Submitted to the Annals of Applied Probability
arXiv: math.PR/0000000

TECHNICAL SUPPLEMENT TO "GOODNESS OF FIT FOR MARKOV MODELS"

BY VANCE MARTIN YOSHIHIKO NISHIYAMA AND JOHN STACHURSKI

This supplement provides some background material and proofs for the paper "Goodness of Fit for Markov Models: A Density Approach". Notation not defined here is defined in that paper. This supplement, the main paper and all computer code is available from the website `http://johnstachurski.net/papers/lae_test.html`.

1. Computing Critical Values by Simulation. We begin with the case of a simple null hypothesis p (i.e., the test without estimated parameters). To compute the critical value c_α^Λ defined in the paper we can use algorithm 1, which simulates the test statistic under the null hypothesis and estimates the critical value as a quantile of the observations.

Algorithm 1: Approximates c_α^Λ corresponding to given size α and kernel p

fix integers M, N;
for $m \in \{1, \ldots, M\}$ **do**
  simulate a p-Markov time series $X_1^*, \ldots, X_N^*$;
  set $T_N^m \leftarrow N^{-1} \int \{\sum_{t=1}^N \bar{p}(X_t^*, y)\}^2 dy$;
end
return the $1 - \alpha$ quantile of $T_N^1, \ldots, T_N^M$;

Turning to the estimated parameter case, we need a means of evaluating $c_\alpha^\Sigma(\theta)$ for given θ. As discussed in the main paper, one possibility is to compute the eigenvalues $\{\sigma_\ell^\theta\}_{\ell \geq 1}$ of Σ_θ, and then the $1 - \alpha$ quantile of $\sum_{\ell=1}^\infty \sigma_\ell^\theta Z_\ell^2$. A simpler method is to use simulation of the test statistic. This is accomplished as in algorithm 2, which returns an approximation to $c_\alpha^\Sigma(\theta)$. The estimator used in the algorithm to estimate θ from the simulated data should be the same estimator used for the observed data.

Algorithm 2: Approximates $c_\alpha^\Sigma(\theta)$ corresponding to given size α and kernel p_θ

fix integers M, N;
for $m \in \{1, \ldots, M\}$ **do**
  simulate a p_θ-Markov time series $X_1^*, \ldots, X_N^*$;
  fit $\hat{\theta}_n^*$ using the simulated data set $X_1^*, \ldots, X_N^*$;
  set $\hat{T}_N^m \leftarrow N^{-1} \int \{\sum_{t=1}^N \bar{p}(\hat{\theta}_n^*, X_t^*, y)\}^2 dy$;
end
return the $1 - \alpha$ quantile of $\hat{T}_N^1, \ldots, \hat{T}_N^M$;

Algorithm 2 computes $c_\alpha^\Sigma(\theta)$ for given θ. To evaluate the particular critical value $c_\alpha^\Sigma(\hat\theta_n)$ needed for the LAE test, the parameter vector θ in algorithm 2 should be replaced with the vector $\hat\theta_n$ estimated from the data. On standard hardware and using compiled C code based on the GNU Scientific Library, a typical calculation with $M = 2000$ and $N = 1000$ completes in several seconds. With interpreted languages the same calculation takes around 20 seconds. (These computations refer to a Gaussian AR(1) null hypothesis, and were implemented on a standard workstation. Numerical integration was based on a 60 point fixed-order Gauss-Legendre integration routine.)

2. The Numerical Experiments. In this section we provide details sufficient to replicate all of the Monte Carlo experiments reported in the paper.

2.1. *The Experiments in Section 5.1.* For details on Aït-Sahalia's test statistic and critical value, see Aït-Sahalia (1996, p. 393). The bandwidth for the nonparametric kernel density estimator used in our simulation was the optimal bandwidth for estimating the stationary density of the Vasicek model with the true parameters. (We experimented with other bandwidths but all choices gave similar rejection rates.) For the LAE test, we first computed the asymptotic critical value $c_\alpha^\Sigma(\theta_0)$ for the test with estimated parameters, setting $\alpha = 0.05$. To compute this value we used algorithm 2 applied to the baseline Vasicek density kernel discussed in section 5.1 and $M = N = 5000$. Next, we simulated 1,000 times series $\{X\}_{t=1}^n$ from the same DGP, where $n = 264$. For each of these simulated time series, we used least squares to obtain an estimate $\hat\theta_n$ for the vector of parameters of the Vasicek model, and then used the resulting density kernel $p_{\hat\theta_n}$ to evaluate the test statistic. Of the 1,000 time series we generated, 4.1% of the test statistics exceeded the asymptotic critical value.

2.2. *The Experiments in Section 5.2.* For the alternative, 1,000 time series of length n were generated from the CIR model with the specified parameters and monthly frequency. In the LAE test, for each time series, the parameters in the Vasicek null were estimated by ordinary least squares, and the estimated parameter version of the LAE test statistic was evaluated. In evaluating the test, the critical value $c_\alpha^\Sigma(\hat\theta_n)$ was calculated via algorithm 2, with $N = n$ and $M = 1000$. Size-adjusted critical values for the Cramér von Mises test and Aït-Sahalia's test where produced by the same method, changing only the definition of the test statistic. Our version of Aït-Sahalia's test statistic was the absolute value of the normalized statistic (cf., e.g., Pritsker, 1998, p. 455). The Cramér von Mises test statistic was $\int(\Psi_n(y) - \Psi(\hat\theta_n, y))^2\Psi(\hat\theta_n, dy)$ where $\Psi(\hat\theta_n, y)$ represents the theoretical stationary cdf under the Vasicek null at the estimated parameters, and Ψ_n denotes the empirical distribution of the data. In the case of Aït-Sahalia's test, we used Silverman's rule for the bandwidth and a standard normal Gaussian density for the kernel.

In simulating the discretized CIR process, we included a reflecting barrier at zero to avoid taking the square root of negative numbers.

3. Computation of ψ. In the main paper it is mentioned that for discrete time Markov models, the stationary density ψ that forms part of the test statistic may be intractable. In this case, one possibility is to approximate ψ via simulation. Since it changes little, we limit the discussion to the case of a simple null hypothesis. Suppose that p is V-mixing and $\{X_t\}$ is p-Markov. Fix $k \in \mathbb{N}$, and let $\{X_t^*\}_{t=1}^{kn}$ be a simulated p-Markov sequence that is independent of the data $\{X_t\}_{t=1}^{n}$. For each $k \in \mathbb{N}$ we have the following result:

THEOREM 3.1. *If $\{Z_\ell\}_{\ell \geq 1}$ is an* IID *sequence of standard normal random variables, then, as $n \to \infty$,*

$$\frac{1}{n} \int \left\{ \sum_{t=1}^{n} p(X_t, y) - \frac{1}{k} \sum_{t=1}^{nk} p(X_t^*, y) \right\}^2 dy \xrightarrow{\mathscr{D}} (1 + 1/k) \sum_{\ell=1}^{\infty} \lambda_\ell Z_\ell^2 \tag{1}$$

This is a version of the test statistic for the simple null hypothesis that does not use ψ. The limit $(1+1/k)\sum_{\ell=1}^{\infty} \lambda_\ell Z_\ell^2$ of the simulation-based test statistic (1) converges almost surely to that of the original test statistic T_n as $k \to \infty$.

Prior to proving theorem 3.1, we make some preliminary comments. Since any covariance operator C is linear, positive, symmetric and Hilbert-Schmidt on L_2 (see, e.g., Bosq, 2000, theorem 1.7), we can obtain the decomposition

$$Ch = \sum_{\ell=1}^{\infty} \lambda_\ell \langle h, v_\ell \rangle v_\ell \qquad (h \in L_2) \tag{2}$$

where $\{v_\ell\}_{\ell \geq 1}$ is an orthonormal basis of L_2 consisting of eigenfunctions of C, and $\{\lambda_\ell\}_{\ell \geq 1}$ is the corresponding eigenvalues. The eigenvalues are real, nonnegative, and satisfy $\sum_{\ell \geq 1} \lambda_\ell < \infty$. It is also well-known that if $\{Z_\ell\}$ are IID and standard normal, then

$$G := \sum_{\ell=1}^{\infty} \lambda_\ell^{1/2} Z_\ell v_\ell \sim N(0, C) \tag{3}$$

PROOF OF THEOREM 3.1. Let data $\{X_t\}_{t=1}^{n}$ be p-Markov. Define $\psi_n(y) := \frac{1}{n} \sum_{t=1}^{n} p(X_t, y)$ and $\psi'_{kn}(y) := \frac{1}{n} \sum_{t=1}^{kn} p(X'_t, y)$. Let $\{U_\ell\}_{\ell \geq 1}$ and $\{(U'_\ell\}_{\ell \geq 1}$ be mutually independent IID sequences of standard normal random variables. Fix $k \in \mathbb{N}$, and consider the decomposition

$$n^{1/2}(\psi_n - \psi'_{kn}) = n^{1/2}(\psi_n - \psi) - k^{-1/2}(kn)^{1/2}(\psi'_{kn} - \psi)$$

Note that $n^{1/2}(\psi_n - \psi)$ and $(kn)^{1/2}(\psi'_{kn} - \psi)$ are independent random functions in L_2. By theorem 2.1 in the main paper and the representation (3) we have

$$n^{1/2}(\psi_n - \psi) \xrightarrow{\mathscr{D}} \sum_\ell \lambda_\ell^{1/2} U_\ell v_\ell \quad \text{and} \quad (kn)^{1/2}(\psi'_{kn} - \psi) \xrightarrow{\mathscr{D}} \sum_\ell \lambda_\ell^{1/2} U'_\ell v_\ell$$

By independence and continuity of addition and scalar multiplication in L_2, we then have

$$n^{1/2}(\psi_n - \psi'_{nk}) \xrightarrow{\mathscr{D}} \sum_\ell \lambda_\ell^{1/2} U_\ell - k^{-1/2} \sum_\ell \lambda_\ell^{1/2} U'_\ell v_\ell = \sum_\ell \lambda_\ell^{1/2} (U_\ell - k^{-1/2} U'_\ell) v_\ell$$

Applying the continuous mapping theorem and the Pythagorean law, we obtain

$$n\|\psi_n - \psi'_{nk}\|^2 \xrightarrow{\mathscr{D}} \| \sum_\ell \lambda_\ell^{1/2}(U_\ell - k^{-1/2}U'_\ell)v_\ell\|^2 = \sum_\ell \lambda_\ell (U_\ell - k^{-1/2}U'_\ell)^2$$

The left-hand side of this equation is equal to the left-hand side of (1). Moreover, if Z_ℓ is standard normal, then $(1+1/k)Z_\ell^2$ and $(U_\ell - k^{-1/2}U'_\ell)^2$ have the same law. This completes the proof of (1). □

4. Local Alternatives. Now we turn to the proof of our result on local alternatives (theorem 3.1). First we recall the definition of geometric ergodicity. For a given density kernel p, ergodicity requires that p has a unique stationary density ψ, and any p-Markov process satisfies the strong law of large numbers (see Meyn and Tweedie, 2009, theorem 17.1.7, and Lindvall, 2002, theorem 21.12 for the many equivalent definitions of ergodicity). Geometric ergodicity requires that, in addition, there exist positive constants $\lambda < 1$ and $L < \infty$ and a weight function $V\colon \mathbb{X} \to \mathbb{R}_+$ such that

$$\int V(x)\psi(x)dx < \infty \quad \text{and} \quad \left| \int_B p^t(x,y)dy - \int_B \psi(y)dy \right| \leq \lambda^t L V(x) \tag{4}$$

for all $B \in \mathscr{X}$, $x \in \mathbb{X}$ and $t \in \mathbb{N}$.

We will make use of the following elementary result. The result seems to be well-known, but a proof is included for completeness.

LEMMA 4.1. *Let $\{X_t\}$ be a geometrically ergodic Markov process on $\mathbb{X}$ and let V be the weight function in (4). If $\hat{X}_t := (X_t, \dots, X_{t+r})$, then $\{\hat{X}_t\}$ is Markov and geometrically ergodic on $\mathbb{X}^{r+1}$ with weight function $\hat{V}(x_0, \dots, x_r) := \sum_{k=0}^r V(x_k)$.*

PROOF. To see that $\{\hat{X}_t\}$ is Markov, pick any bounded measurable $h\colon \mathbb{X}^{r+1} \to \mathbb{R}$. We have

$$\begin{aligned} \mathbf{E}[h(\hat{X}_t) \mid \hat{X}_{t-1}, \dots, \hat{X}_1] &= \mathbf{E}[h(X_t, \dots, X_{t+r}) \mid X_{t+r-1}, \dots, X_1] \\ &= \mathbf{E}[h(X_t, \dots, X_{t+r}) \mid X_{t+r-1}, \dots, X_{t-1}] \end{aligned}$$

where the second equality is due to the Markov property of $\{X_t\}$. Changing notation we can write this as

$$\mathbf{E}[h(\hat{X}_t) \mid \hat{X}_{t-1}, \dots, \hat{X}_1] = \mathbf{E}[h(\hat{X}_t) \mid \hat{X}_{t-1}]$$

Since h is an arbitrary bounded measurable function, we have shown that $\{\hat{X}_t\}$ is Markov as claimed.

Next consider geometric ergodicity. Let V and $\hat{V}$ be as in the statement of the lemma. Pick any measurable $h\colon \mathbb{X}^{r+1} \to \mathbb{R}$ such that $|h| \leq 1$. Fix $(x_0, \dots, x_r) \in \mathbb{X}^{r+1}$. Let $\psi_t := p^{t-r}(x_r, \cdot)$, which is the density of X_t given $X_r = x_r$. Observing that

$$\hat{\psi}(x_t, \dots, x_{t+r}) := \psi(x_t)p(x_t, x_{t+1}) \cdots p(x_{t+r-1}, x_{t+r})$$

is the stationary density of $\hat{X}_t$ and

$$\hat{\psi}_t(x_t, \ldots, x_{t+r}) := \psi_t(x_t) p(x_t, x_{t+1}) \cdots p(x_{t+r-1}, x_{t+r})$$

is the density of $\hat{X}_t$ given $\hat{X}_0 = (x_0, \ldots, x_r)$, we then have

$$\left| \int h\hat{\psi}_t - \int h\hat{\psi} \right| = \left| \int g\psi_t - \int g\psi \right| \tag{5}$$

for $g(x_t) := \int \cdots \int h(x_t, \ldots, x_{t+r}) p(x_t, x_{t+1}) \cdots p(x_{t+r-1}, x_{t+r}) dx_{t+1} \cdots dx_{t+r}$. Given that $|h| \leq 1$, we also have $|g| \leq 1$, and therefore

$$\left| \int g\psi_t - \int g\psi \right| \leq 2 \sup_{B \in \mathcal{X}} \left| \int_B \psi_t - \int_B \psi \right| = 2 \sup_{B \in \mathcal{X}} \left| \int_B p^{t-r}(x_r, y) dy - \int_B \psi(y) dy \right|$$

From this bound, (5) and (4), we then have

$$\left| \int h\hat{\psi}_t - \int h\hat{\psi} \right| \leq 2\lambda^{t-r} L V(x_r) \leq 2\lambda^{t-r} L \sum_{k=0}^{r} V(x_k) = \lambda^t \left(\frac{2L}{\lambda^r} \right) \hat{V}(x_0, \ldots, x_r)$$

Specializing to $h = \mathbb{1}_B$ and recalling (4), we see that $\{\hat{X}_t\}$ satisfies the condition on the right-hand side of (4). Regarding the finiteness condition on the left-hand side of (4), observe that $\mathbf{E}\hat{V}(\hat{X}_t) = \sum_{k=0}^{r} \mathbf{E}V(X_{t+r}) = r \int V d\psi$. This term is finite by geometric ergodicity of the original process $\{X_t\}$. We conclude that $\{\hat{X}_t\}$ is geometrically ergodic with weight function $\hat{V}$. □

Now we return to the local alternatives result. Recall our notation. We set

$$Y_n(y) := n^{-1/2} \sum_{t=1}^{n} \bar{p}(X_t, y) \qquad (y \in \mathbb{X})$$

and let τ be the element of L_2 defined by

$$\tau(y) := \sum_{t=1}^{\infty} \mathbf{E} \left\{ \bar{p}(X_{t+1}, y) \frac{k(X_1, X_2)}{p(X_1, X_2)} \right\}$$

ASSUMPTION 4.1. Together, k and p satisfy the third moment condition

$$\mathbf{E} \sup_{\delta \in [0,1]} \frac{|k(X_1, X_2)|^3}{|p(X_1, X_2) + \delta\, k(X_1, X_2)|^3} < \infty$$

The theorem we aim to prove is restated here for convenience:

THEOREM 4.1. *If H_1 and assumption 4.1 both hold, then $Y_n \xrightarrow{\mathscr{D}} N(\tau, \Lambda)$.*

In assumption 4.1 and in the definition of τ, the expectation is taken under H_0. The exact meaning of the claim in theorem 4.1 can be clarified as follows: Let $(\Omega_n, \mathscr{F}_n)$ be the product space $\mathbb{X}^n := \times_{t=1}^{n} \mathbb{X}$ with its product σ-algebra, let $X_t \colon \Omega_n \to \mathbb{X}$ be the projection $X_t(x_1, \ldots, x_n) = x_t$, let $\mathbf{P}_n$ be the distribution of $(X_1, \ldots, X_n)$ over $\mathbb{X}^n$ constructed from p in H_0, and let $\mathbf{Q}_n$ be the distribution on $\mathbb{X}^n$ constructed from the local alternative p_n. (Construction of $\mathbf{P}_n$ and $\mathbf{Q}_n$ from their respective kernels is via the standard definition—see, e.g., Meyn and Tweedie, ch. 3, 2009.) The claim in theorem 4.1 is that, for all continuous bounded $g \colon L_2 \to \mathbb{R}$, we have $\int g(Y_n) d\mathbf{Q}_n \to \int g d\nu$ as $n \to \infty$, where ν is the L_2 Gaussian $N(\tau, \Lambda)$.

4.1. *Proof of Theorem 4.1.* The proof is long but ultimately very standard. We give full details only for completeness. Let $\mathscr{H}$ be defined as $L_2 \times \mathbb{R}$, with inner product

$$\langle g, h\rangle = \langle g_1, h_1\rangle + g_2 h_2 \qquad (g = (g_1, g_2) \text{ and } h = (h_1, h_2))$$

(Here $\langle g, h\rangle$ is the inner product in $\mathscr{H}$ and $\langle g_1, h_1\rangle$ is the inner product in L_2. The notation does not distinguish between them, but the meaning will be clear from context.) With the norm $\|h\| = \sqrt{\langle h, h\rangle}$, the space $\mathscr{H}$ is a Hilbert space, and the norm topology of $\mathscr{H}$ corresponds to the product topology of $L_2 \times \mathbb{R}$. The next result is an extension of the Cramér-Wold theorem to $\mathscr{H}$:

LEMMA 4.2. *Let $U_n := (Y_n, \ell_n)$ be a random sequence in $\mathscr{H}$, where Y_n is a random element of L_2 and ℓ_n is a random variable for all n. Let U be a Gaussian random element of $\mathscr{H}$ with distribution $N(m, S)$.*

$$\langle U_n, h\rangle \xrightarrow{\mathscr{D}} N(\langle h, m\rangle, \langle h, Sh\rangle) \text{ in } \mathbb{R} \text{ for all } h \in \mathscr{H} \implies U_n \xrightarrow{\mathscr{D}} U \text{ in } \mathscr{H}$$

PROOF. Suppose for the moment that $\{U_n\}$ is tight in $\mathscr{H}$. In this case, to show that U_n converges in distribution to U in $\mathscr{H}$, we need only show that $\langle U_n, h\rangle$ converges in distribution to $\langle U, h\rangle$ in $\mathbb{R}$ for all $h \in \mathscr{H}$ (Bosq, 2000, theorem 2.3). This is immediate, since $U \sim N(m, S)$, and hence $\langle U, h\rangle$ has distribution $N(\langle h, m\rangle, \langle h, Sh\rangle)$.

It remains to show that $\{U_n\}$ is tight in $\mathscr{H}$. To see that this is so, note that, as required in the lemma, $\langle U_n, h\rangle$ converges in distribution for all $h \in \mathscr{H}$. Choosing $h = (0, 1)$, we see that ℓ_n converges in distribution, and is therefore tight. Now fix $\epsilon > 0$. Since $\{Y_n\}$ and $\{\ell_n\}$ are both tight, we can find compact sets $K_a \subset L_2$ and $K_b \subset \mathbb{R}$ with $\mathbf{P}\{Y_n \notin K_a\} < \epsilon/2$ and $\mathbf{P}\{\ell_n \notin K_b\} < \epsilon/2$ for all n. The set $K_a \times K_b$ is compact in the product topology on $\mathscr{H}$, and we have

$$\mathbf{P}\{U_n \notin K_a \times K_b\} \le \mathbf{P}\{Y_n \notin K_a\} \cup \{\ell_n \notin K_b\} \le \mathbf{P}\{Y_n \notin K_a\} + \mathbf{P}\{\ell_n \notin K_b\} < \epsilon$$

We conclude that $\{U_n\}$ is tight in $\mathscr{H}$, completing the proof of lemma 4.2. □

LEMMA 4.3. *Let Y_n, $\mathbf{Q}_n$ and $\mathbf{P}_n$ be as defined above, let*

$$r(x, y) := \frac{k(x, y)}{p(x, y)}, \qquad \sigma^2 := \mathbf{E} r(X_t, X_{t+1})^2 = \int r(X_t, X_{t+1})^2 d\mathbf{P}_n$$

and let $\ell_n \colon \mathbb{X}^n \to \mathbb{R}$ be the log likelihood ratio

$$\ell_n = \log \frac{d\mathbf{Q}_n}{d\mathbf{P}_n} = \log \left\{ \frac{\prod_{t=2}^n p_n(X_{t-1}, X_t)}{\prod_{t=2}^n p(X_{t-1}, X_t)} \right\}$$

If $U_n := (Y_n, \ell_n)$ and $U = (Y, \ell)$ has distribution $N(m, S)$ for $m = (0, -\sigma^2/2)$ and S satisfying

$$\langle h, Sh\rangle = \langle h_1, \Lambda h_1\rangle + 2\langle \tau, h_1\rangle h_2 + h_2^2 \sigma^2 \qquad (h = (h_1, h_2) \in \mathscr{H})$$

then $U_n \xrightarrow{\mathscr{D}} U$ under $\mathbf{P}_n$.

PROOF. In what follows, all probabilities and expectations are evaluated under $\mathbf{P}_n$. We begin by obtaining a more convenient expression for the likelihood ratio ℓ_n. Writing p_t for $p(X_{t-1}, X_t)$ and k_t for $k(X_{t-1}, X_t)$, we have

$$\ell_n = \sum_{t=2}^{n} \{\log(p_t + k_t/\sqrt{n}) - \log(p_t)\}$$

Expanding the log function around p_t yields

$$\ell_n = \frac{1}{\sqrt{n}} \sum_{t=2}^{n} \frac{k_t}{p_t} - \frac{1}{2n} \sum_{t=2}^{n} \frac{k_t^2}{p_t^2} + \frac{1}{3n^{3/2}} \sum_{t=2}^{n} \frac{k_t^3}{[p_t + \lambda n^{-1/2} k_t]^3}$$

For some $\lambda \in [0,1]$. Since (X_{t-1}, X_t) is itself ergodic (see lemma 4.1) and

$$\frac{k_t^3}{[p_t + \lambda n^{-1/2} k_t]^3} \leq \sup_{\delta \in [0,1]} \frac{|k(X_{t-1}, X_t)|^3}{|p(X_{t-1}, X_t) + \delta\, k(X_{t-1}, X_t)|^3}$$

it follows from assumption 4.1 that $\frac{1}{n} \sum_{t=2}^{n} k_t^3 / [p_t + \lambda n^{-1/2} k_t]^3 = O_P(1)$, and hence

$$\ell_n = \frac{1}{\sqrt{n}} \sum_{t=2}^{n} r(X_{t-1}, X_t) - \frac{1}{2n} \sum_{t=2}^{n} r(X_{t-1}, X_t)^2 + o_P(1) \tag{6}$$

Now we return to the proof of lemma 4.3. Taking into account lemma 4.2 and the fact that $\{Y_n\}$ is tight in L_2, it suffices to show that

$$\langle U_n, h \rangle \xrightarrow{\mathscr{D}} N(-h_2\sigma^2/2, \langle h_1, \Lambda h_1 \rangle + 2\langle \tau, h_1 \rangle h_2 + h_2^2 \sigma^2) \tag{7}$$

for arbitrary $h \in \mathscr{H}$. Fixing such an $h = (h_1, h_2)$, the definition of U_n and our expression for ℓ_n in (6) gives

$$\langle U_n, h \rangle = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \langle h_1, \bar{p}(X_t, \cdot) \rangle + h_2 \frac{1}{\sqrt{n}} \sum_{t=2}^{n} r(X_{t-1}, X_t) - h_2 \frac{1}{2n} \sum_{t=2}^{n} r(X_{t-1}, X_t)^2 + o_P(1)$$

Since (X_{t-1}, X_t) is ergodic (lemma 4.1) we have

$$\frac{h_2}{2} \frac{1}{n} \sum_{t=2}^{n} r(X_{t-1}, X_t)^2 \to h_2 \frac{\sigma^2}{2} \quad \text{in probability}$$

As a result of this convergence and Slutsky's theorem, the result (7) will be confirmed if we show that

$$\frac{1}{\sqrt{n}} \sum_{t=2}^{n} q(X_{t-1}, X_t) \xrightarrow{\mathscr{D}} N(0, \langle h_1, \Lambda h_1 \rangle + 2\langle \tau, h_1 \rangle h_2 + h_2^2 \sigma^2) \tag{8}$$

for $q(X_{t-1}, X_t) := \langle h_1, \bar{p}(X_t, \cdot) \rangle + h_2 r(X_{t-1}, X_t)$. To see that this is indeed the case, recall from the main text that $\mathbf{E}\langle h_1, \bar{p}(X_t, \cdot) \rangle = 0$ under H_0. Also,

$$\mathbf{E} r(X_{t-1}, X_t) = \int \int \frac{k(x,y)}{p(x,y)} \psi(x) p(x,y)\, dx dy = \int \left[\int k(x,y) dy \right] \psi(x) dx = 0$$

It follows that $\mathbf{E}q(X_{t-1}, X_t) = 0$, and, as a result of geometric ergodicity of (X_{t-1}, X_t) (lemma 4.1) and the scalar CLT for geometrically ergodic Markov processes (e.g., Meyn and Tweedie, 2009, theorem 17.0.1), we have

$$\frac{1}{\sqrt{n}} \sum_{t=2}^{n} q(X_{t-1}, X_t) \xrightarrow{\mathscr{D}} N(0, v), \quad v := \mathbf{E}q(X_1, X_2)^2 + 2 \sum_{t=2}^{\infty} \mathbf{E}q(X_1, X_2)q(X_t, X_{t+1})$$

It remains only to show that $v = \langle h_1, \Lambda h_1 \rangle + 2\langle \tau, h_1 \rangle h_2 + h_2^2 \sigma^2$, which is the right-hand side of the variance in (8). Prior to proving this, we observe that all of the following statements are valid, and will be used without comment below:

- $\mathbf{E}r(X_1, X_2) = 0$ and $\mathbf{E}[r(X_t, X_{t+1}) \mid X_1] = 0$ for all $t \geq 1$.
- $\mathbf{E}[r(X_1, X_2)r(X_t, X_{t+1})] = 0$ for all $t \geq 2$.
- $\mathbf{E}[\langle h_1, \bar{p}(X_1, \cdot) \rangle r(X_t, X_{t+1}) = 0$ for all $t \geq 1$.

(The first of these statements has already been established above, and the proofs of the rest are similar.) Turning now to the evaluation of v, note that

$$\begin{aligned} \mathbf{E}q(X_1, X_2)^2 &= \mathbf{E}\{\langle h_1, \bar{p}(X_1, \cdot) \rangle^2 + 2\langle h_1, \bar{p}(X_1, \cdot) \rangle h_2 r(X_1, X_2) + h_2^2 r(X_1, X_2)^2\} \\ &= \mathbf{E}\langle h_1, \bar{p}(X_1, \cdot) \rangle^2 + h_2^2 \sigma^2 \end{aligned}$$

while, for any given $t \geq 2$,

$$\mathbf{E}q(X_1, X_2)q(X_t, X_{t+1}) = \mathbf{E}\langle h_1, \bar{p}(X_1, \cdot) \rangle \langle h_1, \bar{p}(X_t, \cdot) \rangle + \mathbf{E}\langle h_1, \bar{p}(X_t, \cdot) h_2 r(X_1, X_2)$$

As a result, we have

$$v = \mathbf{E}\langle h_1, \bar{p}(X_1, \cdot) \rangle^2 + 2 \sum_{t=2}^{\infty} \mathbf{E}\langle h_1, \bar{p}(X_1, \cdot) \rangle \langle h_1, \bar{p}(X_t, \cdot) \rangle + 2 \sum_{t=2}^{\infty} \mathbf{E}\langle h_1, \bar{p}(X_t, \cdot) h_2 r(X_1, X_2) + h_2^2 \sigma^2$$

By the definition of Λ,

$$\mathbf{E}\langle h_1, \bar{p}(X_1, \cdot) \rangle^2 + 2 \sum_{t=2}^{\infty} \mathbf{E}\langle h_1, \bar{p}(X_1, \cdot) \rangle \langle h_1, \bar{p}(X_t, \cdot) \rangle = \langle h_1, \Lambda h_1 \rangle$$

Finally, using the definition of τ, we obtain $v = \langle h_1, \Lambda h_1 \rangle + 2\langle \tau, h_1 \rangle h_2 + h_2^2 \sigma^2$. This verifies (8), and completes the proof of lemma 4.3. $\square$

LEMMA 4.4. *For ℓ_n, ℓ defined in lemma 4.3, we have $\ell_n \xrightarrow{\mathscr{D}} \ell$ and $\mathbf{E} \exp(\ell) = 1$ under $\mathbf{P}_n$.*

PROOF. We saw in lemma 4.3 that, under $\mathbf{P}_n$, we have $\langle h, U_n \rangle \xrightarrow{\mathscr{D}} \langle h, U \rangle$ for all $h \in \mathscr{H}$, where $U \sim N(m, S)$ for m and S defined in lemma 4.3. Specializing to $h = (0, 1)$ obtains the first claim in lemma 4.4. Regarding the second claim in lemma 4.4, for this same h we have $\ell = \langle h, U \rangle = N(\langle m, h \rangle, \langle h, Sh \rangle)$, and given the definitions of m and S in lemma 4.3, we have $N(\langle m, h \rangle, \langle h, Sh \rangle) = N(-\sigma^2/2, \sigma^2)$. Therefore,

$$\mathbf{E} \exp(\ell) = \exp\left(-\frac{\sigma^2}{2} + \frac{\sigma^2}{2}\right) = 1$$

when expectation is taken under $\mathbf{P}_n$. This completes the proof. $\square$

PROOF OF THEOREM 4.1. We saw in lemmas 4.3 and 4.4 that if $\ell_n = d\mathbf{Q}_n / d\mathbf{P}_n$ is the log likelihood ratio, then under $\mathbf{P}_n$ we have

$$\begin{pmatrix} Y_n \\ \ell_n \end{pmatrix} \xrightarrow{\mathscr{D}} \begin{pmatrix} Y \\ \ell \end{pmatrix} \sim N(m, S)$$

and, moreover, $\mathbf{E}\exp(\ell) = 1$. Applying the abstract version of Le Cam's third lemma presented in van der Vaart and Wellner (1996, theorem 3.10.7), we then have $Y_n \xrightarrow{\mathscr{D}} \pi$ under $\mathbf{Q}_n$ when π is the probability measure on L_2 defined by

$$\pi(f) = \mathbf{E}\exp(\ell) f(Y) \quad \text{for all bounded measurable } f \colon \mathbb{X} \to \mathbb{R}$$

To complete the proof of theorem 4.1, we need only show that $\pi := N(\tau, \Lambda)$. To see that this equality holds, let V be a random element on L_2 with $V \sim \pi$. To verify $\pi := N(\tau, \Lambda)$ it suffices to show that, for arbitrary fixed $a \in L_2$, we have

$$\langle a, V \rangle \sim N(\langle a, \tau \rangle, \langle a, \Lambda a \rangle) \tag{9}$$

To establish (9), observe that, from the definition of π, the moment generating function of $\langle a, V \rangle$ is

$$M(t) := \mathbf{E}\exp(t\langle a, V \rangle) = \mathbf{E}\exp(\ell)\exp(t\langle a, Y \rangle) = \mathbf{E}\exp(t\langle a, Y \rangle + \ell)$$

If $h \in \mathscr{H}$ is defined as $h := (ta, 1)$ and $U := (Y, \ell)$, then $\langle h, U \rangle$ is precisely $t\langle a, Y \rangle + \ell$. Since U is Gaussian, we know that $\langle h, U \rangle$ is Gaussian, and therefore $t\langle a, Y \rangle + \ell$ is Gaussian in $\mathbb{R}$. Its expectation and variance are given by

$$\mathbf{E}(t\langle a, Y \rangle + \ell) = \mathbf{E}\langle (ta, 1), U \rangle = \langle (ta, 1), m \rangle = \langle (ta, 1), (0, -\sigma^2/2) \rangle = -\sigma^2/2$$

and

$$\operatorname{var}\langle (ta, 1), U \rangle = \langle (ta, 1), S(ta, 1) \rangle = t^2 \langle a, \Lambda a \rangle + 2t\langle \tau, a \rangle + \sigma^2$$

where the final expression follows from the definition of S given in the statement of lemma 4.3. To finish the proof, we observe that, since $t\langle a, Y \rangle + \ell$ is Gaussian with mean and variance as derived above, we must have

$$\mathbf{E}\exp(t\langle a, Y \rangle + \ell) = \exp\left\{ -\frac{\sigma^2}{2} + t\langle \tau, a \rangle + t^2 \frac{\langle a, \Lambda a \rangle}{2} + \frac{\sigma^2}{2} \right\}$$

Cancelling the two instances of $\sigma^2/2$, we find that the moment generating function of $\langle a, V \rangle$ is

$$M(t) = \exp\left\{ t\langle \tau, a \rangle + t^2 \frac{\langle a, \Lambda a \rangle}{2} \right\} \qquad (t \in \mathbb{R})$$

This is precisely the moment generating function for the $N(\langle a, \tau \rangle, \langle a, \Lambda a \rangle)$ distribution, and hence we have established (9). This completes the proof of theorem 4.1. □

REFERENCES

[1] Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate, *Review of Financial Studies*, 9 (2), 385–426.

[2] Bosq, D. (2000). *Linear Processes in Function Space*, Springer-Verlag.

[3] Lindvall, T. (2002). *Lectures on the Coupling Method*, Dover Publications, Mineola N.Y.

[4] Meyn, S. P. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*, Springer-Verlag: London.

[5] Pritsker, M. (1998). Nonparametric density estimation and tests of continuous time interest rate models, *Review of Financial Studies*, 11 (3), 449–487.

[6] van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes*, Springer.

DEPARTMENT OF ECONOMICS, THE UNIVERSITY OF MELBOURNE,
E-MAIL: vance@unimelb.edu.au

INSTITUTE OF ECONOMIC RESEARCH, KYOTO UNIVERSITY,
E-MAIL: nishiyama@kier.kyoto-u.ac.jp

RESEARCH SCHOOL OF ECONOMICS, AUSTRALIAN NATIONAL UNIVERSITY,
E-MAIL: john.stachurski@anu.edu.au