Functional Regression of Continuous State Distributions

Joon Y. Park
Department of Economics
Texas A&M University

Junhui Qian
School of Economics
Shanghai Jiao Tong University

Abstract

In this paper we propose a regression model that addresses the problem of distributional relationship between two economic variables. Unlike the classical linear regression, which deals with regressive relationships in the mean only, the proposed model describes and can be used to test on dependence structure between entire distributions. We focus on the case when density functions exist for both left and right hand side variables and are time varying. The case when the variables of interest can be characterized by discrete distributions reduces to traditional vector regression under our framework. Technically, we treat density functions as random elements taking values in the Hilbert spaces of square integrable functions on a compact interval. The regression relationship is described by a compact linear operator mapping from one Hilbert space, where the right hand side density functions reside, to another Hilbert space where left hand side densities reside. We describe how we estimate the model and establish consistency of our estimator. And we develop a hypothesis testing procedure and derive the asymptotic distribution for our test statistic. We investigate finite sample performance of our tests using Monte Carlo simulations. In the end of the paper we offer an empirical illustration of our methodology.

This version: August 4, 2008

JEL classification codes: C12, C13, C22.

Key words and phrases: functional regression, time-varying density, moment dependence

1This paper is part of Junhui Qian’s Ph.D. dissertation at Rice University. We like to thank Yoosoon Chang for many helpful discussions. The research of Joon Y. Park was supported by NSF Grant SES-0518619, which is gratefully acknowledged.
1. Introduction

Linear regression is by far the most popular statistical model applied in economics. Graphically speaking, it assumes the means of dependent variables conditional on the independent variables fall on a line or a linear surface. The linear least square can be readily applied to estimate this nice by often surreal relationship.

To get a more complete, hence more robust, picture of the dependent variables, Koenker and Bassett (1978) proposed quantile regression, which examines different quantiles of the dependent variable conditional on the independent variables. While the least square of linear regression minimizes the sum of square errors of estimation, the estimation of quantile regression boils down to the minimization of asymmetrically weighted sum of absolute error.

In both linear regression and quantile regression, the independent variables that are conditioning the dependent variables are scalars. In economics, however, we are often interested in explaining economic variables conditioning on a distribution. As a Chinese saying goes, scarceness does not matter, inequality does. We may want to look at, for example, how consumptions in an economy depends upon the distribution of income or wealth. In finance, also, distributions of returns, whether of one stock/index across a time span or of different stocks in the market or a cohort, carry important information that are relevant to pricing and risk control.

This paper proposes a regression model that promises to accommodate such need or interest. The regressand in the proposed model is a density function of a distribution. This is in the same spirit of quantile regressions, as the density is equivalent to the quantile function. More importantly, the regressor in the model is a time-varying density function, instead of a scalar variable in linear and quantile regressions.

Technically, we regard the time-varying densities as a functional stochastic process, each element of which is a random density taking value in the Hilbert space of square integrable functions. And the regression relation, between the two sequences of densities in the model, is captured by a linear operator from one Hilbert space where the regressor take value to another Hilbert space where the regressand is defined. This linear operator carries all the information on the dependence of the regressand distribution on the regressor. With the hypothesis testing strategy proposed also in this chapter, we are able to test on the moment dependence structure between the two distributions.

The proposed regression model, which we call functional regression (FR), may be applied to general function-valued random variables. To our best knowledge, this model has not been studied in the literature. The FR incorporates functional autoregression (FAR) as a special case, which is well studied in statistical literature (See, eg, Bosq (2000)). Many techniques developed under FAR framework can be adapted to the FR model, under weaker conditions.

In the following, we first present our model and explore its implications and properties. Then we outline the estimation procedure and provide the consistency results in each stage of estimation. Next we present a hypothesis testing strategy in our framework. We provide a test statistic and its asymptotic distribution.
2. The Model and Preliminaries

In this section, we present our model and some preliminaries that are necessary for the development of our theory and methodology in the paper.

2.1 The Model

Denote by $L^2(C)$ the Hilbert spaces of square integrable functions on a compact subset $C$. We consider two sequences of probability densities $(g_t, f_t)$. We regard $(g_t)$ and $(f_t)$ as random, taking values in $H_1 = L^2(C_1)$ and $H_2 = L^2(C_2)$, respectively. We call $H$-valued random variables the random elements taking value in $H$. A sequence of $H$-valued r.v.’s is called an $H$-stochastic process.

We consider the following functional regression (FR) model,

$$g_t = c + Af_t + \varepsilon_t, \quad t = 1, 2, \ldots, T,$$

where $c$ is a nonrandom element in $H_1$, $A$ is an operator from $H_2$ to $H_1$, and $(\varepsilon_t)$ is a sequence of $H_1$-valued functional white noise process. The operator $A$ is corresponding to the regression coefficient in classical linear regressions, and the nonrandom component $c$ is corresponding to the constant term. We may suppress the constant term, but that would put unnecessary restrictions on $A$. And as in classical regression, we may call $(f_t)$ the regressor or independent density, and $(g_t)$ the regressand or dependent density.

We restrict our discussion to the case when both $(f_t)$ and $(g_t)$ are infinite-dimensional, that is, there is no way to represent $(f_t)$ and $(g_t)$ by a finite basis. Otherwise the model in (1) reduces to a vector regression.

The regression relation between $(g_t)$ and $(f_t)$ necessarily implies regression relation in moments of the two sequences of distributions. To see this, we fix $v \in H_1$ and consider

$$\langle v, g_t \rangle = \langle v, c \rangle + \langle v, A f_t \rangle + \langle v, \varepsilon_t \rangle = \langle v, c \rangle + \langle A^* v, f_t \rangle + \eta_t,$$

where $A^*$ is the adjoint of $A$ and $\eta_t = \langle v, \varepsilon_t \rangle$ for all $t$. Under the assumptions on $(\varepsilon_t)$, $(\eta_t)$ is a white noise for any choice of $v \in H_1$.

Now if we let $p_k(x) = x^k$, the $k$-th polynomial, and let $v = p_1$, then the regressand in (2) becomes

$$\int_{C_1} x g_t(x) \, dx,$$

i.e., the mean of the distribution represented by $g_t$. In general, we have $A^* v \neq v = p_1$, and if we write $A^* v = A^* p_1 = \sum_{k=1}^{\infty} c_k p_k$ with some numerical sequence $(c_k)$, then the non-constant regressor in (2) reduces to

$$\sum_{k=1}^{\infty} c_k \int_{C_2} x^k f_t(x) \, dx,$$

i.e., a linear combination of all moments of the distribution represented by $f_t$.

It is worth mentioning that although many well known density functions have supports on the whole real line, they can be approximated by densities in $L^2(C)$ considerably well
if in particular we take \( C \) to be a large subset of \( \mathbb{R} \). From a practical perspective, there is seldom a distribution in economics and finance applications that could have unbounded support. Note also that we do not require them to vanish at the boundary. Hence even when we have to think of densities as with unbounded support, we may interpret our model as an FR model of truncated distributions.

The following assumptions will be maintained throughout the paper.

**Assumption 1**  We assume

(a) \( A \) is a compact linear operator,
(b) \((g_t, f_t)\) are jointly stationary and geometrical strong mixing,
(c) \((\varepsilon_t)\) are iid such that \( \mathbb{E}\varepsilon_t = 0 \) and \( \mathbb{E}\|\varepsilon_t\|^4 < \infty \), and independent of \((f_t)\).

A linear operator \( A \) from \( H_2 \) to \( H_1 \) is said to be compact if it can be written as

\[
A = \sum_{k=1}^{\infty} \lambda_k (u_k \otimes v_k)
\]  

where \((u_k)\) and \((v_k)\) are some orthonormal bases of \( H_1 \) and \( H_2 \), respectively, \((\lambda_k)\) are a sequence of numbers tending to zero, and \( \otimes \) denotes the tensor product defined as \( u \otimes v(\cdot) \equiv \langle v, \cdot \rangle u \). The compact linear operator \( A \) is called nuclear if \( \sum_{k=1}^{\infty} |\lambda_k| < \infty \), and Hilbert-Schmidt if \( \sum_{k=1}^{\infty} \lambda_k^2 < \infty \).

Note that if both \( H_1 \) and \( H_2 \) are finite dimensional, say, \( H_1 = \mathbb{R}^m \) and \( H_2 = \mathbb{R}^n \), then \( A \) would be an \( m \) by \( n \) matrix, and the representation in (3) would be nothing but singular value decomposition. In fact many features of the matrix theory for finite dimensional vector space are generalized to the compact linear operator in a nice way, and consequently, we may regard \( A \) essentially as an infinite dimensional matrix.

The conditions on \((g_t, f_t)\) may be relaxed for some results developed later in this chapter. For example, instead of assuming strong stationarity, \((g_t, f_t)\) can be assumed to be weakly stationary, in that the mean and covariance operators do not change with time. The geometrical strong mixing (GSM) may also be relaxed to strong mixing for some of the results later in the chapter. The same is true for the iid condition on \((\varepsilon_t)\). Many of our subsequent results hold for the sequences that are only serially uncorrelated.

Next we provide some preliminaries on stochastic processes in functional space. The meaning of above conditions will be made more clear.

### 2.2 Some Preliminaries

To define the distribution of an \( H \)-valued more formally, we let \( w : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow H \), where \((\Omega, \mathcal{F}, \mathbb{P})\) is the underlying probability space, and assume that \( \langle v, w \rangle \) is a random variable (and hence \( \mathbb{P} \)-measurable) for any \( v \in H \). The distribution of \( w \) is then completely characterized by the joint distribution of \( \langle v_1, w \rangle, \ldots, \langle v_n, w \rangle \) for an arbitrary selection of \( v_1, \ldots, v_n \in H \).
The expected value of an $H$-valued random variable is defined by a Pettis type integral. We say that a random element $w$ taking values in $H$ has mean value $Ew$, if $\langle v, w \rangle$ has finite expectation and if there exists an element $Ew$ in $H$ satisfying

$$\langle v, Ew \rangle = E\langle v, w \rangle,$$

(4)

for all $v \in H$. It is known that $Ew$ exists if and only if $E\|w\| < \infty$. From the definition we can further infer in particular that if $Ew$ exists and $A$ is bounded linear operator on $H$, then $EAw$ exists and $EAw = AEw$. We may also easily deduce that $\|Ew\| \leq E\|w\|$.

The expectation operation $E$ on the random elements taking values in $H$ satisfy many well known properties of the usual expectation.

By stationarity, $(g_t)$ have a common expectation denoted as $Eg \in H_1$ and $(f_t)$ have a common expectation $Ef \in H_2$. Under the assumption on $(\varepsilon_t)$, we may write the regression in (1) in the mean-correction form,

$$m_t = A(w_t) + \varepsilon_t,$$

(5)

where

$$m_t = g_t - Eg, \quad w_t = f_t - Ef.$$

(6)

Next we introduce the covariance operator for two random elements $m$ and $w$ taking values in $H_1$ and $H_2$, respectively. Recall that for vector-valued random variables, the second moments are covariance matrices, which are finite-dimensional linear operators. The covariance operator of $m$ and $w$, which we denote by $E(m \otimes w)$, is an operator from $H_2$ to $H_1$:

$$E(m \otimes w)v = E\langle w, v \rangle m$$

(7)

for all $v \in H_2$. Note that $E\langle w, v \rangle m$ in the right hand side of (7) is defined to be an element in $H_1$ such that

$$\langle u, E\langle w, v \rangle m \rangle = E\langle u, \langle w, v \rangle m \rangle = E\langle u, m \rangle \langle w, v \rangle$$

for all $u \in H_1$, according to (4) and the linearity of the inner product. Naturally, we may call $E(w \otimes w)$ the variance operator of $w$. It is well known that $E(m \otimes w)$ is compact and

$$\|E(m \otimes w)\| \leq \|E\|\|m\|\|w\|$$

for any random elements $m$ and $w$ taking values in $H_1$ and $H_2$, respectively. Furthermore, $E(w \otimes w)$ is self-adjoint and nonnegative (strictly positive if $w$ is non-degenerate).

For brevity of exposition, we will denote

$$Q = E(w_t \otimes w_t), \quad P = E(m_t \otimes w_t), \quad W = E(m_t \otimes m_t)$$

(8)

in our subsequent development. Note that the Assumption 1(b) ensures that none of $Q$, $P$, and $W$ depends on $t$. $Q$ and $W$ are of course self-adjoint and nonnegative and allow for the following spectral representations,

$$Q = \sum_{k=1}^{\infty} \lambda_k (v_k \otimes v_k) \quad \text{and} \quad W = \sum_{k=1}^{\infty} \gamma_k (u_k \otimes u_k)$$

(9)
where \((\lambda_k, v_k)\) and \((\gamma_k, u_k)\) are the pairs of eigenvalue and eigenvector of \(Q\) and \(W\), respectively. Furthermore, it is clear that

\[ P = AQ, \]  

and this relationship will be used to estimate \(A\).

One of the difficulties in analyzing FR is that it is generally impossible to use the relationship in (10) and define the autoregressive operator \(A\) as \(A = PQ^{-1}\). This will be explained below. As is clear from the spectral representation of \(Q\) in (9), if the kernel of \(Q\) is \(\{0\}\) and \(\lambda_k \neq 0\) for all \(k\), then \(Q^{-1}\) is well defined and given by \(Q^{-1} = \sum_{k=1}^{\infty} \lambda_k^{-1}(v_k \otimes v_k)\).

However, even in this case, \(Q^{-1}\) is not defined on the entire space \(H_2\). Indeed, its domain is restricted to \(D(Q^{-1}) = \{u \in H_2 \mid \sum_{k=1}^{\infty} \langle u, v_k \rangle^2/\lambda_k^2 < \infty \}\), which is a proper subset of \(H_2\).

Consequently, we have \(PQ^{-1} = A\) only on the restricted domain \(D(Q^{-1})\). This problem is often referred to an ill-posed inverse problem.

The standard method to circumvent this problem used in the literature is to restrict the definition of \(A\) in a finite subset of \(H_2\). To explain the details of this method, we let

\[ \lambda_1 > \lambda_2 > \cdots > 0 \]  

and define \(V_K\) to be the subspace of \(H_2\) spanned by the \(K\)-eigenvectors \(v_1, \ldots, v_K\) associated with the eigenvalues \(\lambda_1, \ldots, \lambda_K\). We also denote by \(\Pi_K\) the projection on \(V_K\). Subsequently, we let \(Q_K = \Pi_K Q \Pi_K\) and define

\[ Q_K^+ = \sum_{k=1}^{K} \lambda_k^{-1}(v_k \otimes v_k), \]  

e.i., the inverse of \(Q\) on \(V_K\).

Now we let

\[ A_K = PQ_K^+, \]  

which is the autoregressive operator \(A\) restricted to the subspace \(V_K\) of \(H_2\). Note that \(V_K\) is generated by the \(K\)-principal components of \((w_t)\), i.e., it is the subspace of \(H_2\) spanned by vectors yielding \(K\)-largest variations in \((w_t)\). Since \((\lambda_k)\) decreases down to zero, we may well expect that \(A_K\) approximates \(A\) well if the dimension \(K\) of \(V_K\) increases. The estimator of \(A\), which will be introduced in the next section, is indeed the sample analogue estimator of \(A_K\) defined in (13), and we let \(K\) increase as \(T\) increases.

3. Estimation

In most applications, densities are not directly observed and should therefore be estimated before we look at the FR model specified in (1) or (5) and (6). We suppose that \(N\) observations from the probability density \(g_t\) or \(f_t\) are available so that we may estimate \(g_t\) and \(f_t\) consistently for each \(t = 1, 2, \ldots, T\). In what follows, we denote by \((\hat{g}_t)\) and \((\hat{f}_t)\) the consistent estimators for \((g_t)\) and \((f_t)\), respectively. We define \(\Delta_{g,t} = \hat{g}_t - g_t\) and \(\Delta_{f,t} = \hat{f}_t - f_t\) for \(t = 1, \ldots, T\). Note that this notation suppresses the dependence of \(\hat{g}\) and \(\hat{g}\) on \(N\). As before, we let \((\lambda_k)\) and \((\gamma_k)\) be the ordered eigenvalues of \(Q\) and \(W\), respectively.
Assumption 2. We assume
(a) $\lambda_k > 0$ for all $k$,
(b) $\|f_t\|, \|g_t\| \leq M$ a.s. for some constant $M > 0$,
(c) $\|\Delta_t\| \leq M$ a.s. and $\sup_{t \geq 1} E\|\Delta_t\|^2 = O(N^{-r})$ for some $r > 0$, and
(d) $E\langle v_k, w_t \rangle^4 \leq M\lambda_k^2$, $k \geq 1$, for some constant $M > 0$ independent of $k$.

As long as the distribution of $(f_t)$ is non-degenerate, the condition in part (a) holds. The boundedness of $(g_t)$, $(f_t)$, and their estimates in part (b) is expected to hold for most practical applications. The condition in part (c) is also satisfied in many cases. For instance, if the densities are estimated by the usual kernel estimator from $N$ observations, we may expect for all $t$ that $E\|\Delta_t\|^2 = O(N^{-r})$ for some $0 < r < 2/5$ under very general regularity conditions which allow in particular for dependency among $N$ observations. See, e.g., Bosq (1998, Theorem 2.2). The condition in (d) is equivalent to $E\xi_k^4 < M$, where $M$ is a positive constant independent of $k$ and where $\xi_k$ is from the well known Karhunen-Loève (KL) decomposition of random element $w$,

$$w = \sum_{k=1}^{\infty} \xi_k \sqrt{\lambda_k} v_k. \quad (14)$$

Here $=d$ denotes equality of distributions and $\xi_k$’s are uncorrelated real random variables with zero mean and unit variance. Condition (d) is met whenever the tail probability of $\langle v_k, w_t \rangle$ decreases fast enough. For instance, if $\langle v_k, w_t \rangle$ is Gaussian, the condition holds with $M = 3$. Note that $\langle v_k, w_t \rangle$ has mean zero and variance $\lambda_k$ for $k = 1, 2, \ldots$. It may not hold if the distribution of $\langle v_k, w_t \rangle$ for some $k$ has the thicker tail.

3.1 Mean and Autocovariance Operators

First we estimate $Eg$ and $Ef$ by their sample averages. For the brevity of exposition we focus on $Ef$ only. The results derived in this subsection applies to $Eg$. We estimate $Ef$ by

$$\hat{f} = \frac{1}{T} \sum_{t=1}^{T} \hat{f}_t. \quad (15)$$

We may deduce that

Lemma 1. Let Assumptions 1 and 2 hold. If $N \geq cT^{1/r}$ for some constant $c > 0$, then we have

$$E\|\hat{f} - Ef\|^2 = O(T^{-1})$$

for large $T$. Furthermore, if there exist $a > 0$ and $0 < b < 1$ such that $\lambda_k \leq ab^k$, and if $N$ is such that $N > cT^{2/r} \log^{s-2/r} T$ for some constants $c > 0$ and $s > 0$, we have

$$\|\hat{f} - Ef\| = o(T^{-1/2} \log^{3/2} T) \quad \text{a.s.}$$

for large $T$. 

Lemma 1 establishes the $L^2$ and a.s. consistency for the sample mean. It shows in particular that using estimated density does not affect the convergence rates as long as the number $N$ of observations that we use to estimate the densities is sufficiently large.

The condition of GSM is stronger than necessary for the proof of $L^2$ consistency. Under strong mixing, for example, we may prove the same result if we assume $\alpha(k) = O(k^{-2})$, where $\alpha(k)$’s are the strong mixing coefficients.

The rate of a.s. convergence may be higher if we impose more restrictions on $(f_t)$. If, for instance, $(f_t)$ follows a FAR(1) process introduced in the previous chapter, then we would have

$$\|\bar{f} - \mathbb{E} f\| = o(T^{-1/2} \log^{1/2} T) \text{ a.s.}$$

If $(f_t)$ are iid, the same optimal a.s. convergence rate can also be attained.

Similarly, we may use the sample analogue estimators to consistently estimate the co-variance operators $Q$ and $P$ introduced in (8). Let

$$\hat{m}_t = \hat{g}_t - \bar{g} \quad \text{and} \quad \hat{w}_t = \hat{f}_t - \bar{f}$$

for $t = 1, 2, \ldots$, and define

$$\hat{Q} = \frac{1}{T} \sum_{t=1}^{T} (\hat{w}_t \otimes \hat{w}_t) \quad \text{and} \quad \hat{P} = \frac{1}{T} \sum_{t=1}^{T} (\hat{m}_t \otimes \hat{w}_t).$$

(16)

Then we have

**Theorem 2** Let Assumptions 1 and 2 hold. If $N \geq cT^{1/r}$ for some constant $c$, then

$$\mathbb{E}\|\hat{Q} - Q\|^2, \mathbb{E}\|\hat{P} - P\|^2 = O(T^{-1})$$

for large $T$. Moreover, if there exist $a > 0$ and $0 < b < 1$ such that $\gamma_k \leq ab^k$ and $\lambda_k \leq ab^k$, and if $N$ is such that $N > cT^{2/r} \log^{k-2/r} T$ for some constants $c > 0$ and $s > 0$, then

$$\|\hat{Q} - Q\|, \|\hat{P} - P\| = o(T^{-1/2} \log^{3/2} T) \text{ a.s.}$$

for large $T$.

Theorem 2 establishes the $L^2$ consistency of $\hat{Q}$ and $\hat{P}$. The same result certainly holds for $\hat{W}$. Again, as in Lemma 3.1, the rate of a.s. convergence may be higher if we impose more restrictions on $(g_t, f_t)$ such as iid or FAR(1).

### 3.2 Eigenvalues and Eigenvectors

The implementation of our methodology requires the estimation of eigenvalues and eigenvectors of $Q$, which can be consistently estimated by $\hat{Q}$ defined in (16). Naturally, the eigenvalues and eigenvectors of $Q$ are estimated by those of $\hat{Q}$, which we denote by $(\hat{\lambda}_k, \hat{v}_k)$. 
We order \((\hat{\lambda}_k)\) so that \(\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots\) as in (11). The eigenvalues and eigenvectors of \(\hat{Q}\) satisfy the condition \(\hat{Q}v = \lambda v\), i.e.,

\[
\int_C \hat{V}(x,y)v(x)\,dx = \lambda v(y),
\]

(17)

where \(\hat{V}(x,y)\) is given by

\[
\hat{V}(x,y) = \frac{1}{T} \sum_{t=1}^{T} \hat{w}_t(x)\hat{w}_t(y).
\]

Note that (17) is the first order condition for the maximization problem

\[
\max_{\|v\|=1} \frac{1}{T} \sum_{t=1}^{T} \left( \int_C \hat{w}_t(x)v(x)\,dx \right)^2,
\]

which appears in the functional principal component analysis (FPCA).

We may solve (17) either by directly discretizing \((\hat{w}_t)\) or expanding them into a linear combination of orthonormal bases of \(H\). In either way we transform the original problem into the problem of computing eigenvalues and eigenvectors of a matrix. Readers are referred to Ramsay and Silberman (1997) for more details.

Let us assume that the eigensubspaces \((V_k)\) corresponding to \((\lambda_k)\) are one dimensional, and we define \(v'_k = \text{sgn}(\hat{v}_k, v_k)v_k\). Note that since both \((v_k)\) and \((-v_k)\) are eigenvectors corresponding to \((\lambda_k)\), the introduction of \((v'_k)\) is essential for definitiveness of eigenvectors.

The following lemma from Bosq (2000) associates \(|\hat{\lambda}_k - \lambda_k|\) and \(\|\hat{v}_k - v'_k\|\) to \(\|\hat{Q} - Q\|\).

**Lemma 3** We have

\[
\sup_{k \geq 1} |\hat{\lambda}_k - \lambda_k| \leq \|\hat{Q} - Q\|,
\]

\[
\|\hat{v}_k - v'_k\| \leq \tau_k \|\hat{Q} - Q\|,
\]

where \(\tau_1 = 2\sqrt{2}(\lambda_1 - \lambda_2)^{-1}\) and \(\tau_k = 2\sqrt{2}\max\{(\lambda_{k-1} - \lambda_k)^{-1}, (\lambda_k - \lambda_{k+1})^{-1}\}\) for \(k \geq 2\).

Let \(\tau(K) = \sup_{1 \leq k \leq K}(\lambda_k - \lambda_{k+1})^{-1}\) and set \(K\) to be a function of \(T\), i.e., \(K_T\) such that \(K_T \to \infty\) as \(T \to \infty\). Now it follows directly from Theorem 2 and Lemma 3 that

**Theorem 4** Let Assumptions 1 and 2 hold. If \(N \geq cT^{1/r}\) for some constant \(c\), then

\[
\mathbb{E}\left(\sup_{k \geq 1} |\hat{\lambda}_k - \lambda_k|^2\right) = O(T^{-1}).
\]

And if \(\tau(K) = o(T^{1/2})\) for large \(T\), we have

\[
\mathbb{E}\left(\sup_{1 \leq k \leq K} \|\hat{v}_k - v'_k\|^2\right) = o(1).
\]
Moreover, if there exist $a > 0$ and $0 < b < 1$ such that $\gamma_k \leq ab^k$ and $\lambda_k \leq ab^k$, and if $N$ is such that $N > cT^{2/r} \log^{s-2/r} T$ for some constants $c > 0$ and $s > 0$, we have

$$\sup_{k \geq 1} |\hat{\lambda}_k - \lambda_k| = O(T^{-1/2} \log^{3/2} T) \text{ a.s.}$$

and

$$\sup_{1 \leq k \leq K} \|\hat{v}_k - v'_k\| = o(1) \text{ a.s.}$$

for large $T$ if $\tau(K) = o(T^{1/2} \log^{-3/2} T)$.

### 3.3 Regression and Error Variance Operators

Using the estimates we introduce in previous sections, we may consistently estimate the autoregressive operator $A$ by an estimate for $A_K$ given earlier in (13). To define the estimate more precisely, we first denote by $(\hat{\lambda}_k, \hat{v}_k)$ be the pairs of estimated eigenvalues and eigenvectors of $Q$ ordered as in (11) and by $\hat{P}$ the estimate of $P$ given in (16). Then we define

$$\hat{Q}_K^+ = \sum_{k=1}^K \hat{\lambda}_k^{-1} (\hat{v}_k \otimes \hat{v}_k),$$

and subsequently,

$$\hat{A}_K = \hat{P} \hat{Q}_K^+$$

(18)

analogously as in (13).

**Theorem 5** Let Assumptions 1 and 2 hold. Moreover, we assume that there exist $a > 0$ and $0 < b < 1$ such that $\gamma_k \leq ab^k$ and $\lambda_k \leq ab^k$, $N > cT^{2/r} \log^{s-2/r} T$ for some constants $c > 0$ and $s > 0$, that $A$ is Hilbert-Schmidt, and that $T\lambda^2_K/\log^3 T(\sum_1^K \tau_k)^2 \to \infty$, where $(\tau_k)$ are defined in Lemma 3. Then we have

$$\|\hat{A}_K - A\| \to a.s. 0$$

as $T \to \infty$.

Theorem 5 establishes the consistency of $\hat{A}_K$. Our result here is comparable to Theorem 8.8 in Bosq (2000).

Now we define

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T (\hat{\varepsilon}_t \otimes \hat{\varepsilon}_t),$$

(19)

where $(\hat{\varepsilon}_t)$ are the fitted residuals from FR. As is well expected from Theorem 5, we have
Corollary 6  Let the conditions in Theorem 5 hold. Then we have
\[ \hat{\Sigma} \rightarrow_{a.s.} \Sigma \]
as \( T \to \infty \).

Corollary 6 establishes the strong consistency of \( \hat{\Sigma} \).

4. Hypothesis Testing

As mentioned earlier, the functional regression of density allows intricate dependence structures of regressand density on regressor density. It is important that we be able to extract such information from the estimator for the regressive operator \( A \). One way to do this is to test hypotheses on \( A \) or its Hilbert adjoint \( A^* \). We here propose a general form of testable hypotheses and then derive weak convergence results that are necessary in testing.

We introduce Assumption 3, which is required for the asymptotic theory that we are going to develop. As before, \( Q \) denotes the variance operator of \((f_t)\), and \( \lambda \) is the function on \( \mathbb{R} \) that is given by \( \lambda_k = \lambda(k) \), where \((\lambda_k)\) are eigenvalues of \( Q \) ordered as in (11).

Assumption 3  We assume
(a) \( \lambda_k \) is convex in \( k \),
(b) \((\Delta_t)\) are iid and are independent of \( w_t \), \( \sup_{k \geq 1} \mathbb{E}(v_k, \Delta_t)^2/\lambda_k = O_p(N^{-r}) \),
(c) \( w_t \otimes \varepsilon_t \) is a martingale difference sequence.

Since our analysis is asymptotic, we only require the condition in part (a) hold for large values of \( k \). This assumption is very weak and satisfied by many sequences of \((\lambda_k)\) including \( \lambda_k = c/k^{1+a} \), \( \lambda_k = c/k^a \log^{1+b} k \) and \( \lambda_k = c \exp(-ak) \), where \( a, b \) and \( c \) are some positive constants.

We denote respectively by \( \Pi_K \) and \( \hat{\Pi}_K \) the projections on the subspaces \( V_K \) and \( \hat{V}_K \) of \( H_2 \) generated by the \( K \)-principal eigenvectors \( v_1, \ldots, v_K \) of \( Q \) and \( \hat{v}_1, \ldots, \hat{v}_K \) of \( \hat{Q} \).

We consider the null hypothesis given by
\[ H_0 : A^*v \in U, \]  (20)
where \( U \) is a subspace of \( H_2 \). Let \( U \) be \( L \)-dimensional and spanned by \( u_1, \ldots, u_L \), so that we may write
\[ A^*v = \sum_{k=1}^{L} c_k u_k \]
with some coefficients \((c_k)\). Then we have
\[ \langle v, g_t \rangle = \sum_{k=1}^{L} c_k \langle u_k, f_t \rangle + \eta_t, \]  (21)
which follows readily from (2). If in particular \( u_k = p_k \), where \( p_k(x) = x^k \), i.e., the \( k \)-th polynomial, then the null hypothesis implies that the \( v \)-moment of \( (g_t) \) only depends on the first \( L \) moments of \( (f_t) \).

Let \( \hat{A}_K^* \) be the adjoint of \( \hat{A}_K \) defined in (18), and let

\[ \hat{u}_{L+1}, \ldots, \hat{u}_K \]

be the set of orthonormal vectors which are orthogonal to \( \hat{Q}_K^{1/2}U \) on the subspace \( \hat{V}_K \) of \( H \) spanned by the \( K \)-principal eigenvectors of \( \hat{Q}_K \). Note that \( \hat{Q}_K^{1/2}U \) is an \( L \)-dimensional subset of \( \hat{V}_K \). Moreover, we let \( \hat{\Sigma} \) be defined as in (19). Now we define our test statistic \( Z \) by

\[
Z = \frac{1}{\sqrt{2(K-L)}} \sum_{k=L+1}^{K} \left( \frac{T(\hat{u}_k, \hat{Q}_K^{1/2} \hat{A}_K^* v)^2}{\langle v, \hat{\Sigma} v \rangle} - 1 \right).
\]

Then we have

**Theorem 7** Let Assumptions 1, 2 and 3 hold, and assume that (a) \( N \geq cT^{1/r} \) for some constant \( c > 0 \), (b) \( K^3 \log^2 K/T \to 0 \), and (c) \( K^2/(\delta_K T) \to 0 \) as \( T \to \infty \). Under the null hypothesis (20), we have

\[ Z \to_d N(0, 1) \]

as \( T \to \infty \).

If the null hypothesis (20) does not hold on \( \hat{V}_K \), the projection of \( \hat{Q}_K^{1/2}A^* v \) on the subspace spanned by \( \hat{u}_{L+1}, \ldots, \hat{u}_K \) is non-vanishing and we have

\[
\sum_{k=L+1}^{K} (\hat{u}_k, \hat{Q}_K^{1/2} A^* v)^2 \neq 0
\]

for any \( v \in H \). In this case, the statistic \( Z \) therefore diverges at the rate of \( T/\sqrt{K} \). The test based on \( Z \) is therefore consistent as long as \( K = o(T^2) \).

The asymptotic normal distribution in Theorem 7 may be used for obtaining critical values for our tests. In practice, however, the asymptotic distribution may be a rather crude approximation. Here the major issue is not that \( T \) or \( N \) is not big enough, but that \( K \) increases rather slowly as \( T \) increases. To see this, examine the following equation from the proof of Theorem 7,

\[
Z = \frac{1}{\sqrt{2(K-L)}} \sum_{k=L+1}^{K} \left[ N_k(0,1)^2 - 1 \right] + o_p(1),
\]

where \( =d \) denotes equality in distribution and \( N_k(0,1) \) are independent standard normal random variables indexed by \( k \). The asymptotics is obtained by averaging \( K - L \) terms, which in practice is usually chosen to be a small integer. Hence the asymptotic normal distribution may be a poor approximation. We may use bootstrap or subsampling methods.
to obtain more accurate critical values for our test. But this would involve intense if not prohibitive computations. Instead we may pretend that the integer of $K - L$ is a parameter (degree of freedom) of the distribution of $Z$ and obtain approximate quantiles of $Z$ using the $\chi^2$ quantile table.

5. Simulations

In this section we investigate the finite sample behavior of the test statistic $Z$ defined in (22). We focus on a specific testing problem: given a panel or pseudo-panel data set, test whether the mean of the cross-section distribution is systematically dependent on some structure of past moments. We work on following four hypotheses:

$$H_0^L : A^*v \in U^L, \quad L = 0, \cdots, 3$$

where $v$ is the first polynomial and $U^L$ is the space spanned by the first $L$ polynomials.

For the convenience of exposition, we replicate the equation (1) here, $g_t = c + Af_t + \epsilon_t$.

Under $H_0^0$, the mean of $g$ does not depend on any moments of $f$. Under $H_0^1$, the mean of $g$ does not depend on the moments of $f$ that are higher than the first. Under $H_0^2$, the mean of $g$ does not depend on the moments of $f$ that are higher than the second. If both $H_0^0$ and $H_0^1$ are rejected, and $H_0^2$ is maintained, then we accept the statement that the mean depend on past mean and variance only. Under $H_0^3$, the mean does not depend on the moments of $f$ that are higher than three.

We use Gaussian density to generate density functions $g_t$ and $f_t$. Let the first two moments of $g$ be $\mu_{1,t}$ and $\mu_{2,t}$, respectively, and the first two moments of $f$ be $\nu_{1,t}$ and $\nu_{2,t}$, respectively. And we set

$$\begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \Phi \begin{bmatrix} \nu_{1,t} \\ \nu_{2,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}. \tag{24}$$

$\epsilon_{i,t}$ are iid across $i$ and $t$, generated from a uniform distribution, Uniform($-0.2, 0.2$). We let $\nu_{1,t}$ be drawn from Uniform($-1, 1$) and $\nu_{2,t}$ from Uniform($1, 2$). And we let $c_1 = 0.75$ and $c_2 = 0.5$. The matrix $\Phi$ gives the regression structure,

$$\Phi = \begin{bmatrix} 0.5 & -0.5 \\ 0 & 0.5 \end{bmatrix}.$$ 

Under this specification, the hypotheses $H_0^0$ and $H_0^1$ are false, while $H_0^2$ and $H_0^3$ are true. The percentages of rejections under $H_0^0$ and $H_0^1$ then give empirical power, and those under $H_0^2$ and $H_0^3$ give empirical size.

A few practical issues are worth noting. First, the choice of $K$ plays an analogous role with that of the smoothing parameter in generic nonparametric tests. There are two possible approaches to control $K$. One is to keep $K$ constant, say, $K = 10$. As $L$ varies from 0 to 3 for hypotheses from $H_0^0$ to $H_0^3$, $L - K$, which is the number of independent $\chi^2_1$ variables in the formation of $Z$-statistic in (22), varies from 10 to 7. The other is to keep $K_0 = K - L$ constant, say, $K_0 = 7$. Then $K$ for the hypotheses from $H_0^0$ to $H_0^3$ varies from 7
to 10. We adopt the latter approach. Generally, as $T$ increases, $K_0$ should slowly increase. A rough guide for this choice is the integer part of $K_0 = -11.96 + 1.82\sqrt{T}$.

Second, the support of density functions is another choice a practitioner should make. However, simulation results indicate that this is not a crucial decision as long as the support is wide enough. In the following simulations we choose the centered support that preserves 99.99995% of the total probability mass of the average of $f_t$ and $g_t$, that is, $[-6.1560, 6.1560]$ for $g$ and $[-5.6196, 5.6196]$ for $f$.

Third, to obtain the set of orthonormal vectors that are orthogonal to $\hat{Q}_{K/2}^1 U$ on the subspace $\hat{V}_K$, we first estimate $\hat{Q}_K$ and obtain $\hat{u}_k = \hat{Q}_{K/2}^1 p_k$, where $p_k(x) = x^k, k = 1, \cdots, K$. Then we use a functional modified Gram-Schmidt procedure to orthonormalize the set of functions $(\hat{u}_1, \cdots, \hat{u}_K)$. The modified Gram-Schmidt procedure is theoretically identical to the classic Gram-Schmidt but numerically more stable. By the definition of Gram-Schmidt procedure, the last $K - L$ vectors of the orthonormalized set are orthogonal to $(\hat{u}_1, \cdots, \hat{u}_L)$ and thus $\hat{Q}_{K/2}^1 U$.

The results are summarized in Table 1. For each $T$, there are four rows of rejection percentages corresponding to null hypotheses from $H_0^0$ to $H_3^0$. According to our design, the first two rows give empirical powers, while the following two give empirical sizes. We can see that all empirical sizes in the table are close to the nominal values. This remains true when we run simulations with different values for $\Phi$, which controls “signal-to-noise ratio”.

We now turn to power of the tests. In the table, “$N = \infty$” corresponds to the case in which densities are directly observable or, more realistically, the case when $N$ is a very large number. In this case, the tests enjoys reasonable power even when $T$ is only 50. We note that the combination of a short time horizon and a very long cross-section is typical for panel or pseudo-panel data in applied microeconomic studies. When $N$ is finite, the tests also perform well, especially when $T$ is relatively long. For example, the combination of $N = 75$ and $T = 75$ already achieves good power for our tests. The power of the tests increases as either $T$ or $N$ increases.

6. An Empirical Application

This section offers an illustration of our methodology. We investigate how high-frequency intraday variation in the stock index of South Korea depends on that of the United States.

We study the main tracking index in South Korea, which is the Korean Composite Stock Price Index, or KOSPI for short. The KOSPI Index is calculated from 200 of the largest and most liquid stocks with weights of market capitalization. For the US data we choose Dow Jones Industrial Average (DJIA).

We calculate 5-minute log returns (5-min return hereafter) on each index on each business day. In average, there are 77 5-min returns each business day in the US market and 69 5-min returns each day in the South Korean market. Table 2 gives summary statistics of the calculated returns. All figures are in percentage units.

We estimate the density of 5-min returns on each day using kernel smoothing. The implicit assumption here is that on each day, the 5-min returns on an index is strictly
<table>
<thead>
<tr>
<th>( N = \infty )</th>
<th>( N = 75 )</th>
<th>( N = 150 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha = 1 )</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>5</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>10</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

\( T = 50, \ K_0 = 2 \)

\( H_0^0 \): 100.0 100.0 100.0

\( H_0^1 \): 87.3 91.2 92.8

\( H_0^2 \): 1.1 4.7 9.6

\( H_0^3 \): 0.9 5.3 10.1

\( T = 75, \ K_0 = 3 \)

\( H_0^0 \): 100.0 100.0 100.0

\( H_0^1 \): 100.0 100.0 100.0

\( H_0^2 \): 1.2 5.4 9.7

\( H_0^3 \): 1.3 5.3 11.0

\( T = 100, \ K_0 = 5 \)

\( H_0^0 \): 100.0 100.0 100.0

\( H_0^1 \): 100.0 100.0 100.0

\( H_0^2 \): 1.4 5.3 10.2

\( H_0^3 \): 0.9 5.1 10.1

\( T = 125, \ K_0 = 6 \)

\( H_0^0 \): 99.9 99.9 99.9

\( H_0^1 \): 99.9 99.9 99.9

\( H_0^2 \): 1.2 5.0 9.4

\( H_0^3 \): 0.9 4.7 9.6

\( T = 150, \ K_0 = 8 \)

\( H_0^0 \): 99.8 99.8 99.8

\( H_0^1 \): 99.8 99.8 99.8

\( H_0^2 \): 1.1 4.9 9.6

\( H_0^3 \): 1.1 4.7 9.5

---

**Table 2: Summary Statistics**

<table>
<thead>
<tr>
<th>year</th>
<th>n</th>
<th>Mean</th>
<th>STD</th>
<th>Skew</th>
<th>Kurt</th>
<th>Min</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOSPI</td>
<td>2003</td>
<td>17513</td>
<td>-0.000494</td>
<td>0.147</td>
<td>-0.492</td>
<td>10.5</td>
<td>-1.79</td>
<td>-0.503</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>17655</td>
<td>-0.000837</td>
<td>0.141</td>
<td>-0.0508</td>
<td>11.9</td>
<td>-1.52</td>
<td>-0.508</td>
</tr>
<tr>
<td>DJIA</td>
<td>2003</td>
<td>19513</td>
<td>-0.000638</td>
<td>0.0973</td>
<td>-0.0245</td>
<td>7.27</td>
<td>-0.784</td>
<td>-0.322</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>19620</td>
<td>-0.000223</td>
<td>0.0672</td>
<td>-0.000707</td>
<td>6.20</td>
<td>-0.613</td>
<td>-0.214</td>
</tr>
</tbody>
</table>
stationary. Figure 1 shows density functions for the intraday 5-min returns on KOSPI (South Korea) and DJIA (US) in 2003 and 2004.

Recall that we denote \( k \)-th polynomial as \( p_k \). We conduct the following tests. Only the null hypotheses are listed. The alternatives are the opposites of the nulls.

(T.1) \( H_0 : A^* p_1 \in \{0\} \)
(T.2) \( H_0 : A^* p_1 \in \text{span}(p_1) \)
(T.3) \( H_0 : A^* p_1 \in \text{span}(p_1, p_2) \)
(T.4) \( H_0 : A^* p_1 \in \text{span}(p_1, p_2, p_3) \)
(T.5) \( H_0 : A^* p_2 \in \{0\} \)
(T.6) \( H_0 : A^* p_2 \in \text{span}(p_2) \)
(T.7) \( H_0 : A^* p_2 \in \text{span}(p_2, p_1) \)
(T.8) \( H_0 : A^* p_2 \in \text{span}(p_2, p_1, p_3) \)
Table 3: Tests on Moment Dependence of 5-min Returns of KOSPI on that of DJIA.

<table>
<thead>
<tr>
<th></th>
<th>2003</th>
<th>2004</th>
<th>2003-2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_0$</td>
<td>15</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>T.1</td>
<td>0.1852</td>
<td>0.0991</td>
<td>4.4548***</td>
</tr>
<tr>
<td>T.2</td>
<td>-0.6449</td>
<td>-0.3808</td>
<td>3.8596***</td>
</tr>
<tr>
<td>T.3</td>
<td>-0.2107</td>
<td>-0.3093</td>
<td>1.2163</td>
</tr>
<tr>
<td>T.4</td>
<td>-0.1385</td>
<td>-0.3099</td>
<td>1.1963</td>
</tr>
<tr>
<td>T.6</td>
<td>0.3509</td>
<td>0.7537</td>
<td>1.8201</td>
</tr>
<tr>
<td>T.7</td>
<td>0.5759</td>
<td>0.5121</td>
<td>2.1436*</td>
</tr>
<tr>
<td>T.8</td>
<td>0.2165</td>
<td>0.5666</td>
<td>3.3305***</td>
</tr>
</tbody>
</table>

Note: * denotes rejection at 10% significance level, ** denotes rejection at 5% significance level, and *** denotes rejection at 1% significance level.

The first four tests concern how the mean of the 5-min return of KOSPI (South Korea) depends on the distribution of the 5-min returns on DJIA (US). The last four tests concern how the second moment of the 5-min return on KOSPI depends on the corresponding distribution of DJIA. The results are summarized in Table 3.

We first discuss results on the tests from T.1 to T.4. In the year of 2003, we may conclude that the mean of the intraday 5-min returns on KOSPI (South Korea) does NOT depend on any moments of the corresponding returns on DJIA (US). On a particular trading day, the mean of the intraday 5-min returns corresponds to the color and the length of the body on the popular Candlestick chart of the underlying stock price or index. In the language of technical analysis, a hollow body (close higher) generally indicates buying pressure, while a filled body (close lower) indicates selling pressure. A long body (large absolute mean) indicates more intense buying or selling pressure, while a short body (small absolute mean) means little upward or downward movement and generally indicates “consolidation”. The results on the tests from T.1 to T.4 for the 2003 data may be translated into saying that the body on the 1-day chart of KOSPI does not depend on any moments of the intraday 5-min return on DJIA, which include the first moment that corresponds to the body on the DJIA chart. This is a somewhat strong conclusion. It is well known that the US market “leads” other international markets. Our results may imply that the leadership of the US stock market over the South Korean counterpart is adequately reflected in the opening price of KOSPI.

In the year of 2004, however, we reject the hypotheses T.1 and T.2 and accept T.3 and T.4, implying that the mean of the intraday 5-min return on KOSPI, to some degree, depends on the first two moments of the 5-min return on DJIA. We also run the same regression on the joint data of 2003 and 2004. The results are similar with the case of 2003. Viewing together, we may safely say that the dependent relationship of intraday 5-min return on KOSPI with that on DJIA may change over time and that the mean of the former at most weakly depend on the first two moments of the latter.
We next discuss results on the tests from T.5 to T.8. All these tests deal with the second moment of the 5-min return on KOSPI, which captures at least part of the intraday risk. The 2003 data reveal that the second moment of 5-min return on KOSPI depends ONLY on the second moment of DJIA 5-min return. The 2004 data tells a less strong story, in which the second moment of the 5-min return on KOSPI may depend on higher moments of the 5-min return on DJIA. The joint data of 2003 and 2004 tell a similar story.

7. Conclusions

In this paper we have proposed a model of functional regression of continuous-state distributions. We explore the properties of the model and outline the estimation procedure. We show that the proposed estimator, which uses estimated densities, is consistent under mild conditions. And we develop a hypothesis testing procedure and derive the asymptotic distribution for our test statistic. Simulation results show that our test performs well in terms of size and power in finite samples. Finally, we investigate how the distribution of intraday 5-min return on the South Korean stock index KOSPI depends on its counterpart on the Dow Jones Industrial Average using our methodology.

References


Appendix: Mathematical Proofs

Throughout the proof, $M$ denotes a generic constant.

**Proof for Lemma 1** We prove the statements concerning $\bar{f}$. The same holds for $\bar{g}$. First we have

$$\bar{f} - \mathbb{E}f = \bar{w} + R,$$

where

$$\bar{w} = \frac{1}{T} \sum_{t=1}^{T} w_t \quad \text{and} \quad R = \frac{1}{T} \sum_{t=1}^{T} \Delta_{f,t}.$$ 

Hence

$$\mathbb{E}\|\bar{f} - \mathbb{E}f\|^2 \leq 2\mathbb{E}\|\bar{w}\|^2 + 2\mathbb{E}\|R\|^2. \quad (25)$$

It can be easily checked that that

$$\mathbb{E}\|R\|^2 = O(N^{-r}).$$

Hence under the condition on $N$, we have

$$\mathbb{E}\|R\|^2 = O(T^{-1}).$$

For the first term,

$$\|\bar{w}\|^2 = \sum_{k=1}^{\infty} \langle \bar{w}, v_k \rangle^2 = \sum_{k=1}^{\infty} \frac{1}{T^2} \left( \sum_{t=1}^{T} \langle w_t, v_k \rangle \right)^2.$$

Assumption 2 (b) implies $\xi_k < M$ a.s., where $\xi_k$ is defined in (14). Hence

$$\langle w_t, v_k \rangle \leq \frac{1}{2} M \lambda_k \quad \text{a.s.}$$

Then by moment inequality of strong mixing process,

$$|\mathbb{E}\langle w_1, v_k \rangle \langle w_{1+j}, v_k \rangle| \leq 4M^2 \lambda_k \alpha(j),$$
where $\alpha(j)$ are the strong mixing coefficients. Then,

$$
\mathbb{E} \frac{1}{T^2} \left( \sum_{t=1}^{T} (w_t, v_k) \right)^2 \leq \lambda_k/T + \frac{2}{T^2} \sum_{j=1}^{T-1} (T-j)\left| \mathbb{E}(w_1, v_k)(w_{1+j}, v_k) \right|
$$

$$
\leq \lambda_k/T + \frac{8M^2}{T} \sum_{j=1}^{T-1} (1 - j/T) \alpha(j) \lambda_k
$$

$$
= M\lambda_k/T.
$$

Hence

$$
\mathbb{E} \| \tilde{w} \|^2 \leq \frac{M}{T} \sum_{k=1}^{\infty} \lambda_k = O\left( \frac{1}{T} \right),
$$

(26)
since $Q$ is nuclear. The first part of Lemma 1 is now proved. For the a.s. convergence in the second part, we note that

$$
\bar{f} - Ef \leq \| \tilde{w} \| + \| R \|.
$$

It is established in Corollary 2.4 in Bosq (2000) that

$$
\| \tilde{w} \| = o\left( \frac{T^{-1/2} \log^{3/2} T}{T} \right).
$$

To see $\| R \| = o\left( T^{-1/2} \log^{3/2} T \right)$ a.s., we look at

$$
\mathbb{P} \left( \frac{T^{1/2}}{\log^{3/2} T} \| R \| \geq \epsilon \right) \leq \frac{1}{\epsilon^2 \log^3 T} \mathbb{E} \| R \|^2
$$

$$
= \frac{T}{\log^3 T} O(N^{-r})
$$

$$
= O\left( \frac{1}{T \log^{1+s} T} \right)
$$

By Borel-Cantelli Lemma we have

$$
\mathbb{P} \left( \lim_{T \to \infty} \frac{T^{1/2}}{\log^{3/2} T} \| R \| \geq \epsilon \right) = 0.
$$

This is,

$$
\| R \| = o\left( T^{-1/2} \log^{3/2} T \right) \text{ a.s.}
$$

The proof is now complete. □

**Proof for Theorem 2**

Let

$$
\tilde{Q} = \frac{1}{T} \sum_{t=1}^{T} w_t \otimes w_t,
$$

We have

$$
\| \tilde{Q} - Q \|^2 = \| \tilde{Q} - \tilde{Q} + \tilde{Q} - Q \|^2 \leq 2( \| \tilde{Q} - \tilde{Q} \|^2 + \| \tilde{Q} - Q \|^2).
$$

(27)
By Lemma A.1 in Park and Qian (2007),

\[ \mathbb{E}\|\hat{Q} - \tilde{Q}\|^2 = O(N^{-r}). \]

So under the condition on \( N \) we have

\[ \mathbb{E}\|\hat{Q} - \tilde{Q}\|^2 = O(T^{-1}). \]

We now prove \( \mathbb{E}\|\hat{Q} - Q\|^2 = O(T^{-1}) \). First we prove a crucial result,

\[
\sup_{m,p} T \mathbb{E} \frac{\langle (\hat{Q} - Q)(v_p), v_m \rangle}{\lambda_p \lambda_m} \leq M \tag{28}
\]

\[
\sup_{m,p} T \mathbb{E} \frac{\langle (\hat{P} - P)(v_p), u_m \rangle}{\lambda_p \gamma_m} \leq M \tag{29}
\]

where \( M \) is some constant. The result in (28) is proved in Mas (2006) for the case when \((w_t)\) follows an FAR(1) process. Here we show that it holds when \((w_t)\) is GSM. First note that,

\[
\langle (\hat{Q} - Q)(v_p), v_m \rangle \leq \langle \hat{Q}(v_p), v_m \rangle \text{ a.s.}
\]

Then by stationarity of \((w_t)\) we have

\[
\mathbb{E}\langle (\hat{Q} - Q)(v_p), v_m \rangle^2 \leq \frac{1}{T} \mathbb{E}\langle w_t, v_p \rangle^2 \langle w_t, v_m \rangle^2 + \frac{2}{T^2} \sum_{1 \leq s < t \leq T} \mathbb{E}\langle w_s, v_p \rangle \langle w_s, v_m \rangle \langle w_t, v_p \rangle \langle w_t, v_m \rangle
\]

It is obvious from KL decomposition in (14) and Assumption 2(d) that

\[
\frac{1}{T} \mathbb{E}\langle w_t, v_p \rangle^2 \langle w_t, v_m \rangle^2 = \frac{1}{T} \lambda_p \lambda_m \mathbb{E}(\xi_p^2 \xi_m^2) \leq M \lambda_p \lambda_m T.
\]

For the second term, when \( m \neq p \),

\[
C := \frac{2}{T^2} \sum_{1 \leq s < t \leq T} \mathbb{E}\langle w_s, v_p \rangle \langle w_s, v_m \rangle \langle w_1, v_p \rangle \langle w_1, v_m \rangle
\]

\[
= \frac{2}{T} \sum_{j=1}^{T-1} (1 - j/T) \mathbb{E}\langle w_1, v_p \rangle \langle w_1, v_m \rangle \langle w_{1+j}, v_p \rangle \langle w_{1+j}, v_m \rangle
\]

\[
= \frac{2}{T} \sum_{j=1}^{T-1} (1 - j/T) \text{cov}\left( \langle w_1, v_p \rangle \langle w_1, v_m \rangle, \langle w_{1+j}, v_p \rangle \langle w_{1+j}, v_m \rangle \right)
\]

\[
\leq \lambda_m \lambda_p \frac{8M^4}{T} \sum_{j=1}^{T-1} (1 - j/T) \alpha(j), \tag{30}
\]

where \( \alpha(j) \)'s are the strong mixing coefficients. Note that \( \langle w_1, v_p \rangle \langle w_1, v_m \rangle \leq M^2 \sqrt{\lambda_m \lambda_p} \) a.s. and we applied the moment inequality of mixing process.
When \( m = p \), we have

\[
\mathbb{E} \langle (\hat{Q} - Q)(v_p), v_p \rangle^2 = \mathbb{E} \langle \hat{Q}(v_p), v_p \rangle^2 - \lambda_p^2,
\]

and

\[
\mathbb{E} \langle \hat{Q}(v_p), v_p \rangle^2 = \frac{1}{T} \mathbb{E} \langle w_t, v_p \rangle^4 + \frac{2}{T^2} \sum_{1 \leq s < t \leq T} \mathbb{E} \langle w_s, v_p \rangle^2 \langle w_t, v_p \rangle^2.
\]

Again we have \( \frac{1}{T} \mathbb{E} \langle w_t, v_p \rangle^4 \leq \frac{M}{T} \lambda_p^2 \). Now let

\[
C := \frac{2}{T^2} \sum_{1 \leq s < t \leq T} \mathbb{E} \langle w_s, v_p \rangle^2 \langle w_t, v_p \rangle^2 - \lambda_p^2
\]

\[
= \frac{2}{T^2} \sum_{1 \leq s < t \leq T} \text{cov}(\langle w_s, v_p \rangle^2, \langle w_t, v_p \rangle^2)
\]

\[
= \frac{2}{T} \sum_{j=1}^{T-1} (1 - j/T) \text{cov}(\langle w_1, v_p \rangle^2, \langle w_{1+j}, v_p \rangle^2)
\]

\[
\leq \lambda_p^2 8M^4 \frac{T-1}{T} \sum_{j=1}^{T-1} (1 - j/T) \alpha(j)
\]

(31)

In both (30) and (31), \( \sum_{j=1}^{T-1} (1 - j/T) \alpha(j) \) is bounded as \( (w_t) \) is GSM (In fact, GSM is more than necessary here). Hence

\[
\frac{\mathbb{E} \langle (\hat{Q} - Q)(v_p), v_m \rangle^2}{\lambda_p \lambda_m} \leq \frac{M}{T}.
\]

Hence (28) is established. (29) can be similarly established. Then we have

\[
\mathbb{E} \| \hat{Q} - Q \|^2 = \mathbb{E} \sum_{m,p} \langle (\hat{Q} - Q)(v_p), v_m \rangle^2
\]

\[
\leq \frac{1}{T} \left( \sum_{m,p} \lambda_p \lambda_m \right) \sup_{m,p} T \mathbb{E} \langle (\hat{Q} - Q)(v_p), v_m \rangle^2 \lambda_p \lambda_m
\]

\[
\leq \frac{M}{T} \sum_{m,p} \lambda_p \lambda_m = \frac{M}{T} \left( \sum_m \lambda_m \right)^2
\]

\[
= O(T^{-1}),
\]

using the fact that \( Q \) is nuclear. So \( \mathbb{E} \| \hat{Q} - Q \|^2 = O(T^{-1}) \). Basically the same argument obtains \( \mathbb{E} \| \hat{P} - P \|^2 = O(T^{-1}) \).

For the second part, let

\[
Z_t = w_t \otimes w_t - Q.
\]
It is clear that \((Z_t)\) are zero-mean operator-valued random variables and geometrically strongly mixing. Furthermore \(\|Z_t\| < M\) a.s. Hence Corollary 2.4 in Bosq (2000) is applicable to \((Z_t)\) and we have
\[
\|\hat{Q} - Q\| = \left\| \frac{1}{T} \sum_{t=1}^{T} Z_t \right\| = o(T^{-1/2} \log^{3/2} T).
\]
And it follows an application of Chebyshev’s Inequality and Borel-Cantelli Lemma that
\[
\|\hat{Q} - \tilde{Q}\| = o(T^{-1/2} \log^{3/2} T)
\]
as \(N > cT^{2/r} \log^{s-2/r} T\) for some constants \(c > 0\) and \(s > 0\). So \(\|\hat{Q} - Q\| = o(T^{-1/2} \log^{3/2} T)\). The statement on \(\hat{P}\) is similarly established.

Proof for Lemma 3  
See Lemma 4.2 and Lemma 4.3 in Bosq (2000). These two lemmas are valid for all compact linear operators.

Proof for Theorem 4  
It follows immediately from Theorem 2 and Lemma 3.

Proof for Theorem 5  
We first prove:
\[
\|\hat{P}(\hat{v}_k)\| \leq \hat{\lambda}_k^{1/2} \left( \frac{1}{T} \sum_{t=1}^{T} \|\hat{m}_t\|^2 \right)^{1/2}.
\]  (32)
Note that \(\hat{P}\) is a mapping from \(H_2\) to \(H_1\). Let \((\hat{\gamma}_k, \hat{u}_k)\) be the estimated eigen-pairs for \(W\). For each \(j\),
\[
|\langle \hat{P}(\hat{v}_k), \hat{u}_j \rangle| = \frac{1}{T} \left| \sum_{t=1}^{T} \langle \hat{w}_t, \hat{v}_k \rangle \langle \hat{m}_t, \hat{u}_j \rangle \right| \\
\leq \left( \frac{1}{T} \sum_{t=1}^{T} \|\hat{w}_t\|^2 \right)^{1/2} \left( \frac{1}{T} \sum_{t=1}^{T} \|\hat{m}_t\|^2 \right)^{1/2} \\
= \langle \hat{Q}(\hat{v}_k), \hat{v}_k \rangle^{1/2} \langle \hat{W}(\hat{u}_j), \hat{u}_j \rangle^{1/2} \\
= \hat{\lambda}_k^{1/2} \hat{\gamma}_j^{1/2}.
\]
Then
\[
\|\hat{P}(\hat{v}_k)\|^2 = \sum_{j=1}^{\infty} |\langle \hat{P}(\hat{v}_k), \hat{u}_j \rangle|^2 \\
\leq \hat{\lambda}_k \sum_{j=1}^{\infty} \hat{\gamma}_j = \hat{\lambda}_k \|\hat{W}\|_N = \hat{\lambda}_k \frac{1}{T} \sum_{t=1}^{T} \|\hat{m}_t\|^2,
\]
where $\| \cdot \|_N$ denotes nuclear norm. Hence (32) holds. Now we write
\[
(\hat{A}_K - A)(x) = a(x) + b(x) + c(x),
\]
where
\[
a(x) = \hat{P}\hat{Q}_K^+(x) - A\Pi_K(x), \quad b(x) = A\Pi_K(x) - A\hat{\Pi}_K(x), \quad c(x) = A\hat{\Pi}_K(x) - A(x).
\]
We first examine $a(x)$.

\[
a(x) = \hat{P}\sum_{k=1}^K \hat{\lambda}_k^{-1}\langle \hat{v}_k, x \rangle \hat{v}_k - P\sum_{k=1}^K \lambda_k^{-1}\langle v_k, x \rangle v_k = a_1(x) + a_2(x) + a_3(x) + a_4(x),
\]
where
\[
a_1(x) = \hat{P}\sum_{k=1}^K (\hat{\lambda}_k^{-1} - \lambda_k^{-1})\langle \hat{v}_k, x \rangle \hat{v}_k,
\]
\[
a_2(x) = \hat{P}\sum_{k=1}^K \lambda_k^{-1}\langle \hat{v}_k - \hat{v}_k', x \rangle \hat{v}_k,
\]
\[
a_3(x) = \hat{P}\sum_{k=1}^K \lambda_k^{-1}\langle \hat{v}_k, x \rangle (\hat{v}_k - \hat{v}_k'),
\]
\[
a_4(x) = (\hat{P} - P)\sum_{k=1}^K \lambda_k^{-1}\langle v_k, x \rangle v_k.
\]

\[
\|a_1(x)\| \leq \sum_{k=1}^K \frac{|\hat{\lambda}_k - \lambda_k|}{\hat{\lambda}_k \lambda_k} \|\langle \hat{v}_k, x \rangle\| \|\hat{P}(\hat{v}_k)\|
\]
\[
\leq \sup_k |\hat{\lambda}_k - \lambda_k| \hat{\lambda}_K^{-1/2} \lambda_K^{-1} \left( \sum_{k=1}^K |\langle \hat{v}_k, x \rangle| \right) \left( \frac{1}{T} \sum_{t=1}^T \|\hat{m}_t\|^2 \right)^{1/2}
\]
\[
\leq \|\hat{Q} - Q\| \hat{\lambda}_K^{-1/2} \lambda_K^{-1} \|x\| \left( \frac{1}{T} \sum_{t=1}^T \|\hat{m}_t\|^2 \right)^{1/2},
\]

We have
\[
\lambda_K \left( \sum_{k=1}^K \tau_k \right)^{-1} = O(T^{-1/2} \log^{3/2} T).
\]

Then from Theorem 2 we may deduce that $\|\hat{Q} - Q\| \leq \lambda_K / 2$ for large enough $T$ a.s.. This implies
\[
\hat{\lambda}_K \geq \lambda_K - \|\hat{Q} - Q\| \geq \lambda_K / 2.
\]
Hence
\[ \|a_1\| \leq \sqrt{2} \|Q - Q\| \lambda_K^{-3/2} K^{1/2} \left( \frac{1}{T} \sum_{t=1}^{T} \|\hat{m}_t\|^2 \right)^{1/2} \]

Since \( \|\hat{m}_t\| < M \) a.s. and \( \lambda_K^{-3/2} K^{1/2} \|Q - Q\| \to 0 \) using the a.s. convergence result in Theorem 2, we conclude \( \|a_1\| \to 0 \) a.s.. Similar calculations yields
\[ \|a_2\| \leq \sqrt{3/2} \|Q - Q\| \lambda_1^{1/2} \lambda_K^{-1} \left( \sum_{k=1}^{K} \tau_k \right) \left( \frac{1}{T} \sum_{t=1}^{T} \|\hat{m}_t\|^2 \right)^{1/2} \to 0 \text{ a.s.} \]
as \( T \) goes to infinity. And
\[ \|a_3\| \leq \|\hat{P}\| \left( \sum_{k=2}^{K} \tau_k \right) K^{1/2} \lambda_K^{-1} \|Q - Q\| \to 0 \text{ a.s.} \]

And finally,
\[ \|a_4\| \leq \|P - P\| \lambda_K^{-1} \to 0 \text{ a.s.} \]

So we have \( \|a\| \to 0 \) a.s.

Next we deal with \( b(x) \) and \( c(x) \) in (33). First observe that
\[ A = \sum_{k=1}^{\infty} \langle v_k, \cdot \rangle A(v_k) = \sum_{k=1}^{\infty} \langle \hat{v}_k, \cdot \rangle A(\hat{v}_k). \]

Hence
\[ b(x) = \sum_{k=1}^{K} \langle v_k, x \rangle A(v_k) - \sum_{k=1}^{K} \langle \hat{v}_k, x \rangle A(\hat{v}_k) = \sum_{k>K} \langle \hat{v}_k, x \rangle A(\hat{v}_k) - \sum_{k>K} \langle v_k, x \rangle A(v_k). \]

So
\[
\|b(x)\| \leq \sum_{k>K} \|\langle \hat{v}_k, x \rangle\| \|A(\hat{v}_k)\| + \sum_{k>K} \|\langle v_k, x \rangle\| \|A(v_k)\|
\leq \left( \sum_{k>K} \|\hat{v}_k, x\|^2 \right)^{1/2} \left( \sum_{k>K} \|A(\hat{v}_k)\|^2 \right)^{1/2} + \left( \sum_{k>K} \|v_k, x\|^2 \right)^{1/2} \left( \sum_{k>K} \|A(v_k)\|^2 \right)^{1/2}
\leq \|x\| \left( \sum_{k>K} \|A(\hat{v}_k)\|^2 \right)^{1/2} + \|x\| \left( \sum_{k>K} \|A(v_k)\|^2 \right)^{1/2}.
\]

Hence
\[ \|b\|^2 \leq 2 \left( \sum_{k>K} \|A(\hat{v}_k)\|^2 + \sum_{k>K} \|A(v_k)\|^2 \right). \]
Since $A$ is Hilbert-Schmidt, $\sum_{k>K} \|A(v_k)\|^2 \to 0$. And it can be shown that $\sum_{k>K} \|A(\hat{v}_k)\|^2 \to 0$ a.s. (See the proof of Lemma 8.2 in Bosq (2000)). Hence

$$\|b\| \to 0 \text{ a.s.}$$

For $c(x)$, we have

$$c(x) = -\sum_{k>K} (\hat{v}_k, x) A(\hat{v}_k).$$

So

$$\|c(x)\| \leq \left( \sum_{k>K} (\hat{v}_k, x)^2 \right)^{1/2} \left( \sum_{k>K} \|A(\hat{v}_k)\|^2 \right)^{1/2} \leq \|x\| \left( \sum_{k>K} \|A(\hat{v}_k)\|^2 \right)^{1/2}.$$ 

Hence

$$\|c\| \leq \left( \sum_{k>K} \|A(\hat{v}_k)\|^2 \right)^{1/2} \to 0 \text{ a.s.}$$

The proof of 5 is now complete.

**Proof for Corollary 6** First we note that

$$\hat{w}_t = (\hat{f}_t - f_t) + (f_t - \mathbb{E}f) + (\mathbb{E}f - \bar{f}) = \Delta_t + w_t + (\mathbb{E}f - \bar{f}).$$

Under the condition on $N$, by Lemma 1,

$$\|\bar{f} - \mathbb{E}f\| = o(T^{-1/2} \log^{3/2} T) = o(1) \text{ a.s.}$$

And

$$P\left\{ T^{1/2} \log^{-3/2} T \|\Delta_t\| \geq \epsilon \right\} \leq \frac{1}{c^2} T \log^{-3} T \mathbb{E}\|\Delta_t\|^2 \leq \frac{1}{c^2} T \log^{-3} T \sup_t \mathbb{E}\|\Delta_t\|^2 = T \log^{-3} T O(N^{-\delta}) = O \left( \frac{1}{T \log^{\delta} T} \right),$$

where $\delta = 1 + sr > 1$. Using Borel-Cantelli Lemma we have

$$\|\Delta_t\| = o(T^{-1/2} \log^{3/2} T) = o(1), \text{ a.s.}.$$ 

Hence,

$$\|\hat{w}_t - w_t\| = o(1) \text{ a.s.}$$
Similarly we have
\[ \| \hat{m}_t - m_t \| = o(1) \text{ a.s.} \]

Then by adding and subtracting terms, we write
\[
\hat{\epsilon}_t = \hat{m}_t - \hat{A}\hat{w}_t
\]
\[
= m_t + (\hat{m}_t - m_t) - A\hat{w}_t - (\hat{A}_K - A)w_t - A(\hat{w}_t - w_t)
\]
\[
- (\hat{A}_K - A)(\hat{w}_t - w_t)
\]
\[
= \epsilon_t + (\hat{m}_t - m_t) - (\hat{A}_K - A)w_t - A(\hat{w}_t - w_t) - (\hat{A}_K - A)(\hat{w}_t - w_t).
\]

It is then clear that
\[
\hat{\epsilon}_t = \epsilon_t + o(1), \text{ a.s.}
\]

Hence
\[
\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t \otimes \hat{\epsilon}_t
\]
\[
= \frac{1}{T} \sum_{t=1}^{T} \epsilon_t \otimes \epsilon_t + o(1) \longrightarrow \text{a.s.} \Sigma.
\]

The last strong convergence again invokes Theorem 2.4 in Bosq (2000). \qed

**Proof for Theorem 7** Write
\[
\hat{m}_t = A\hat{w}_t + \epsilon_t + (\hat{m}_t - m_t) - A(\hat{w}_t - w_t),
\]
from which we may readily deduce that
\[
\hat{P}^* = \hat{Q}A^* + S + (R_1 + R_2 + R_3), \tag{34}
\]
where
\[
S = \frac{1}{T} \sum_{t=1}^{T} w_t \otimes \epsilon_t \tag{35}
\]
and
\[
R_1 = \frac{1}{T} \sum_{t=1}^{T} (\hat{w}_t - w_t) \otimes \epsilon_t
\]
\[
R_2 = \frac{1}{T} \sum_{t=1}^{T} \hat{w}_t \otimes (\hat{m}_t - m_t)
\]
\[
= \frac{1}{T} \sum_{t=1}^{T} w_t \otimes (\hat{m}_t - m_t) + \frac{1}{T} \sum_{t=1}^{T} (\hat{w}_t - w_t) \otimes (\hat{m}_t - m_t)
\]
\[
R_3 = - \left[ \frac{1}{T} \sum_{t=1}^{T} \hat{w}_t \otimes (\hat{w}_t - w_t) \right] A^*
\]
\[
= - \left[ \frac{1}{T} \sum_{t=1}^{T} w_t \otimes (\hat{w}_t - w_t) \right] A^* - \left[ \frac{1}{T} \sum_{t=1}^{T} (\hat{w}_t - w_t) \otimes (\hat{w}_t - w_t) \right] A^*.
\]
It can be shown that the terms $R_1, R_2$ and $R_3$ in (34) are asymptotically negligible. Therefore, we ignore them in our subsequent asymptotic analysis.

Under this convention, we may deduce from (34) and

$$\hat{A}_K^* = \hat{Q}_K^{-1} \hat{P}^*$$

that

$$\hat{A}_K^* = \hat{\Pi}_K A^* + \hat{Q}_K^{-1} S.$$

Consequently, it follows that

$$\hat{Q}_K^{1/2} (\hat{A}_K^* - A^*) = \hat{Q}_K^{1/2} S.$$  \hspace{1cm} (36)

Note in particular that $\hat{Q}_K^{1/2} \hat{\Pi}_K = \hat{Q}_K^{1/2}$. Furthermore, we have

$$\hat{Q}_K^{1/2} S = Q_K^{1/2} S + o_p(K^{-1/2})$$ \hspace{1cm} (37)

as we will show subsequently.

To establish (37), we rely on the results from perturbation theory established in Cardot, Mas, and Sarda (2005) and Mas (2006). Denote by $B_k$ and $\hat{B}_k$, respectively, the oriented circle of the complex plane with centers and radii $(\lambda_k, \delta_k/2)$ and $(\hat{\lambda}_k, \delta_k/2)$, and define

$$B = \bigcup_{k=1}^{K} B_k \quad \text{and} \quad \hat{B} = \bigcup_{k=1}^{K} \hat{B}_k.$$

Results from perturbation theory yield

$$Q_K^{1/2} = \frac{1}{2\pi i} \int_{B} z^{-1/2} (z - Q)^{-1} dz \quad \text{and} \quad \hat{Q}_K^{1/2} = \frac{1}{2\pi i} \int_{\hat{B}} z^{-1/2} (z - \hat{Q})^{-1} dz$$

and it can be shown that

$$\left( \hat{Q}_K^{1/2} - Q_K^{1/2} \right) S = \frac{1}{2\pi i} \int_{B} \left[ (z - Q)^{-1} - (z - Q)^{-1} \right] S dz \quad \text{and} \quad \hat{Q}_K^{1/2} S = Q_K^{1/2} S + o_p(K^{-1/2})$$

with probability approaching to one. The reader is referred to the proof of Proposition 6.1 in Cardot, Mas and Sarda (2005) and the proof of Lemma 5.6 in Mas (2006) for the details. Our framework is different from theirs in that we rely on estimated functional data, but this does not affect the proof.

We further analyze the right hand side of the equation in (38). After some simple algebra, we may readily deduce

$$(z - \hat{Q})^{-1} - (z - Q)^{-1} = (z - Q)^{-1} (Q - \hat{Q}) (z - \hat{Q})^{-1}$$

$$= (z - Q)^{-1} \left[ (z - Q)^{-1/2} (Q - \hat{Q}) (z - Q)^{-1/2} \right] \cdot \left[ (z - Q)^{1/2} (z - \hat{Q})^{-1} (z - Q)^{-1/2} \right] (z - Q)^{-1/2}. \hspace{1cm} (39)$$
Moreover, it follows that
\[
(z - Q)^{1/2}(z - \hat{Q})^{-1}(z - Q)^{1/2} = 1 + (z - Q)^{-1/2}(Q - \hat{Q})(z - Q)^{-1/2},
\] (40)
whenever it is invertible. The reader is referred to e.g., Cardot, Mas and Sarda (2005) [see the proof of Lemma 6.4, pages 17-18], for the derivation. Note that the inverse in (40) exists if \|((z - Q)^{-1/2}(Q - \hat{Q})(z - Q)^{-1/2})\| < 1.

Using (28) we can easily establish that, for all large \(z\),
\[
E \sup_{z \in B_k} \| (z - Q)^{-1/2}(Q - \hat{Q})(z - Q)^{-1/2} \|^2 \leq c \frac{(k \log k)^2}{T},
\]
(41)
where \(c > 0\) is some constant independent of \(k\). See Lemma 6.3 of Cardot, Mas and Sarda (2005). Therefore, if we define
\[
E_k = \left\{ \sup_{z \in B_k} \| (z - Q)^{-1/2}(Q - \hat{Q})(z - Q)^{-1/2} \| < 1/2 \right\}
\]
for \(k = 1, \ldots, K\), then it follows from Chebyshev inequality and (41) that
\[
P \left( \bigcup_{k=1}^K E_k^c \right) \leq \sum_{k=1}^K P(E_k^c) \leq \frac{4C}{T} \sum_{k=1}^K k^2 \log^2 k = O \left( \frac{K^3 \log^2 K}{T} \right),
\]
which is asymptotically negligible under condition (b). For our subsequent analysis, we may therefore focus on \(E = \bigcap_{k=1}^K E_k\). In particular, we have
\[
(z - Q)^{-1} - (z - Q)^{-1} = (z - Q)^{-1}(Q - \hat{Q})(z - Q)^{-1}
\]
(42)
with probability approaching to one. Note that we have on \(E\)
\[
\max_{1 \leq k \leq K} \sup_{z \in B_k} \left\| (z - Q)^{1/2}(z - \hat{Q})^{-1}(z - Q)^{1/2} \right\|
\]
\[
\leq \max_{1 \leq k \leq K} \left[ 1 - \sup_{z \in B_k} \left\| (z - Q)^{-1/2}(Q - \hat{Q})(z - Q)^{-1/2} \right\| \right]^{-1} \leq 2,
\]
due in particular to (40).

Hence,
\[
\int_{B_k} z^{-1/2} \left[ (z - \hat{Q})^{-1} - (z - Q)^{-1} \right] Sdz
\]
\[
= \int_{B_k} (z - Q)^{-1/2} \left[ (z - Q)^{-1/2}(Q - \hat{Q})(z - Q)^{-1/2} \right]
\]
\[
\cdot \left[ (z - Q)^{1/2}(z - \hat{Q})^{-1}(z - Q)^{1/2} \right] \left[ z^{-1/2}(z - Q)^{-1/2}S \right] dz.
\]
\[
= \int_{B_k} (z - Q)^{-1/2} \left[ (z - Q)^{-1/2}(Q - \hat{Q})(z - Q)^{-1/2} \right] \left[ z^{-1/2}(z - Q)^{-1/2}S \right] dz
\]
(43)
with probability approaching to one, under condition (b).

We have for all large $k$

$$\mathbb{E} \sup_{z \in B_k} \left\| z^{-1/2} (z - Q)^{-1/2} S \right\|^2 \leq c \frac{k \log k}{T}$$

(44)

with some constant $c > 0$ independent of $k$. A similar result was shown in Lemma 5.4 of Mas (2005).

We also have

$$\sup_{z \in B_k} \left\| (z - Q)^{-1/2} \right\| \leq \delta_k^{-1/2}.$$

(45)

It follows from (43), (41), (44) and (45) that

$$C := \mathbb{E} \left\| \int_{B_k} z^{-1/2} \left[ (z - \hat{Q})^{-1} - (z - Q)^{-1} \right] S dz \right\|$$

$$\leq \mathbb{E} \left\| \int_{B_k} (z - Q)^{-1/2} \left\| (z - Q)^{-1/2} (Q - \hat{Q}) (z - Q)^{-1/2} \right\| \left\| z^{-1/2} (z - Q)^{-1/2} S \right\| dz \right\|$$

$$\leq \int_{B_k} (z - Q)^{-1/2} \left( \sup_{z \in B_k} \mathbb{E} \left\| (z - Q)^{-1/2} (Q - \hat{Q}) (z - Q)^{-1/2} \right\|^2 \right)^{1/2}$$

$$\cdot \left( \sup_{z \in B_k} \mathbb{E} \left\| z^{-1/2} (z - Q)^{-1/2} S \right\| \right)^{1/2} dz$$

$$\leq c \frac{\delta_k^{1/2} k^{3/2} \log^{3/2} k}{T}$$

for some constant $c > 0$. Then,

$$\mathbb{E} \sum_{k=1}^{K} \left\| \int_{B_k} z^{-1/2} \left[ (z - \hat{Q})^{-1} - (z - Q)^{-1} \right] S dz \right\| \leq c \frac{1}{T} \sum_{k=1}^{K} \delta_k^{1/2} k^{3/2} \log^{3/2} k$$

$$\leq c \frac{K^{5/2} \log^{3/2} K}{T}.$$

Therefore, (37) follows immediately from under conditions (b) and (c), as $T \to \infty$.

It follows from (35), (36) and (37) that

$$\sqrt{T} \hat{Q}_K^{1/2} \langle \hat{A}_K^* - A^* \rangle v = \sqrt{T} Q_K^{1/2} Sv$$

$$= \sqrt{T} Q_K^{1/2} Sv + o_p(K^{-1/2})$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Q_{t}^{1/2} w_t \langle v, \varepsilon_t \rangle + o_p(K^{-1/2}),$$

(46)

and therefore, we have

$$\sqrt{T} \langle u_k, \hat{Q}_K^{1/2} (\hat{A}_K^* - A^*) \rangle v = \langle u_k, \sqrt{T} Q_K^{1/2} (\hat{A}_K^* - A^*) v \rangle$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \langle u_k, Q_{K}^{1/2} w_t \rangle \langle v, \varepsilon_t \rangle + o_p(K^{-1/2})$$

(47)
uniformly in $k = L + 1, \ldots, K$. In (47), we may readily observe that

(a) $\langle u_k, Q_K^{\frac{1}{2}} w_\ell \rangle$ for $k = L + 1, \ldots, K$ is a martingale difference sequence, and

(b) $\langle u_k, Q_K^{\frac{1}{2}} w_\ell \rangle$ for $k = L + 1, \ldots, K$ is uncorrelated.

For (b), note that

$$\mathbb{E} \left( Q_K^{\frac{1}{2}} u_t \otimes Q_K^{\frac{1}{2}} u_t \right) = \Pi_K,$$

and therefore,

$$\mathbb{E} \langle u_i, Q_K^{\frac{1}{2}} w_\ell \rangle \langle u_j, Q_K^{\frac{1}{2}} w_\ell \rangle = \langle u_i, \Pi_K u_j \rangle = \langle u_i, u_j \rangle = \delta_{ij}$$

for any orthogonal vectors $u_i$ and $u_j$ on the subspace $V_K$ of $H$, where $\delta_{ij}$ is the kronecker delta.

Now we consider

$$M_K^2 = \sum_{k=L+1}^{K} \langle u_k, \hat{Q}_K^{\frac{1}{2}} (\hat{A}_K^*-A^*)v \rangle^2,$$

$$\hat{M}_K^2 = \sum_{k=L+1}^{K} \langle \hat{\nu}_k, \hat{Q}_K^{\frac{1}{2}} (\hat{A}_K^*-A^*)v \rangle^2 = \sum_{k=L+1}^{K} \langle \hat{\nu}_k, \hat{Q}_K^{\frac{1}{2}} \hat{A}_K^* v \rangle^2.$$

Note that $\hat{M}_K$ is the norm of the projection of $\hat{Q}_K^{\frac{1}{2}} (\hat{A}_K^* - A^*)v = \hat{Q}_K^{\frac{1}{2}} \hat{A}_K^* v$ into the orthogonal complement of $\hat{Q}_K^{\frac{1}{2}} U$ on the subspace $\hat{V}_K$ of $H$. Likewise, $M_K$ is the norm of the projection of $\hat{Q}_K^{\frac{1}{2}} \hat{A}_K^* v$ into the orthogonal complement of $Q_K^{\frac{1}{2}} U$ on the subspace $V_K$ of $H$. Since projection is a contraction, we therefore have

$$\left| M_K - \hat{M}_K \right| \leq \left\| \hat{\Pi}_K \hat{Q}_K^{\frac{1}{2}} (\hat{A}_K^* - A^*)v - \Pi_K \hat{Q}_K^{\frac{1}{2}} (\hat{A}_K^* - A^*)v \right\|$$

$$\leq \left\| \hat{\Pi}_K - \Pi_K \right\| \left\| \hat{Q}_K^{\frac{1}{2}} (\hat{A}_K^* - A^*)v \right\|. \quad (48)$$

It follows from (46) that

$$\left\| \hat{Q}_K^{\frac{1}{2}} (\hat{A}_K^* - A^*)v \right\| = O_p(T^{-1/2}). \quad (49)$$

We may also easily deduce that

$$\left\| \hat{\Pi}_K - \Pi_K \right\| \leq K \sup_{1 \leq k \leq K} \| \hat{\nu}_k - v_k \| = O_p(\delta_K^{-1} T^{-1/2} K), \quad (50)$$

since

$$\hat{\Pi}_K - \Pi_K = \sum_{k=1}^{K} \hat{\nu}_k \otimes \hat{\nu}_k - \sum_{k=1}^{K} v_k \otimes v_k$$

$$= \sum_{k=1}^{K} v_k \otimes (\hat{\nu}_k - v_k) + \sum_{k=1}^{K} (\hat{\nu}_k - v_k) \otimes v_k + \sum_{k=1}^{K} (\hat{\nu}_k - v_k) \otimes (\hat{\nu}_k - v_k).$$
Moreover, we have
\[ M_K = O_p(T^{-1/2}K^{1/2}), \]  
(51)
due to (47) and the subsequent remarks. We may now deduce from (48), (49), (50) and (51) that
\[ \left| \hat{M}_K - M_K \right|^2 = \left| M_K - M_K \right|^2 + 2M_K \left| \hat{M}_K - M_K \right| = O_p(\delta_K^{-1}T^{-3/2}K^{3/2}), \]  
(52)
under condition (c).

We have
\[ Z = \frac{1}{\sqrt{2(K-L)}} \sum_{k=L+1}^{K} \left( \frac{T(\hat{u}_k, \hat{Q}_K^{1/2}(\hat{A}_k^*-A)v)^2}{\langle v, \Sigma v \rangle} - 1 \right) \]
\[ = \frac{1}{\sqrt{2(K-L)}} \left[ \frac{T\hat{M}_K}{\langle v, \Sigma v \rangle} - (K-L) \right] \]
\[ = \frac{1}{\sqrt{2(K-L)}} \left[ \frac{TM_K}{\langle v, \Sigma v \rangle} - (K-L) \right] + O_p(\delta_K^{-1}T^{-1/2}K) \]
\[ = \frac{1}{\sqrt{2(K-L)}} \sum_{k=L+1}^{K} \left( \frac{T(u_k, \hat{Q}_K^{1/2}(\hat{A}_k^*-A)v)^2}{\langle v, \Sigma v \rangle} - 1 \right) + o_p(1) \]  
(53)
using (52) and the fact that \( \|\hat{\Sigma} - \Sigma\| = o(1) \) a.s. under our assumption.

Due to (47) and (53), we have
\[ Z = \frac{1}{\sqrt{2(K-L)}} \sum_{k=1}^{K} (Z_k^2 - 1) + o_p(1), \]  
(54)
where we set
\[ Z_k = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\langle u_k, \hat{Q}_K^{1/2}w_t \rangle \langle v, \epsilon_t \rangle}{\langle v, \epsilon_t \rangle} \]
for \( k = L + 1, \ldots, K \). Assume \( \mathbb{E}\|\epsilon_t\|^{4+\epsilon} < \infty \) for some \( \epsilon > 0 \). Then we may have, using the probability embedding as in Lemma A3 of Park (2003) for a martingale difference sequence,
\[ Z_k = d N_k(0,1) + O_p(T^{-1/4}) \]
uniformly in \( k = L + 1, \ldots, K \), where \( (N_k(0,1)) \) for \( k = L + 1, \ldots, K \) are independent standard normal random variates. In what follows, we assume that \( (Z_k) \) and \( (N_k(0,1)) \) are defined in a common probability space. This causes no loss in generality, since we only concern the distribution of our statistic. Since
\[ Z_k^2 - N_k(0,1)^2 = [Z_k - N_k(0,1)]^2 + 2N_k(0,1)[Z_k - N_k(0,1)] \]
for \( k = L + 1, \ldots, K \), and since

\[
\sum_{k=L+1}^{K} |N_k(0,1)||Z_k - N(0,1)| \\
\leq \left( \sup_{L+1 \leq k \leq K} |N_k(0,1)| \right) \sum_{k=L+1}^{K} |Z_k - N(0,1)| = O_p(T^{-1/4}K^{1+\epsilon})
\]

for any \( \epsilon > 0 \), we have

\[
\frac{1}{\sqrt{2(K-L)}} \sum_{k=1}^{K} Z_k^2 = \frac{1}{\sqrt{2(K-L)}} \sum_{k=1}^{K} N_k(0,1)^2 + O_p(T^{-1/4}K^{1/2+\epsilon})
\]

for any \( \epsilon > 0 \). Note that

\[
\sup_{L+1 \leq k \leq K} |N_k(0,1)| = O_p(K^\epsilon),
\]

since \( N_k(0,1) \) has finite moment up to arbitrary integral order. We may now easily deduce from (54) and (55) that

\[
Z =_{d} \frac{1}{\sqrt{2(K-L)}} \sum_{k=L+1}^{K} \left[ N_k(0,1)^2 - 1 \right] + o_p(1) \to_d N(0,1)
\]

as was to be shown. The proof is therefore complete.