}\|h_0 - \pi_J h_0\|_\infty \tag{4.31}$$

and have variance bounded uniformly for $x_n \in \mathcal{S}_n'$ by

$$\begin{aligned}
& E[(n^{-1}\widetilde{b}^K(x_n)D\Delta_{i,J,K})^2] \\
\leq \; & \|h_0 - \pi_J h_0\|_\infty^2 \times n^{-2} E[\widetilde{b}^K(x_n)' D\widetilde{b}^K(X_i)\widetilde{b}^K(X_i)' D\widetilde{b}^K(x_n)] \\
\lesssim \; & \|h_0 - \pi_J h_0\|_\infty^2 \times \frac{K}{n^2} \, .
\end{aligned} \tag{4.32}$$

The result follows for large enough C by Bernstein's inequality for i.i.d. sequences using the bounds (4.31) and (4.32).

This completes the proof. $\square$

Bibliography

Abel, A. B. (1990): "Asset Prices under Habit Formation and Catching up with the Joneses," *American Economic Review*, 80, 38–42.

——— (1999): "Risk Premia and Term Premia in General Equilibrium," *Journal of Monetary Economics*, 43, 3–33.

Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Alvarez, F. and U. J. Jermann (2005): "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73, 1977–2016.

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

——— (2011): "Examples of L2-Complete and Boundedly-Complete Distributions," *Cowles Foundation Discussion Paper No. 1801.*

Attanasio, O. P. and G. Weber (1993): "Consumption Growth, the Interest Rate and Aggregation," *Review of Economic Studies*, 60, 631–649.

——— (1995): "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey," *Journal of Political Economy*, 103, 1121–1157.

Backus, D., M. Chernov, and I. Martin (2011): "Disasters Implied by Equity Index Options," *Journal of Finance*, 66, 1969–2012.

Backus, D., M. Chernov, and S. Zin (2012): "Sources of Entropy in Representative Agent Models," Working paper, New York University.

Bakshi, G. and F. Chabi-Yo (2012): "Variance Bounds on the Permanent and Transitory Components of Stochastic Discount Factors," *Journal of Financial Economics*, 105, 191–208.

Bandtlow, O. F. (2004): "Estimates for Norms of Resolvents and an Application to the Perturbation of Spectra," *Mathematische Nachrichten*, 267, 3–11.

Bansal, R. and B. N. Lehmann (1997): "Growth-Optimal Portfolio Restrictions on Asset Pricing Models," *Macroeconomic Dynamics*, 1, 333–354.

Bansal, R. and S. Viswanathan (1993): "No Arbitrage and Arbitrage Pricing: A New Approach," *The Journal of Finance*, 48, 1231–1262.

Bansal, R. and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Belloni, A., X. Chen, V. Chernozhukov, and K. Kato (2012): "On the Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," Preprint, arXiv:1212.0442v1 [stat.ME].

Bertholon, H., A. Monfort, and F. Pegoraro (2008): "Econometric Asset Pricing Modelling," *Journal of Financial Econometrics*, 6, 407–458.

Beveridge, S. and C. R. Nelson (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, 151–174.

Bickel, P. J. and J. Kwon (2001): "Inference for Semiparametric Models: Some Questions and an Answer (with Discussion)," *Statistica Sinica*, 11, 863–960.

Billingsley, P. (1961): "The Lindeberg-Lévy Theorem for Martingales," *Proceedings of the American Mathematical Society*, 12, 788–792.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613–1669.

Blundell, R. and J. L. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, Cambridge University Press, Cambridge.

Bradley, R. C. (2005): "Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions," *Probability Surveys*, 2, 107–144.

Carroll, R. J. and P. Hall (1988): "Optimal Rates of Convergence for Deconvolving a Density," *Journal of the American Statistical Association*, 83, 1184–1186.

Cavalier, L. (2008): "Nonparametric Statistical Inverse Problems," *Inverse Problems*, 24, 034004.

Cavalier, L., G. K. Golubev, D. Picard, and A. B. Tsybakov (2002): "Oracle Inequalities for Inverse Problems," *The Annals of Statistics*, 30, 843–874.

Cavalier, L. and N. W. Hengartner (2005): "Adaptive Estimation for Inverse Problems with Noisy Operators," *Inverse Problems*, 21, 1345–1361.

Chatelin, F. (1983): *Spectral Approximation of Linear Operators*, Academic Press, New York.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. 6, Part B, chap. 76, 5549–5632.

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014a): “Local Identification of Nonparametric and Semiparametric Models,” *Econometrica*, forthcoming.

Chen, X., J. Favilukis, and S. C. Ludvigson (2013): “An Estimation of Economic Models with Recursive Preferences,” *Quantitative Economics*, 4, 39–83.

Chen, X., L. P. Hansen, and J. A. Scheinkman (2000): “Shape-Preserving Estimation of Diffusions,” Working paper, Univeristy of Chicago.

Chen, X., Z. Liao, and Y. Sun (2014b): “Sieve Inference on Possibly Misspecified Semi-Nonparametric Time Series Models,” *Journal of Econometrics*, 178, 639–658.

Chen, X. and S. C. Ludvigson (2009): “Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models,” *Journal of Applied Econometrics*, 24, 1057–1093.

Chen, X. and D. Pouzo (2009): “Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals,” *Journal of Econometrics*, 152, 46–60.

——— (2012): “Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals,” *Econometrica*, 80, 277–321.

Chen, X. and M. Reiss (2011): “On Rate Optimality for Ill-Posed Inverse Problems in Econometrics,” *Econometric Theory*, 27, 497–521.

Cochrane, J. H. (1992): “Explaining the Variance of Price-Dividend Ratios,” *Review of Financial Studies*, 5, 243–280.

Cohen, A., I. Daubechies, B. Jawerth, and P. Vial (1993a): “Multiresolution Analysis, Wavelets and Fast Algorithms on an Interval,” *Comptes Rendus de l’Académie des Sciences Paris. Série 1, Mathématique*, 316, 417–421.

Cohen, A., I. Daubechies, and P. Vial (1993b): “Wavelets on the Interval and Fast Wavelet Transforms,” *Applied and Computational Harmonic Analysis*, 1, 54–81.

Cohen, A., M. Hoffmann, and M. Reiss (2004): "Adaptive Wavelet Galerkin Methods for Linear Inverse Problems," *SIAM Journal on Numerical Analysis*, 42, 1479–1501.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79, 1541–1565.

Darolles, S., C. Gourieroux, and J. Jasiak (2006): "Structural Laplace Transform and Compound Autoregressive Models," *Journal of Time Series Analysis*, 27, 477–503.

De Boor, C. (2001): *A Practical Guide to Splines*, Springer-Verlag, New York.

de Jong, R. M. (2002): "A Note on "Convergence Rates and Asymptotic Normality for Series Estimators": Uniform Convergence Rates," *Journal of Econometrics*, 111, 1–9.

D'Haultfoeuille, X. (2011): "On the Completeness Condition in Nonparametric instrumental regression," *Econometric Theory*, 27, 460–471.

Ding, J. and L. Huang (1997): "On the Continuity of Generalized Inverses of Linear Operators in Hilbert Spaces," *Linear Algebra and its Applications*, 262, 229–242.

Doukhan, P. (1994): *Mixing: Properties and Examples*, Springer, New York.

Doukhan, P., P. Massart, and E. Rio (1995): "Invariance Principles for Absolutely Regular Empirical Processes," *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, 31, 393–427.

Dunford, N. and J. T. Schwartz (1958): *Linear Operators, Part I: General Theory*, Interscience Publishers, New York.

Efromovich, S. and V. Koltchinskii (2001): "On Inverse Problems with Unknown Operators," *IEEE Transactions on Information Theory*, 47, 2876–2894.

Eraker, B. (2008): "Affine General Equilibrium Models," *Management Science*, 54, 2068–2080.

Escanciano, J. C. and S. Hoderlein (2012): “Nonparametric Identification of Euler Equations,” Working paper, Indiana University.

Fan, J. (1991): “On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems,” *The Annals of Statistics*, 19, 1257–1272.

Fleissig, A. R., A. R. Gallant, and J. J. Seater (2000): “Separability, Aggregation and Euler Equation Estimation,” *Macroeconomic Dynamics*, 4, 547–572.

Florens, J.-P. and A. Simoni (2012): “Nonparametric estimation of an instrumental variables regression: a quasi-Bayesian approach based on regularized posterior,” *Journal of Econometrics*, 170, 458–475.

Gagliardini, P. and O. Scaillet (2012): “Tikhonov Regularization for Nonparametric Instrumental Variable Estimators,” *Journal of Econometrics*, 167, 61–75.

Galí, J. (1994): “Keeping up with the Joneses: Consumption Externalities, Portfolio Choice, and Asset Prices,” *Journal of Money, Credit and Banking*, 26, 1–8.

Gallant, A. R. and G. Tauchen (1989): “Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications,” *Econometrica*, 57, 1091–1120.

Gobet, E., M. Hoffmann, and M. Reiss (2004): “Nonparametric Estimation of Scalar Diffusions Based on Low Frequency Data,” *The Annals of Statistics*, 32, 2223–2253.

Gourieroux, C. and J. Jasiak (2006): “Autoregressive Gamma Processes,” *Journal of Forecasting*, 25, 129–152.

Greenwood, P. E., A. Schick, and W. Wefelmeyer (2001): “Comment [on Bickel and Kwon, 2001],” *Statistica Sinica*, 11, 892–906.

Greenwood, P. E. and W. Wefelmeyer (1995): “Efficiency of Empirical Estimators for Markov Chains,” *Annals of Statistics*, 32, 132–143.

Hall, P. and J. L. Horowtiz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics*, 33, 2904–2929.

Hall, P. and A. Meister (2007): "A Ridge-Parameter Approach to Deconvolution," *The Annals of Statistics*, 35, 1535–1558.

Hansen, L. P. (2012): "Dynamic Valuation Decomposition Within Stochastic Economies," *Econometrica*, 80, 911–967.

Hansen, L. P., J. C. Heaton, and N. Li (2008): "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 116, 260–302.

Hansen, L. P. and E. Renault (2010): *Encyclopedia of Quantitative Finance*, John Wiley & Sons, Ltd, chap. Pricing Kernels.

Hansen, L. P. and S. F. Richard (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55, 587–613.

Hansen, L. P. and J. A. Scheinkman (2009): "Long-Term Risk: An Operator Approach," *Econometrica*, 77, 177–234.

——— (2012a): "Pricing Growth-Rate Risk," *Finance and Stochastics*, 16, 1–15.

——— (2012b): "Recursive Utility in a Markov Environment with Stochastic Growth," *Proceedings of the National Academy of Sciences*, 109, 11967–11972.

——— (2013): "Stochastic Compounding and Uncertain Valuation," Working paper, University of Chicago.

Hansen, L. P. and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

Hoffmann, M. and M. Reiss (2008): "Nonlinear Estimation for Linear Inverse Problems with Error in the Operator," *The Annals of Statistics*, 36, 310–336.

Horowitz, J. L. (2011): “Applied Nonparametric Instrumental Variables Estimation,” *Econometrica*, 79, 347–394.

Huang, J. Z. (1998): “Projection Estimation in Multiple Regression with Application to Functional ANOVA Models,” *The Annals of Statistics*, 26, 242–272.

——— (2003): “Local Asymptotics for Polynomial Spline Regression,” *The Annals of Statistics*, 31, 1600–1635.

Jansson, M. (2004): “The Error in Rejection Probability of Simple Autocorrelation Robust Tests,” *Econometrica*, 72, 937–946.

Johnstone, I. M. (2013): “Gaussian Estimation: Sequence and Wavelet Models,” Manuscript.

Julliard, C. and A. Ghosh (2012): “Can Rare Events Explain the Equity Premium Puzzle?” *Review of Financial Studies*, 25, 3037–3076.

Kato, T. (1980): *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin.

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2000): “Simple Robust Testing of Regression Hypotheses,” *Econometrica*, 68, 695–714.

Kress, R. (1999): *Linear Integral Equations*, Springer-Verlag, Berlin.

Kreĭn, M. G. and M. A. Rutman (1950): *Linear Operators Leaving Invariant a Cone in a Banach Space*, American Mathematical Society, New York.

Lee, J. and P. Robinson (2013): “Series Estimation Under Cross-sectional Dependence,” Preprint, London School of Economics.

Li, B., W. Li, and L. Cui (2013): “New Bounds for Perturbation of the Orthogonal Projection,” *Calcolo*, 50, 69–78.

Liao, Y. and W. Jiang (2011): “Posterior Consistency of Nonparametric Conditional Moment Restricted Models,” *The Annals of Statistics*, 39, 3003–3031.

Linton, O. B., A. Lewbel, and S. Srisuma (2011): "Nonparametric Euler Equation Identification and Estimation," Working paper, Boston College.

Longstaff, F. A. and M. Piazzesi (2004): "Corporate Earnings and the Equity Premium," *Journal of Financial Economics*, 74, 401–421.

Loubes, J.-M. and C. Marteau (2012): "Adaptive Estimation for an Inverse Regression model with Unknown Operator," *Statistics & Risk Modeling*, 29, 215–242.

Lounici, K. and R. Nickl (2011): "Global Uniform Risk Bounds for Wavelet Deconvolution Estimators," *The Annals of Statistics*, 39, 201–231.

Mehra, R. and E. C. Prescott (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

Monfort, A. and F. Pegoraro (2007): "Switching VARMA Term Structure Models," *Journal of Financial Econometrics*, 5, 105–153.

Müller, U. K. (2007): "A Theory of Robust Long-Run Variance Estimation," *Journal of Econometrics*, 141, 1331–1352.

Newey, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168.

Newey, W. K. and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Parzen, E. (1957): "On Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics*, 28, 329–348.

Peligrad, M. (1985): "An Invariance Principle for Phi-Mixing Sequences," *Annals of Probability*, 13, 1304–1313.

PHILLIPS, P. C. B. (2005): "HAC Estimation by Automated Regression," *Econometric Theory*, 21, 116–142.

ROSENBLATT, M. (1971): *Markov Processes. Structure and Asymptotic Behavior*, Springer-Verlag, Berlin.

ROSS, S. A. (2013): "The Recovery Theorem," *Journal of Finance*, forthcoming.

SCHAEFER, H. H. (1960): "Some Spectral Properties of Positive Linear Operators," *Pacific Journal of Mathematics*, 10, 1009–1019.

——— (1974): *Banach Lattices and Positive Operators*, Springer-Verlag, Berlin.

——— (1999): *Topological Vector Spaces*, Springer-Verlag, New York.

SCHMITT, B. A. (1992): "Perturbation Bounds for Matrix Square Roots and Pythagorean Sums," *Linear Algebra and its Applications*, 174, 215–227.

SCHUMAKER, L. L. (2007): *Spline Functions: Basic Theory*, Cambridge University Press, Cambridge.

SONG, K. (2008): "Uniform Convergence of Series Estimators over Function Spaces," *Econometric Theory*, 24, 1463–1499.

STONE, C. J. (1982): "Optimal Global Rates of Convergence for Nonparametric Regression," *Annals of Statistics*, 10, 1040–1053.

SUN, Y., P. C. B. PHILLIPS, AND S. JIN (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, 76, 175–194.

TRIEBEL, H. (2006): *Theory of Function Spaces III*, Birkhäuser, Basel.

——— (2008): *Function Spaces and Wavelets on Domains*, European Mathematical Society, Zürich.

Tropp, J. A. (2012): "User-Friendly Tail Bounds for Sums of Random Matrices," *Foundations of Computational Mathematics*, 12, 389–434.

Tsybakov, A. B. (2009): *Introduction to Nonparametric Estimation*, Springer, New York.

van de Geer, S. (1995): "Exponential Inequalities for Martingales, with Application to Maximum Likelihood Estimation for Counting Processes," *The Annals of Statistics*, 23, 1779–1801.

Wefelmeyer, W. (1999): "Efficient Estimation in Markov Chain Models: An Introduction," in *Asymptotics, Nonparametrics and Time Series*, ed. by S. Ghosh, Dekker, New York, no. 158 in Statistics: Textbooks and Monographs.

Weiss, G. (1975): "Time-Reversibility of Linear Stochastic Processes," *Journal of Applied Probability*, 12, 831–836.

Zhang, C.-H. (1990): "Fourier methods for estimating mixing densities and distributions," *The Annals of Statistics*, 18, 806–831.