Testing the Drift-Diffusion Model

Drew Fudenberg[a], Whitney Newey[a], Philipp Strack[b], and Tomasz Strzalecki[c]

[a]MIT; [b]Yale; [c]Harvard

This manuscript was compiled on June 3, 2020

The drift diffusion model (DDM) is a model of sequential sampling with diffusion signals, where the decision maker accumulates evidence until the process hits either an upper or lower stopping boundary, and then stops and chooses the alternative that corresponds to that boundary. In perceptual tasks the drift of the process is related to which choice is objectively correct, whereas in choice tasks the drift is related to the relative appeal of the alternatives. The simplest version of the DDM assumes that the stopping boundaries are constant over time. More recently a number of papers have used non-constant boundaries to better fit the data. This paper provides a statistical test for DDMs with general, nonconstant boundaries. As a byproduct, we show that the drift and the boundary are uniquely identified. We use our condition to nonparametrically estimate the drift and the boundary and construct a test statistic based on finite samples.

response times | drift-diffusion model | statistical test

DRAFT

The *drift diffusion model* (DDM) is a model of sequential sampling with diffusion (Brownian) signals, where the decision maker accumulates evidence until the process hits a stopping boundary, and then stops and chooses the alternative that corresponds to that boundary. This model has been widely used in psychology, neuroeconomics, and neuroscience to explain the observed patterns of choice and response times in a range of binary choice decision problems. One class of papers study "perception tasks"with an objectively correct answer e.g. "are more of the dots on the screen moving left or moving right?"; here the drift of the process is related to which choice is objectively correct (1, 2). The other class of papers study "consumption tasks" such as "which of these snacks would you rather eat?"; here the drift is related to the relative appeal of the alternatives (3–11).

The simplest version of the DDM assumes that the stopping boundaries are constant over time (12–15). More recently a number of papers use non-constant boundaries to better fit the data, and in particular the observed correlation between response times and choice accuracy, i.e., that correct responses are faster than incorrect responses (16–19).

Constant stopping boundaries are optimal for perception tasks where the volatility of the signals and the flow cost of sampling are both constant, and the prior belief is that the drift of the diffusion has only two possible values, depending on which decision is correct. Even with constant volatility and costs, non-constant boundaries are optimal for other priors, for example when the difficulty of the task varies from trial to trial and some decision problems are harder than others. (17) show how to computationally derive the optimal boundaries in this case. (18) characterize the optimal boundaries for the consumption task: the decision maker is uncertain about the utility of each choice, with independent normal priors on the value of each option.

This paper provides a statistical test for DDMs with general boundaries, without regard to their optimality. We first prove a characterization theorem: we find a condition on choice probabilities that is satisfied if and only if the choice probabilities are generated by some DDM. Moreover, we show that the drift and the boundary are uniquely identified. We then use our condition to nonparametrically estimate the drift and the boundary and construct a test statistic based on finite samples.

Recent related work on DDM includes (17) who conducted a Bayesian estimation of a collapsing boundary model and (18) who conducted a maximum likelihood estimation. (20) estimate collapsing boundaries in a parametric class, allowing for a random nondecision time at the start. (21) estimate a version of DDM with constant boundaries but random starting point of the signal accumulation process; (22) estimates a similar model where other parameters are made random. (23) partially characterize DDM with constant boundary.*

Other work on DDM-like models includes the decision field theory of (24–26), which allows the signal process to be mean-reverting. (27) and (28) study models where response time is a deterministic function of the utility difference. (29–34) study dynamic costly optimal information acquisition.

1. The Stochastic Choice Function

Let X be the universe of alternatives (actions) and $T = \mathbb{R}_+$ be time. For every pair of objects $\{x, y\}$ the analyst observes pairwise stochastic choices and decision times. In the limit as the sample size grows large, the analyst will have access to the joint distribution over which object is chosen and at which time a choice is made. We denote by $F^{xy}(t)$ the probability that the agent makes a choice by time t, and let $p^{xy}(t)$ the

*They ignore the issue of correlation between response times choices by looking only at marginal distributions, which makes their conditions necessary but not sufficient.

Significance Statement

The drift diffusion model (DDM) has been widely used in psychology and neuroeconomics to explain observed patterns of choices and response times. This paper provides the first identification and characterization theorems for this model: we show that the parameters are uniquely pinned down and determine which data sets are consistent with some form of DDM. We then develop a statistical test of the model based on finite data sets using spline estimation. These results establish the empirical content of the model and provide a way for researchers to see when it is applicable.

All authors designed research, performed research, contributed new analytic tools, and wrote the paper. DF, PS, and TS contributed Theorems 1 and 2; WN contributed Theorem 3.

The authors declare no conflict of interest.

[1]DF, WN, PS, and TS contributed equally to this work.

[1]To whom correspondence should be addressed. E-mail: author.twoemail.com

probability that the agent picks x conditional on stopping at time t. Throughout, we restrict attention to cases where F has full support and no atoms at time 0, so that $F(0) = 0$, and we assume that F is strictly increasing with $\lim_{t\to\infty} F(t) = 1$. These restrictions imply the agent never stops immediately, that there is a positive probability of stopping in every time interval, and that the agent always eventually stops. We call (p^{xy}, F^{xy}), the *stochastic choice function*.

An immediate restriction on the stochastic choice function is that the choices of the agent are unaffected by which object we consider to be the first and which object we consider to be the second. This is formally equivalent to

$$p^{xy}(t) \equiv 1 - p^{yx}(t) \text{ for all } t \text{ and } F^{xy} \equiv F^{yx} \text{ for all } x, y \in X.$$

Without loss of generality we only consider stochastic choice functions which satisfy this restriction. We also assume that each option is chosen with positive probability $0 < p^{xy}(t) < 1$ for all t.

Given (p^{xy}, F^{xy}) we define the *choice imbalance* at each time t to be

$$I^{xy}(t) := p^{xy}(t) \log\left(\frac{p^{xy}(t)}{1-p^{xy}(t)}\right) + (1-p^{xy}(t)) \log\left(\frac{1-p^{xy}(t)}{p^{xy}(t)}\right).$$

This is the Kullback-Leibler divergence (or relative entropy) between the Binomial distribution of the agent's time t choice $(p^{xy}(t), 1-p^{xy}(t))$ and the permuted choice distribution $(1-p^{xy}(t), p^{xy}(t))$. As the Kullback-Leibler divergence is a statistical measure of the similarity between distributions $I(t)$ captures the imbalance of the agent's choice at time t. Note that $I = 0$ means that both choices are equally likely; $I = \infty$ when p equals 0 or 1, and that I is symmetric about 0.5. We define $\bar{I}^{xy}$ to be the average choice imbalance,

$$\bar{I}^{xy} := \int_0^\infty I^{xy}(t)\, dF^{xy}(t),$$

and we define $\bar{T}^{xy}$ to be the average decision time,

$$\bar{T}^{xy} := \int_0^\infty t\, dF^{xy}(t),$$

and define $\bar{p}^{xy}$ to be the average choice probability,

$$\bar{p}^{xy} := \int_0^\infty p^{xy}(t)\, dF^{xy}(t),$$

and assume that all of these integrals exist. Finally, we relabel objects as needed so that the first object is chosen weakly more often, i.e. $\bar{p}^{xy} \geq 0.5$ for all x, y.

2. DDM representation

The drift diffusion model (DDM) is commonly used to explain the stochastic choice data in neuroscience and psychology. The two main ingredients of a DDM are the stimulus process Z_t and a time-dependent stopping boundary $b(t)$. In the DDM representation, the stimulus process Z_t is a Brownian motion with drift δ and volatility α:

$$Z_t = \delta\, t + \alpha\, B_t, \qquad [1]$$

where B_t is a standard Brownian motion, so in particular $Z_0 = 0$. Define the hitting time τ

$$\tau = \inf\{t \geq 0 : |Z_t| \geq b(t)\}, \qquad [2]$$

i.e., the first time the absolute value of the process Z_t hits the boundary b. Let $F^*(t; \delta, b, \alpha) := \mathbb{P}[\tau \leq t]$ be the distribution of the stopping time τ. Likewise, let $p^*(t; \delta, b, \alpha)$ be the conditional choice probability induced by Eq. (1) and Eq. (2) and a decision rule that chooses x if $Z_\tau = b(\tau)$ and y if $Z_\tau = -b(\tau)$.

Our goal in this paper is to determine which data is consistent with a DDM representation, and when it is, when the representation is unique. When the drift $\delta = 0$, each alternative will be chosen half of the time regardless of the shape of the boundary, so we will exclude this case going forward.

The original formulation of the DDM was for "perception tasks" where the drift δ is either $+1$ or -1 depending on which decision is correct; more generally there can be a distinct drift δ^{xy} for each pair x, y. In consumption-choice problems (otherwise known as value-based problems, see, e.g., (16)) it is natural to assume that the net drift δ^{xy} is the difference between two signals, an x-signal with drift $u(x)$ equal to the utility of x and a y-signal with drift $u(y)$ equal to the utility of y, so that $\delta^{xy} = u(x) - u(y)$. This imposes some consistency conditions that we discuss below.

Definition 1 (DDM Representation)**.** Stochastic choice data $(p^{xy}, F^{xy})_{x,y\in X}$ has a DDM representation if there exists a utility function $u : X \to \mathbb{R}$, a volatility parameter $\alpha > 0$ as well as a boundary $b : \mathbb{R}_+ \to \mathbb{R}_+$ such that for all $x, y \in X$ and $t \in \mathbb{R}$

$$p^{xy}(t) = p^*\Big(t, u(x) - u(y), b, \alpha\Big)$$
$$\text{and } F^{xy}(t) = F^*\Big(t, u(x) - u(y), b, \alpha\Big).$$

Note that this definition requires that the data from all of the menus $\{x, y\}$ is generated with the *same* boundary function b. This corresponds to cases where the agent treats each decision problem as a random draw from a fixed environment.[†] We are interested in characterizing which stochastic choice functions admits a DDM representation. The following result follows immediately from rescaling δ and b.

Lemma 1. *If a stochastic choice function exhibits a DDM representation for some α, then it also exhibits a DDM representation for $\alpha = 1$.*

We will thus without loss of generality only consider the DDM model where we normalized $\alpha = 1$. We write $p^*(t, \delta, b)$ and $F^*(t, \delta, b)$ as short-hands for $p^*(t, \delta, b, 1)$ and $F^*(t, \delta, b, 1)$.

3. Characterization

Given a stochastic choice function (p^{xy}, F^{xy}), define the *revealed drift*

$$\widetilde{\delta}^{xy} := \sqrt{\frac{\bar{I}^{xy}}{2\bar{T}^{xy}}}. \qquad [3]$$

When the revealed drift is is non zero, we define the *revealed boundary* as

$$\widetilde{b}^{xy}(t) := \frac{\ln p^{xy}(t) - \ln(1 - p^{xy}(t))}{2\widetilde{\delta}^{xy}}. \qquad [4]$$

The revealed drift is high for a pair x, y whenever the agent either makes very imbalanced choices or decides quickly, and the revealed drift is low for choices that are slow and close to

[†] In an optimal stopping model, the shape of the boundary is determined by the agent's prior over these draws.

DRAFT

50-50. Over time the boundary at time t follows the log-odds ratio of the agent's choice at time t which is zero whenever the agent's choice is balanced and and increases in the imbalance of the agent's choice. The revealed boundary is smaller for pairs with a larger revealed drift. In the knife-edge case when the revealed drift is 0 the revealed boundary is not defined and our results do not apply.

Theorem 1 below says that if the true data generating process is a DDM, then the revealed drift and boundary will exactly match the true parameters. Moreover, Theorem 1 allows us to test whether the true data generating process is indeed a DDM.

A. Characterization for a fixed pair. Our first result characterizes the DDM for a fixed pair $x, y \in X$.

Theorem 1. *For a fixed pair x, y with $\tilde{\delta}^{xy} \neq 0$ the stochastic choice function (p^{xy}, F^{xy}) admits a DDM representation if and only if for all $t \geq 0$*

$$F^{xy}(t) = F^*(t; \tilde{\delta}^{xy}, \tilde{b}^{xy}).$$

Moreover, if such a representation exists, it is unique (up to the choice of α) and given by $\tilde{\delta}^{xy}, \tilde{b}^{xy}$.

Thus, the stochastic choice function (p^{xy}, F^{xy}) is consistent with DDM whenever the observed distribution of stopping times F^{xy} equals to the distribution of hitting times generated by the revealed drift $\tilde{\delta}^{xy}$ and revealed boundary $\tilde{b}^{xy}$. Theorem 1 shows that the revealed drift and boundary are the unique candidate for a DDM representation. It thus allows us to identify the parameters of the DDM model directly from choice data. This permits the model to be calibrated to the data without computing the likelihood function, which requires computationally costly Monte-Carlo simulations. More substantially, as Theorem 1 connects the primitives of the model directly to data it allows us to better understand their economic meaning. The drift in the DDM model is a measure of how imbalanced and quick the agent's choices are and the shape of the boundary follows the imbalance of the agent's choices over time. We hope that this interpretation makes the empirical content of the parameters of DDM model more transparent and the model thus more useful.

Note that this theorem shows that the distribution of stopping times contains additional information that is not captured by the mean. For example, choice data where $p^{xy}(t)$ and $\bar{T}^{xy}$ are any 2 given constants is only consistent with one possible distribution of stopping times F^{xy} However a test based only on the mean choice probability and mean stopping time will accept any model that matches those two numbers, and in particular regardless of F^{xy} the data is consistent with a constant stopping boundary. (See (23)).

B. Characterization for menus of pairs. Our next result extends the characterization to all pairs $x, y \in X$.

Theorem 2. *The stochastic choice function $(\{p^{xy}\}, \{F^{xy}\})_{x,y \in X})$ has a DDM representation iff*

(i) $F^{xy}(t) = F^*(t; \tilde{\delta}^{xy}, \tilde{b}^{xy})$ for all $t \geq 0$,

(ii) $\widetilde{b}^{xy}(t) = \widetilde{b}^{xz}(t)$ for all $x, y, z \in X$ and all $t \geq 0$.

(iii) $\widetilde{\delta}^{xy} + \widetilde{\delta}^{yz} = \widetilde{\delta}^{xz}$ for all $x, y, z \in X$,

Thus, in addition to satisfying the condition from Theorem 1 pairwise, we have two additional consistency conditions imposed across pairs. Condition (ii) follows from our assumption that the agent uses the same stopping boundary in every menu. Condition (iii) comes from the assumption that the drift in a given menu depends on the difference of utilities, that is $\delta^{xy} = u(x) - u(y)$.[‡]

4. An Econometric Test for a Fixed Pair of Alternatives

The idea for the test is based on Theorem 1, which requires that the observed distribution of stopping times matches the distribution induced by the revealed boundary $\tilde{b}$ and drift $\tilde{d}$. We first describe a nonparametric estimator of $\tilde{b}$ and $\tilde{\delta}$ based on a finite data set. Next, we show how to test the distribution matching condition. This test could be extended to multiple-alternatives settings along the lines of Theorem 2, but we do not do so here.

A. Estimation of drift and boundary. Suppose that we have a fixed pair $x, y \in X$. Define

$$\gamma_\tau := \begin{cases} 1, & \text{when choice } x \text{ is made,} \\ 0, & \text{when choice } y \text{ is made.} \end{cases}$$

Each data point consists of the time τ_i at which the choice is made and the choice γ_i made at time τ_i.

Assumption 1. The data $(\tau_1, \gamma_1), \ldots, (\tau_n, \gamma_n)$ are i.i.d.

The unknown features of the DDM model are the drift δ and the boundary $b(t)$. We use estimators based on equations Eq. (3) and Eq. (4) that identify the revealed drift and boundary. Both of them depend on the choice probability, so we first give an estimator of that. Here $p^{xy}(t) := \Pr(\gamma_i = 1 | \tau_i = t)$ is the probability of choice x conditional on the choice being made at t.

The nonparametric estimator we construct is a spline regression: that is, a least squares regression of γ_i on approximating functions of τ_i. For simplicity, we use a linear probability estimator of $p^{xy}(t)$.[§]

We first transform τ_i to the unit interval[¶] by choosing a random variable whose density is positive on all positive reals, and setting

$$G_i = G(\tau_i).$$

Because G_i lies in the unit interval we can use standard series estimation to estimate $p^{xy}(t)$. We consider regression spline estimation of $p^{xy}(t)$. For this purpose let

$$q^K(G) = (q_{1K}(G), \ldots, q_{KK}(G))'$$

be a B-spline vector, say for evenly spaced knots on $(0, 1)$. Let $\hat{\beta}$ be OLS coefficients from regressing γ_i on $q_i^K = q^K(G_i)$. The choice probability estimator we consider is

$$\hat{p}(t) := q^K(G(t))' \hat{\beta}, \ \hat{\beta} := \left[\sum_{i=1}^n q_i^K q_i^{K\prime} \right]^{-1} \sum_{i=1}^n q_i^K \gamma_i.$$

[‡] The proof of Theorem 2 follows from Theorem 1 and the Sincov functional equation, see, e.g., (35).

[§] We reserve consideration of other estimators of the choice probability to future work, including logit or probit with a series approximation inside the logit or probit CDF.

[¶] In DDM models where $b(t)$ does not reach zero, there is no uniform bound on realized decision times τ_i. Because τ_i is the conditioning variable (i.e. regressor) in the choice probability, it is important to allow for an unbounded regressor.

DRAFT

We give conditions for this estimator to be consistent and have other important large sample properties in Assumptions 2 and 3 to follow.

A choice of $G(\tau)$ will be important in practice for constructing $\hat{p}(t)$. One possible choice of $G(\tau)$ is the CDF for the first passage time of a Brownian motion with drift crossing a single boundary, with mean and variance matched to that of the τ_i observations. Figure 1 gives a histogram for $G(\tau_i)$ from 100,000 simulations of τ_i for drift $\delta = .5$ and a constant boundary of -1 and 1.

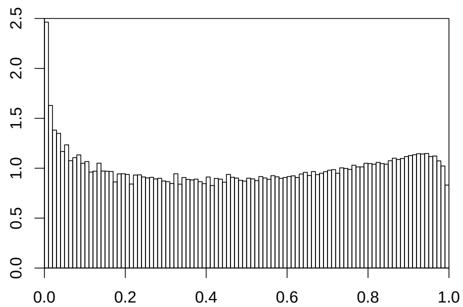


Fig. 1. Density of transformed FPT (τ)

This histogram is bounded well away from zero and infinity over most of its range so that we expect asymptotic approximations to work well that are based on a bounded regressor density that is bounded away from zero. The histogram does suggest that the density may grow as $G(\tau)$ approaches zero and shrink and $G(\tau)$ approaches 1. We expect this tail behavior to have little effect on finite sample performance of the estimator. It could also be controlled for if the boundary is constant as τ approaches zero and infinity and that restriction is imposed on the boundary estimator.

We can estimate the drift δ by plugging in $\hat{p}(t)$ for $p^{xy}(t)$ in formula Eq. (3) and replacing expectations with sample averages. Let

$$\hat{I}(t) := \hat{p}\,(t) \ln\left[\frac{\hat{p}\,(t)}{1-\hat{p}\,(t)}\right] + [1-\hat{p}\,(t)] \ln\left[\frac{1-\hat{p}\,(t)}{\hat{p}\,(t)}\right],$$

$$\bar{I} := \frac{1}{n}\sum_{i=1}^{n} \hat{I}\,(\tau_i)\,,\ \bar{\tau} := \frac{1}{n}\sum_{i=1}^{n} \tau_i.$$

The estimator of δ is then

$$\hat{\delta} := \sqrt{\frac{\bar{I}}{2\bar{\tau}}}.$$

The estimator of the boundary $b\,(t)$ is obtained by plugging in $\hat{\delta}$ and $\hat{p}(t)$ in the expression of equation Eq. (4), giving

$$\hat{b}(t) := \frac{1}{2\hat{\delta}} \ln\left[\frac{\hat{p}\,(t)}{1-\hat{p}\,(t)}\right].$$

B. Testing. We now have to test whether the observed distribution of stopping times matches the one induced by the revealed drift and boundary. We do this by comparing sample moments of functions of the decision time with estimators of the moments predicted by the model. To describe such a test let $m_J(\tau) = (m_{1J}(\tau), ..., m_{JJ}(\tau))'$ be a $J \times 1$ vector of functions of τ. Examples include indicator functions for intervals and B-splines in $G(\tau)$. The sample average vector will be $\bar{m} = \sum_{i=1}^{n} m_J(\tau_i)/n$.[‖] We use simulation to obtain model prediction. To describe the simulated predictions, let $\{B_t^1, ..., B_t^S\}$ be S independent copies of Brownian motion and ,

$$\hat{\tau}_s = \inf\{t \geq 0 : \left|\hat{\delta}t + B_t^s\right| \geq \hat{b}(t)\}.$$

A moment vector predicted by the model is $\hat{m}_S = \sum_{s=1}^{S} m_J(\hat{\tau}_s)/S$. A test of the model can be based on comparing $\bar{m}$ and $\hat{m}$. Let $\hat{V}$ be a consistent estimator of the asymptotic variance of $\sqrt{n}(\bar{m} - \hat{m}_S)$ when the model is correctly specified, as we will describe below. A test statistic can be formed as

$$\hat{A} := n(\bar{m} - \hat{m}_S)'\hat{V}^{-1}(\bar{m} - \hat{m}_S).$$

The model would be rejected if $\hat{A}$ exceeds the critical value of a $\chi^2(J)$ distribution. If J is allowed to grow with n and the $m_J(\tau)$ is allowed to grow in dimension and richness as n grows then this approach will test all the restrictions implied by DDM as n grows.

It is straightforward to construct $\hat{V}$ using the nonparametric bootstrap. Each bootstrap replication would start with a random sample $Z_n^j = (\tau_1^j, y_1^j), ..., (\tau_n^j, y_n^j)$ that consists of i.i.d. observations (τ_i^j, y_i^j), $(i = 1, ..., n)$, that are drawn at random with replacement from the data observations. Here j is a positive integer that identifies the bootstrap replication with $(j = 1, ..., B)$. For the j^{th} replication G_i^j, $\hat{p}^j(t)$, $\hat{\delta}^j$, $\hat{b}^j(t)$, and $\bar{m}^j$ are computed exactly as describe above with Z_n^j replacing the actual data. Using drift coefficient $\hat{\delta}^j$ and the estimated boundary $\hat{b}^j(t)$ from the j^{th} bootstrap replication S simulations $\hat{\tau}_s^b$, $(s = 1, ..., S)$, would be constructed as described above, resimulating for each bootstrap replication, and $\hat{m}_S^j = \sum_{s=1}^{S} m_J(\hat{\tau}_s^j)/S$ would be obtained. For $\hat{\Delta}^j = \bar{m}^j - \hat{m}_S^j$ and $\bar{\Delta}^j = \sum_{j=1}^{B} \hat{\Delta}^j/B$ a bootstrap variance estimator $\hat{V}_B$ would be

$$\hat{V}_B = \frac{n}{B}\sum_{j=1}^{B}(\hat{\Delta}^j - \bar{\Delta}^j)(\hat{\Delta}^j - \bar{\Delta}^j)'.$$

In Section 3 of SI we also give another estimator $\hat{V}_n$ based on asymptotic theory. In Monte Carlo simulations to follow we find that the bootstrap estimator $\hat{V}_B$ leads to rejection frequencies that are closer to their nominal values, so we recommend the bootstrap estimator variance estimator $\hat{V} = \hat{V}_B$ for constructing $\hat{A}$ in practice.

In formulating conditions for the asymptotic distribution of this test we will let $m_{jJ}(\tau)$, $(j = 1, ..., J)$ be indicator functions for disjoint intervals. Let $\tau_{jJ} = G^{-1}(j/(J+1))$, $(j = 0, ..., J)$, $\tau_{J+1,J} = \infty$. Consider

$$m_{jJ}(t) = \sqrt{J+1} \cdot \mathbb{1}(\tau_{j,J} \leq t < \tau_{j+1,J}),\ (j = 1, ..., J).$$

The test based on these functions will be based on comparing empirical probabilities of intervals with those predicted by the model. The normalization of multiplying by $\sqrt{J+1}$ is convenient in making the second moment of these functions of the same magnitude for different values of J. Note that we have left out the indicator for the interval $(0, 1/(J+1))$. We

[‖] The Kolmogorov–Smirnoff test uses indicator functions but instead of the the average of m it takes the supremum. The Cramer–von Mises test takes the sum of squares. We look at the average of m because the target cdf we are comparing with is not fixed, but involves estimates of the boundary and drift, see (36).

DRAFT

have done this to account for the fact that the estimator of the drift parameter uses some information about τ_i, so that we are not able to test all of the implications of the DDM for the distribution of τ_i; we can only test overidentifying restrictions. Also in the Monte Carlo results we left out the indicator for the interval $(J/(J+1), 1)$. Leaving out this other endpoint makes actual rejection rates closer to the nominal ones in our Monte Carlo study.

We derive results under the following conditions:

Assumption 2. The pdf of $G(\tau_i)$ is bounded and bounded away from zero.

This assumption is equivalent to the ratio of the pdf of τ_i to $dG(t)/dt$ being bounded and bounded away from zero. It is straightforward to weaken this condition to allow it to only requiring it on a compact, connected interval that is a subset of $(0, 1)$, if we assume the $b(t)$ is constant on known intervals near 0 and where τ is large.

We also make a smoothness assumption on the boundary function.

Assumption 3. $b(G^{-1}(g))$ is bounded and $s \geq 1$ times differentiable with bounded derivatives on $g \in [0, 1]$ and the $q_{kK}(G)$, $k = 1, ..., K$ are b-splines of order $s - 1$.

This condition requires that the derivatives of $b(t)$ go to zero in the tails of the distribution of τ_i as fast as the pdf of $G(t)$ does. We also require that the drift parameter be nonzero.

Assumption 4. $\delta \neq 0$.

We need to add other conditions about the smoothness of CDF of τ_i as a function of the drift δ and the boundary and about rates of growth of J and K. They involve much notation, so we state them in Assumption 5 in Appendix A.

We can now state the following result on the limiting distribution of $\hat{A}$ for the asymptotic variance estimator $\hat{V} = \hat{V}_n$ described in SI, Section 3.

Theorem 3. *Suppose that Assumptions 1, 2, 3, 4 and Assumption 5 in Appendix A are satisfied. Then for the* $1 - \alpha$ *quantile* $c(\alpha, J)$ *of a chi-square distribution with* J *degrees of freedom*

$$\Pr\left(\hat{A} \geq c(\alpha, J)\right) \longrightarrow \alpha.$$

Appendix

A. Smoothness Conditions for the CDF of τ_i.

To obtain the limiting distribution of the test statistic we make use of smoothness conditions for the CDF of τ_i as $F(t|\delta, b)$ as a function of the drift δ and boundary $b(\cdot)$. The three key primitive regularity conditions that will be useful involve a Frechet derivative $D(\tilde{\delta} - \delta, \tilde{b} - b; \delta, b, t)$ of $F(t|\delta, b)$ with respect to δ and b. We collect these conditions in the following assumption. Let $\varepsilon_{pn} = \sqrt{n^{-1}K\ln(K)} + K^{-s}$.

Assumption 5. For $\left|\tilde{b}\right| = \sup_t \left|\tilde{b}(t)\right|$ there is $C > 0$ not depending on δ, b, t such that

a)

$$|F(t|\tilde{\delta}, \tilde{b}) - F(t|\delta, b) + D(\tilde{\delta} - \delta, \tilde{b} - b; \delta, b, t)| \leq C(|\tilde{\delta} - \delta|^2 + |\tilde{b} - b|^2);$$

b) for each t there is a constant D^{δ}_{0t} and function $\alpha_{0t}(t)$ such that $|\alpha_{0t}(\tau_i)| \leq C$, $\left|D^{\delta}_{0t}\right| \leq C$, $|d^s\alpha_{0t}(t)/dt^s| \leq C$ for s equal to the order of the spline plus 1, and

$$D(\tilde{\delta} - \delta, \tilde{b} - b; \delta, b, t) = D^{\delta}_{0t}(\tilde{\delta} - \delta) + E[\alpha_{0t}(\tau_i)\{\tilde{b}(\tau_i) - b(\tau_i)\}];$$

c)

$$|D(\delta, b; \tilde{\delta}, \tilde{b}, t) - D(\delta, b; \delta_0, b_0, t)| \leq C(|\delta| + |b|)(|\tilde{\delta} - \delta_0| + |\tilde{b} - b_0|).$$

d) there is $C > 0$ such that for $\psi_{i\delta x} = I(\tau_i) - E[I(\tau_i)] - \delta^2\{\tau_i - E[\tau_i]\}$ and all J,

$$(J+1)E[1(\tau_i < 1/(J+1))\psi^2_{i\delta x}] \geq C.$$

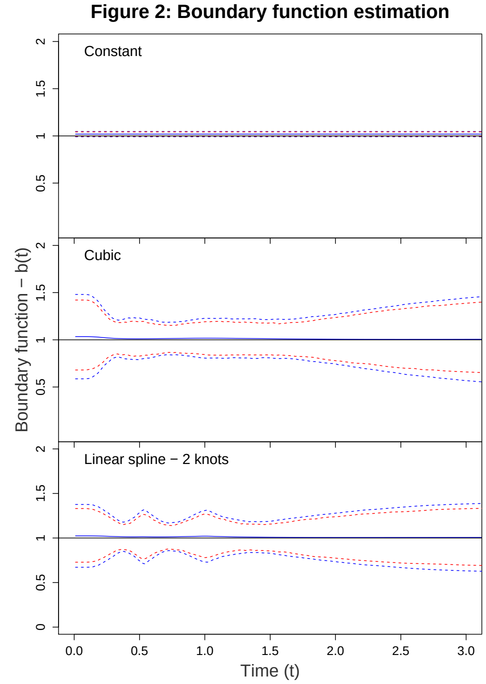
e) Each of the following converge to zero: $\sqrt{n}J\varepsilon^2_{pn}$, nJ^3/S, $J^{7/2}K/(\sqrt{S}\Delta)$, $J^{7/2}K\Delta$, $J^{7/2}K^{3/2}\varepsilon_{pn}$, $J^{5/2}K^{-s_\alpha}$

Part a) is Frechet differentiability of the CDF of τ_i in the drift and boundary, b) is implied by mean square continuity of the derivative and the Riesz representation Theorem, and c) is continuity of the functional derivative D in δ and b. The test statistic will continue to be asymptotically chi-squared for a stronger norm for b under corresponding stronger rate conditions for J, K, and Δ.

B. Monte Carlo Examples

To consider how the estimators and test might work in practice we consider some Monte Carlo examples. For all the Monte Carlo results we assume that the true boundary is constant at -1 and 1. In the supplemental information we find that $\hat{\delta}$ is a precise estimator of the drift parameter for sample size $n = 1000$.

Figure 2: Boundary function estimation



DRAFT

Figure 2 plots the mean and pointwise (inner) and uniform (outer) .025 and .975 quantile bands for the estimated boundary function. We set $\delta_0 = .5$ throughout and consider three different boundary estimators: a constant boundary estimator, a cubic approximation to $p(t)$ where $q^K(G) = (1, G, G^2, G^3)'$, and a linear spline approximation with two knots at .33 and .66 respectively. The quantile bands for the boundary estimator with cubic and spline approximations seem large but are consistent with delta method calculations, as discussed in the Supplemental Information.

Table 1 reports Monte Carlo rejection frequencies for the test statistic with bootstrap variance estimator. We use a linear regression spline estimator for $\hat{p}(t)$ with no knots, 1 knot at .5, or two knots at .33 and .66. We consider the test statistic with bootstrap variance estimator $\hat{V}_B$ for $B = 250$ bootstrap replications. We set $J = 5$ with only the middle three intervals included and $J = 8$ where we only keep the middle six intervals in the test statistic. Rejection frequencies are given when critical values are chosen to give asymptotic levels 1, 5, 10, and 20 percent.

Table 1: Reject Rates for Test Statistic

	Model	20%	10%	5%	1%
$J = 5$	Constant	.172	.078	.048	.014
	Linear	.216	.104	.042	.012
	1 Knot	.194	.108	.070	.018
	2 Knots	.224	.142	.080	.030
$J = 8$	Constant	.192	.106	.054	.008
	Linear	.214	.116	.066	.020
	1 Knot	.212	.128	.076	.026
	2 Knots	.248	.158	.112	.060

We find some tendency of the test statistic to reject too often when the number of intervals J is larger and the number of series terms K is larger. When we used power series up to cubic order or the analytic estimator of the asymptotic variance., the test statistic tended to overreject even more, especially for the analytic variance estimator.

The tendency to overreject for larger J and/or more flexible boundary specifications indicates some difficulty in reliably testing the many implications of the DDM model from 1000 observations. This difficulty is not surprising given the high variance of the boundary estimator, which could lead to the local approximation used in the asymptotic theory not working well. Imposing restrictions on the boundary could help with this problem as it does in Table 2, where more parsimonious specifications have less tendency to overreject. One potentially useful nonparametric restriction is monotonicity of the boundary. One could impose such a restriction and carry out inference using the approach of (37). This avenue seems potentially fruitful to explore but is beyond the scope of this paper.

ACKNOWLEDGMENTS. This research was supported by NSF grants SES-1643517, SES-1757140, and SES-1255062. David Hughes provided excellent research assistance.

1. Roger Ratcliff and Gail McKoon. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural computation*, 20(4):873–922, 2008.
2. Michael N Shadlen and Roozbeh Kiani. Decision making as a window on cognition. *Neuron*, 80(3):791–806, 2013.
3. Ernst Fehr and Antonio Rangel. Neuroeconomic foundations of economic choice—recent advances. *The Journal of Economic Perspectives*, 25(4):3–30, 2011.
4. Robert M Roe, Jermone R Busemeyer, and James T Townsend. Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological review*, 108 (2):370, 2001.
5. John A Clithero and Antonio Rangel. Combining response times and choice data using a neuroeconomic model of the decision process improves out-of-sample predictions. *mimeo*, 2013.
6. Ian Krajbich, Carrie Armel, and Antonio Rangel. Visual fixations and the computation and comparison of value in simple choice. *Nature neuroscience*, 13(10):1292–1298, 2010.
7. Ian Krajbich and Antonio Rangel. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, 108(33):13852–13857, 2011.
8. Ian Krajbich, Dingchao Lu, Colin Camerer, and Antonio Rangel. The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in psychology*, 3:193, 2012.
9. Milica Milosavljevic, Jonathan Malmaud, Alexander Huth, Christof Koch, and Antonio Rangel. The drift diffusion model can account for value-based choice response times under high and low time pressure. *Judgement & Decision Making*, 5:437–449, 2010.
10. I Krajbich, B Bartling, T Hare, and E Fehr. Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications*, 6(7455):700, 2015.
11. Elena Reutskaja, Rosemarie Nagel, Colin F Camerer, and Antonio Rangel. Search dynamics in consumer choice under time pressure: An eye-tracking study. *The American Economic Review*, 101(2):900–926, 2011.
12. Abraham Wald. *Sequential analysis*. John Wiley & Sons, 1947.
13. Mervyn Stone. Models for choice-reaction time. *Psychometrika*, 25(3):251–260, 1960.
14. Ward Edwards. Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, 2(2): 312–329, 1965.
15. Roger Ratcliff. A theory of memory retrieval. *Psychological review*, 85(2):59, 1978.
16. Milica Milosavljevic, Jonathan Malmaud, Alexander Huth, Christof Koch, and Antonio Rangel. The drift diffusion model can account for value-based choice response times under high and low time pressure. *Judgement & Decision Making*, 5:437–449, 2010.
17. Jan Drugowitsch, Rubén Moreno-Bote, Anne K Churchland, Michael N Shadlen, and Alexandre Pouget. The cost of accumulating evidence in perceptual decision making. *The Journal of Neuroscience*, 32(11):3612–3628, 2012.
18. Drew Fudenberg, Philipp Strack, and Tomasz Strzalecki. Speed, accuracy, and the optimal timing of choices. *American Economic Review*, 108(12):3651–84, 2018.
19. Satohiro Tajima, Jan Drugowitsch, Nisheet Patel, and Alexandre Pouget. Optimal policy for multi-alternative decisions. *Nature neuroscience*, 22(9):1503–1511, 2019.
20. Guy E Hawkins, Birte U Forstmann, Eric-Jan Wagenmakers, Roger Ratcliff, and Scott D Brown. Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience*, 35(6):2476–2484, 2015.
21. Khai Chiong, Matthew Shum, Ryan Webb, and Richard Chen. Split-second decision-making in the field: Response times in mobile advertising. *Available at SSRN*, 2018.
22. Roger Ratcliff. A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic bulletin & review*, 9(2):278–291, 2002.
23. Carlo Baldassi, Simone Cerreia-Vioglio, Fabio Maccheroni, and Massimo Marinacci. An axiomatization of the drift diffusion model and its extension to multi-alternative choice. *mimeo*, 2018.
24. Jerome R Busemeyer and James T Townsend. Fundamental derivations from decision field theory. *Mathematical Social Sciences*, 23(3):255–282, 1992.
25. Jerome R Busemeyer and James T Townsend. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, 100(3):432, 1993.
26. Jerome R Busemeyer and Joseph G Johnson. Computational models of decision making. *Blackwell handbook of judgment and decision making*, pages 133–154, 2004.
27. Carlos Alós-Ferrer, Ernst Fehr, and Nick Netzer. Time will tell: recovering preferences when choices are noisy. *University of Zurich, Department of Economics, Working Paper*, (306), 2018.
28. Federico Echenique and Kota Saito. Response time and utility. *Journal of Economic Behavior & Organization*, 139:49–59, 2017.
29. Benjamin Hebert and Michael Woodford. Rational inattention when decisions take time. *mimeo*, 2019.
30. Michael Woodford. An optimizing neuroeconomic model of discrete choice. *Columbia University working paper*, 2014.
31. Yeon-Koo Che and Konrad Mierendorff. Optimal dynamic allocation of attention. *American Economic Review*, 109(8):2993–3029, 2019.
32. A. Liang, X. Mu, and V. Syrgkanis. Dynamically aggregating diverse information. *mimeo*, 2012.
33. Annie Liang and Xiaosheng Mu. Complementary information and learning traps. *The Quarterly Journal of Economics*, 135(1):389–448, 2020.
34. Weijie Zhong. Optimal dynamic information acquisition. *mimeo*, 2019.
35. Joseph Aczél. *Lectures on functional equations and their applications*, volume 19. Academic press, 1966.
36. Whitney K Newey. The asymptotic variance of semiparametric estimators. *Econometrica*, pages 1349–1382, 1994.
37. Victor Chernozhukov, Whitney K Newey, and Andres Santos. Constrained conditional moment restriction models. *arXiv preprint arXiv:1509.06311*, 2015.

DRAFT

PNAS
www.pnas.org

Supplementary Information for

Testing the Drift-Diffusion Model

Drew Fudenberg, Whitney Newey, Philipp Strack, Tomasz Strzalecki

Corresponding Author Drew Fudenberg.
E-mail: drewf@mit.edu

This PDF file includes:

Supplementary text
SI References

Supporting Information Text

1. Monte Carlo Examples

Table 1 reports Monte Carlo mean and standard deviation for the estimator $\hat{\delta}$ of the drift parameter δ_0, where the true boundary is constant at -1 and 1, $\delta_0 \in \{.25, .5, 1.0\}$, $p^K(G) = (1, G, G^2, G^3)'$, 1000 Monte Carlo replications, and sample sizes $n = 100$ and $n = 1000$.

Table 1: Mean and Std Dev of $\hat{\delta}$

	$\delta = .25$	$\delta = .50$	$\delta = 1.0$
$n = 100$	.31	.55	1.07
	(.10)	(.10)	(.14)
$n = 1000$	.26	.50	1.02
	(.03)	(.03)	(.04)

Here we see that $\hat{\delta}$ is slightly upward biased for $n = 100$ but the bias disappears for $n = 1000$. The drift parameter is quite precisely estimated for $n = 1000$. We expect that this small variance results from averaging over observed τ_i values. The delta method implies that averaging lowers the sample variance to be equal to the estimator of the unconditional log odds, which is smaller than the variance of the log odds for the regression.

The large size of the quantile bands for the boundary estimator in Figure 2 are consistent with delta method calculations. When estimating a constant boundary the numerator and denominator of the boundary estimator $\hat{b}(t)$ are highly positively correlated leading to a precise boundary estimator. When the boundary is allowed to depend on t the variance of the slope is much larger than the variance of a constant when t is far from the middle of the distribution of τ. Furthermore, the variance of the slope is magnified by the fact that the boundary depends on a log odds ratio. Note that $\partial \ln(p/[1-p])/\partial p = 1/[p(1-p)] \geq 4$ so that the standard deviation of a log odds ratio is at least 4 times the standard deviation of an estimator of p. If $\delta = .5$, $n = 1000$, the true probability is constant, is estimated by a linear regression of γ_i on $G(\tau_i)$, and $G(\tau_i)$ is approximately uniformly distributed as in the simulation, then in the tails of the distribution of τ_i the boundary estimator has standard deviation of about $\sqrt{12/1000} \approx .11$, with a corresponding distance between upper and lower quantiles of about .44, consistent with Figure 2. Thus we see that both Monte Carlo results and delta method calculations deliver the conclusion that the boundary estimator is quite variable. We do not think this results from the choice of the least squares estimator of the probability, as other estimators would have similar variances. The high variance of the boundary seems to come instead from the fact it depends directly on a log odds ratio, which is quite variable.

2. Proofs from Section 4

A. Proof of Lemma 1. Dividing Eq. (1) in the paper by α and observing that $\inf\{t \geq 0 : |Z_t| \geq b(t)\} = \inf\{t \geq 0 : |\frac{Z_t}{\alpha}| \geq \frac{b(t)}{\alpha}\}$ yields that $p^*\left(\delta(x,y), b, \alpha\right) = p^*\left(\frac{1}{\alpha}\delta(x,y), \frac{b}{\alpha}, 1\right)$ and thus the result. **Q.E.D.**

B. Proof of Theorem 1. (1) We first show that these conditions are necessary for (p, F) to admit a DDM representation for a given pair $\{x, y\}$.

By equation (4) in (1) we have $\frac{p^{xy}(t)}{1-p^{xy}(t)} = \exp\left(2\delta(x,y)\, b(t)\right)$. Thus, we have that

$$b(t) = \frac{1}{2\delta(x,y)} \log\left(\frac{p^{xy}(t)}{1-p^{xy}(t)}\right). \qquad [1]$$

This proves that $b(t) = \tilde{b}(t)$ which is defined in Eq. (4) in the paper. By the definition of τ in Eq. (2) in the paper we have $Z_\tau = \operatorname{sgn}(Z_\tau) b(\tau)$. By Eq. (1) in the paper, $Z_\tau = \delta(x,y)\tau + B_\tau$. Combining these two equations and taking expectations, it follows from Doob's optional sampling theorem that

$$\delta(x,y)\, \mathbb{E}^{xy}[\tau] = \mathbb{E}^{xy}[\operatorname{sgn}(Z_\tau) b(\tau)] \qquad [2]$$

Plugging (1) into (2) yields

$$\delta(x,y)\, \mathbb{E}^{xy}[\tau] = \mathbb{E}^{xy}\left[\operatorname{sgn}(Z_\tau)\frac{1}{2\delta(x,y)} \log\left(\frac{p^{x}(\tau)}{1-p^{xy}(\tau)}\right)\right]$$

Dividing by $\mathbb{E}^{xy}[\tau]$ and multiplying by $2\delta(x,y)$ yields

$$
\begin{aligned}
2\delta(x,y)^2 &= \frac{\mathbb{E}^{xy}\left[\operatorname{sgn}(Z_\tau)\log\left(\frac{p^x(\tau)}{1-p^{xy}(\tau)}\right)\right]}{\mathbb{E}^{xy}[\tau]} = \frac{\mathbb{E}^{xy}\left[[\mathbf{1}_{Z_\tau>0}-\mathbf{1}_{Z_\tau<0}]\log\left(\frac{p^x(\tau)}{1-p^{xy}(\tau)}\right)\right]}{\mathbb{E}^{xy}[\tau]}\\
&= \frac{\mathbb{E}^{xy}\left[\int_0^\infty \mathbf{1}_{\tau=t}[\mathbf{1}_{Z_\tau>0}-\mathbf{1}_{Z_\tau<0}]\log\left(\frac{p^{xy}(t)}{1-p^{xy}(t)}\right)\mathrm{d}t\right]}{\int_0^\infty t\,\mathrm{d}F^{xy}(t)}\\
&= \frac{\mathbb{E}^{xy}\left[\int_0^\infty \mathbf{1}_{\tau=t}\mathbb{E}^{xy}\left[[\mathbf{1}_{Z_\tau>0}-\mathbf{1}_{Z_\tau<0}]\log\left(\frac{p^{xy}(t)}{1-p^{xy}(t)}\right)\mid \tau=t\right]\mathrm{d}t\right]}{\int_0^\infty t\,\mathrm{d}F^{xy}(t)}\\
&= \frac{\mathbb{E}^{xy}\left[\int_0^\infty \mathbf{1}_{\tau=t}[\mathbb{E}^{xy}[\mathbf{1}_{Z_\tau>0}\mid\tau=t]-\mathbb{E}^{xy}[\mathbf{1}_{Z_\tau<0}\mid\tau=t]]\log\left(\frac{p^{xy}(t)}{1-p^{xy}(t)}\right)\mathrm{d}t\right]}{\int_0^\infty t\,\mathrm{d}F^{xy}(t)}\\
&= \frac{\mathbb{E}^{xy}\left[\int_0^\infty \mathbf{1}_{\tau=t}[p^{xy}(t)-(1-p^{xy}(t))]\log\left(\frac{p^{xy}(t)}{1-p^{xy}(t)}\right)\mathrm{d}t\right]}{\int_0^\infty t\,\mathrm{d}F^{xy}(t)}\\
&= \frac{\int_0^\infty [p^{xy}(t)-(1-p^{xy})(t)]\log\left(\frac{p^{xy}(t)}{1-p^{xy}(t)}\right)\mathrm{d}F^{xy}(t)}{\int_0^\infty t\,\mathrm{d}F^{xy}(t)}\\
&= \frac{\int_0^\infty [2\,p^{xy}(t)-1]\log\left(\frac{p^{xy}(t)}{1-p^{xy}(t)}\right)\mathrm{d}F^{xy}(t)}{\int_0^\infty t\,\mathrm{d}F^{xy}(t)}\,.
\end{aligned}
$$

This proves that $\tilde{\delta}^{xy}=\delta^{xy}$ where $\tilde{\delta}$ is defined in Eq. (3) in the paper. Finally, we know that $\delta>0$ if and only if the probability with which the first object is chosen $\mathbb{P}^{xy}[Z_\tau>0]=\int_0^\infty p^{xy}(t)\mathrm{d}F^{xy}(t)$ is greater $\frac{1}{2}$ which yields the result.

To show sufficiency, consider the DDM model with parameters $(\tilde{\delta}^{xy},\tilde{b}^{xy})$ given by Eq. (3,4) in the paper. It follows from the assumption of the Theorem that F^{xy} equals the distribution over stopping times in the DDM model with boundary $\tilde{b}^{xy}$ and drift δ^{xy}. Finally, we will show that this DDM model also generates the correct conditional stopping probabilities p^{xy}. By equation (4) in (1), the conditional probability of stopping in the DDM model $\tilde{p}^{xy}$ satisfies

$$
\frac{\tilde{p}^{xy}(t)}{1-\tilde{p}^{xy}(t)}=\exp\left(2\tilde{\delta}(x,y)\,\tilde{b}^{xy}(t)\right)=\frac{p^{xy}(t)}{1-p^{xy}(t)}\,,
$$

which completes the proof as we have argued that each stochastic choice function is uniquely identified by the associated pair (p,F). **Q.E.D.**

3. Construction of $\hat{V}$

To construct $\hat{V}$ we use the fact that there are three asymptotically independent sources of variation in $\bar{m}-\hat{m}$. These sources are the variation in τ_i, the variation in $\hat{\beta}$, and the variation from simulation. The variation in τ_i affects both $\bar{m}$ and $\hat{\delta}$ and the variation in $\hat{\delta}$ has an effect through $\hat{m}$. Generally $\hat{m}$ will not be differentiable in $\hat{\delta}$ so we use a difference quotient to estimate the derivative of $\hat{m}$ with respect to δ. To describe how this source of variation can be estimated let

$$
\tau_s(\delta,\beta)=\inf\{t\geq 0:|\delta t+B_t^s|\geq\frac{1}{\delta}\ln\left[\frac{q^K(G(t))'\beta}{1-q^K(G(t))'\beta}\right]\},\ \hat{m}(\delta,\beta)=\frac{1}{S}\sum_{s=1}^S m_J(\tau_s(\delta,\beta)).
$$

denote one simulation $\tau_s(\delta,\beta)$ of τ_s when δ is the true drift and $q_K(G(t))'\beta$ the true $p(t)=p^{xy}(t)$ and $\hat{m}(\delta,\beta)$ denote the average over S simulations. Let

$$
\hat{M}_\delta=\frac{\hat{m}(\hat{\delta}+\Delta,\hat{\beta})-\hat{m}(\hat{\delta}-\Delta,\hat{\beta})}{2\Delta}
$$

be the difference quotient that serves as an estimator of the derivative of the the expectation of the model moments with respect to the drift. Then

$$
\hat{\psi}_{i1}=m_J(\tau_i)-\bar{m}-\hat{M}_\delta\frac{1}{2\hat{\delta}\bar{\tau}}[\hat{I}(\tau_i)-\bar{I}-\hat{\delta}^2\{\tau_i-\bar{\tau}\}]
$$

will estimate the influence of τ_i on the difference of moments coming from the effect of τ_i on the sample moments as well as on $\hat{\delta}$. An estimator of the variance of the moment differences due to variation in τ_i is then

$$
\hat{V}_1=\frac{1}{n}\sum_{i=1}^n\hat{\psi}_{i1}\hat{\psi}_{i1}'.
$$

To estimate the component of the variance due to $\hat{\beta}$ we use

$$\hat{M}_k = \frac{\hat{m}(\hat{\delta}, \hat{\beta} + e_k\Delta) - \hat{m}(\hat{\delta}, \hat{\beta} - e_k\Delta)}{2\Delta}, \quad \hat{M}_\beta = [\hat{M}_1, ..., \hat{M}_K].$$

to estimate the derivative of $E[m_J(\tau_s(\delta, \beta))]$ with respect to β at $\hat{\delta}$ and $\hat{\beta}$, where e_k is the k^{th} unit vector. Let $\hat{p}_i = \hat{p}(\tau_i)$ and $d(p) = d\ln[p/(1-p)]/dp = p^{-1}(1-p)^{-1}$. Accounting also for the effect of β on $\hat{\delta}$, an estimator of the Jacobian of $E[m_J(\tau_s(\delta, \beta))]$ with respect to β is

$$\hat{D}_\beta = \hat{M}_\delta \frac{1}{2\hat{\delta}\bar{\tau} n} \sum_{i=1}^{n} d(\hat{p}_i) q_i^{K\prime} + \hat{M}_\beta.$$

The variation in $\bar{m} - \hat{m}$ due to $\hat{\beta}$ can then be estimated by

$$\hat{V}_2 = \hat{D}_\beta \hat{\Sigma}^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} q_i^K q_i^{K\prime} (\gamma_i - \hat{p}_i)^2 \right] \hat{\Sigma}^{-1} \hat{D}_\beta', \; \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} q_i^K q_i^{K\prime}.$$

This is a delta method estimator of the asymptotic variance of $E[m_J(\tau_s(\delta, \beta))]$ due to the $\hat{\beta}$ in the nonparametric estimator $\hat{p}(t)$. As in (2), it is formed by treating $\hat{m}$ as depending on the vector of parameters $\hat{\beta}$ and applying the delta method as if K were fixed and not growing with the sample size.

The variation due to simulation is easy to estimate as $\hat{V}_3 = (n/S^2) \sum_{s=1}^{S} [m_J(\hat{\tau}_s) - \hat{m}] [m_J(\hat{\tau}_s) - \hat{m}]'$. In the theory we assume that the number of simulations is large enough so that we can replace this $\hat{V}_3$ by zero without affecting the results. Computing $\hat{V}_3$ in practice may still be a good idea check whether the number of simulations is large enough to make $\hat{V}_3$ negligible.

The estimators of the variance from independent sources of variation can then be combined into an asymptotic variance estimator for $\sqrt{n}[\bar{m} - \hat{m}_S]$ as

$$\hat{V} = \hat{V}_1 + \hat{V}_2 + \hat{V}_3.$$

We give conditions in Theorem 3 sufficient for the chi-squared approximation to the distribution of $\hat{A}$ to be correct for n, J, and S growing and Δ shrinking in specific ways.

4. Lemmas for Theorem 3

We will use two Lemmas on the asymptotic behavior of quadratic forms to prove the properties of the test statistic. For the first Lemma let h_i be a $J \times 1$ vector of random variables with $E[h_i] = 0$ and $h_1, \ldots, h_n$ i.i.d. Let

$$\Omega = E\left[h_i h_i'\right], \; \bar{h} = \frac{1}{n} \sum_i h_i.$$

Consider $\hat{h}$ that is approximately equal to $\bar{h}$ in the sense that $\hat{h} - \bar{h}$ is small. Also consider an estimator $\hat{\Omega}$ of Ω and let $\|A\| = \sqrt{tr\,(A'A)}$ be the L_2 norm on matrices.

Lemma 2: If i) $\lambda_{\min}(\Omega) \geq c > 0$, ii) $J^{-1/2}\sqrt{n} tr\,(\Omega)^{1/2} \left\| \hat{h} - \bar{h} \right\| \xrightarrow{p} 0$, iii) $J^{-1/2} tr\,(\Omega) \left\| \hat{\Omega} - \Omega \right\| \xrightarrow{p} 0$, and iv) $E\left[(h_i' h_i)^2\right]/nJ \longrightarrow 0$ then for the $1-\alpha$ quantile $c\,(\alpha, J)$ of a chi-square distribution with J degrees of freedom

$$\Pr\left(n \hat{h}' \hat{\Omega}^{-1} \hat{h} \geq c\,(\alpha, J)\right) \longrightarrow \alpha.$$

Proof: By i) we have $\lambda_{\min}(\Omega) \geq c$, so that $J^{-1/2} tr\,(\Omega)^{1/2} \geq c$. Then iii) implies $\left\| \hat{\Omega} - \Omega \right\| \xrightarrow{p} 0$ and hence w.p.a.1,

$$\lambda_{\min}\left(\hat{\Omega}\right) \geq c.$$

Since this event occurs w.p.a.1 we can assume it is true henceforth. Define

$$T_1 = n' \hat{h} \left(\hat{\Omega}^{-1} - \Omega^{-1}\right) \hat{h}, \; T_2 = n\left[\hat{h}' \Omega^{-1} \hat{h} - \bar{h}' \Omega^{-1} \bar{h}\right]$$

Note that $E[n \left\| \bar{h} \right\|^2] = nE[\bar{h}'\bar{h}] = tr(\Omega)$.Then by the Markov inequality we have

$$\sqrt{n} \left\| \bar{h} \right\| = O_p(tr(\Omega)^{1/2}).$$

Also by ii) $\sqrt{n} \left\| \hat{h} - \bar{h} \right\| \leq C J^{-1/2} tr(\Omega)^{1/2} \sqrt{n} \left\| \hat{h} - \bar{h} \right\| \xrightarrow{p} 0$. Then by the triangle inequality

$$\sqrt{n} \left\| \hat{h} \right\| \leq \sqrt{n} \left\| \bar{h} \right\| + \sqrt{n} \left\| \hat{h} - \bar{h} \right\| = O_p(tr(\Omega)^{1/2}).$$

It therefore follows that

$$|T_1| = \left|n\hat{h}'\hat{\Omega}^{-1}\left(\Omega-\hat{\Omega}\right)\Omega^{-1}\hat{h}\right| \leq \left\|\sqrt{n}\hat{h}'\hat{\Omega}^{-1}\right\|\left\|\hat{\Omega}-\Omega\right\|\left\|\sqrt{n}\hat{h}'\Omega^{-1}\right\| \leq cn\left\|\hat{h}\right\|^2\left\|\hat{\Omega}-\Omega\right\|$$
$$= O_p(tr(\Omega))\left\|\hat{\Omega}-\Omega\right\| = o_p(J^{1/2}).$$

Similarly we have

$$|T_2| = n\left|\left(\hat{h}-\bar{h}\right)'\Omega^{-1}\hat{h}+\bar{h}'\Omega^{-1}\left(\hat{h}-\bar{h}\right)\right| \leq n(\left\|\hat{h}-\bar{h}\right\|(\left\|\hat{h}\right\|+\left\|\bar{h}\right\|))$$
$$= O_p(tr(\Omega)^{1/2}\sqrt{n}\left\|\hat{h}-\bar{h}\right\|) = o_p(J^{1/2}).$$

It then follows by the triangle inequality that

$$n'\hat{h}\hat{\Omega}^{-1}\hat{h} - n\bar{h}\Omega^{-1}\bar{h} = T_1+T_2 = o_p(J^{1/2}).$$

In addition, by iv) and Lemma A.15 of (3),

$$\frac{n\bar{h}'\Omega^{-1}\bar{h}-J}{\sqrt{2J}} \xrightarrow{d} N(0,1).$$

Also, by standard results for the chi-squared distribution, as $J\rightarrow\infty$ we have $(c(\alpha,J)-J)/\sqrt{2J}$ converges to the $1-\alpha$ quantile of a $N(0,1)$. Hence

$$\Pr\left(n\bar{h}'\Omega^{-1}\bar{h}\geq c(\alpha,J)\right) = \Pr\left(\frac{n\bar{h}'\Omega^{-1}\bar{h}-J}{\sqrt{2J}} \geq \frac{c(\alpha,J)-J}{\sqrt{2J}}\right) \longrightarrow \alpha.$$

The conclusion then follows by the Slutzky Lemma. **Q.E.D.**

The next Lemma gives a rate of growth for the number of simulation draws to ensure that the limiting distribution of the test statistic based on $\hat{m}_S$ is the same as that based on $\hat{m} = \int m\left(\tau_s\left(\hat{\delta},\hat{b}\right)\right)dF(s)$.

Let h_s be simulated moments. Then we have

Lemma 3: If $\max_{1\leq j\leq J}\sup_{\tau>0}|m_{jJ}(\tau)| \leq C\sqrt{J}$ and $nJtr(\Omega)/S \longrightarrow 0$ then

$$J^{-1/2}\sqrt{n}tr(\Omega)^{1/2}\|\hat{m}_S-\hat{m}\| \xrightarrow{p} 0,$$

Proof: Let $Z = ((\gamma_1,\tau_1),\ldots,(\gamma_n,\tau_n))$ denote the data. Note that by definition, $E[\hat{m}_S|Z] = \hat{m}$. Then for any constant ℓ

$$\lim Prob(\|\hat{m}_S-\hat{m}\|>\ell) = E\left[\Pr\left(\|\hat{m}_S-\hat{m}\|>\ell \mid Z\right)\right].$$

By the Markov inequality

$$\Pr(\|\hat{m}_S-\hat{m}\|>\ell\mid Z) = \Pr\left(\|\hat{m}_S-\hat{m}\|^2>\ell^2\mid Z\right) \leq E\left[\sum_{j=1}^{J}(\hat{m}_{Sj}-\hat{m}_j)^2 \mid Z\right]/\ell^2$$
$$\leq \frac{1}{S}\sum_{j=1}^{J}E\left[\hat{m}_j\left(\tau_s\left(\hat{\delta},\hat{\beta}\right)\right)^2\mid Z\right]/\ell^2 \leq \frac{C^2J^2}{S\ell^2}.$$

By iterated expectations we then have

$$\Pr(\|\hat{m}_S-\hat{m}\|>\ell) \leq \frac{C^2J^2}{S\ell^2}.$$

Let $\ell = J^{1/2}tr(\Omega)^{-1/2}n^{-1/2}\varepsilon$. Then

$$\Pr\left(J^{-1/2}tr(\Omega)^{1/2}\sqrt{n}\|\hat{m}_s-\hat{m}\|\geq\varepsilon\right) = \Pr(\|\hat{m}_s-\hat{m}\|\geq\ell) \leq C^2J^2\left[SJtr(\Omega)^{-1}n^{-1}\varepsilon^2\right]^{-1}$$
$$= \frac{J^2tr(\Omega)n}{SJ\varepsilon^2} = \frac{nJtr(\Omega)}{S}\frac{1}{\varepsilon^2} \longrightarrow 0.$$

Q.E.D.

We next give a uniform convergence rate for $\hat{p}(t)$. For notational simplicity we let $p(t) := p^{xy}(t)$.

Lemma 4: If Assumptions 2 and 3 are satisfied then

$$\sup_t|\hat{p}(t)-p(t)| = O_p\left(\sqrt{\frac{K\ln(K)}{n}}+K^{-s}\right).$$

Proof: Follows from Theorem 4.3 and Comments 4.5 and 4.6 of (4). **Q.E.D.**

We next give an asymptotic expansion for $\hat{\delta}$. Define

$$\begin{aligned} I(p) &= p\ln\left(\frac{p}{1-p}\right) + (1-p)\ln\left(\frac{1-p}{p}\right) = (1-2p)\ln\left(\frac{1-p}{p}\right), \\ \psi_i^\delta &= \frac{1}{2E[\tau_i]\delta}\left\{I(p_i) - I_0 + I_p(p_i)(\gamma_i - p_i) - \delta^2(\tau_i - E[\tau_i])\right\}. \end{aligned}$$

Lemma 5: If Assumptions 2 and 3 are satisfied and $\sqrt{n}\varepsilon_{pn}^2 \longrightarrow 0$ then

$$\hat{\delta} - \delta = \frac{1}{n}\sum_i \psi_i^\delta + O_p(\varepsilon_{pn}^2) = \frac{1}{n}\sum_i \psi_i^\delta + o_p(1/\sqrt{n}) = O_p(1/\sqrt{n}).$$

Proof: Equation (4) and Assumption 3 imply that $p(t)$ is bounded away from zero and one. It then follows from Lemma 4 that with probability approaching one (w.p.a.1) there is $\varepsilon > 0$ with $\varepsilon \le \hat{p}(t) \le 1-\varepsilon$. It is straightforward to check that $I(p)$ is twice continuously differentiable in $p \in (0,1)$ with first and second deriatives that are bounded when p is bounded away from zero and one. It then follows by an expansion and Lemma 4 that

$$I(\hat{p}_i) = I(p_i) + I_p(p_i)(\hat{p}_i - p_i) + \hat{R}_i, \left|\hat{R}_i\right| \le C|\hat{p}_i - p_i|^2.$$

Therefore we have

$$\hat{I} = \frac{1}{n}\sum_i I(\hat{p}_i) = \frac{1}{n}\sum_i [I(p_i) + I_p(p_i)(\hat{p}_i - p_i)] + \hat{R},\ \hat{R} = O_p(\varepsilon_{pn}^2).$$

Define

$$\begin{aligned} \Gamma &= (\gamma_1, ..., \gamma_n)',\ P = (p_1, ..., p_n)',\ Q = [q^K(G_1), ..., q^K(G_n)]',\ I_p = (I_p(p_1), ..., I_p(p_n)), \\ H &= I - Q(Q'Q)^- Q. \end{aligned}$$

Note that derivatives of $I_p(p)$ to any order are bounded on $[\varepsilon, 1-\varepsilon]$, so that by the fact that the approximation rate of a general s differentiable function by a b-spline of at least order $s-1$ is K^{-s} we have

$$\frac{1}{n}P'HP = O(K^{-2s}),\ \frac{1}{n}I_p'HI_p = O(K^{-2s}).$$

Note also that

$$\frac{1}{n}\sum_i I_p(p_i)(\hat{p}_i - p_i) - \frac{1}{n}\sum_i I_p(p_i)(\gamma_i - p_i) = -\frac{1}{n}I_p'H\Gamma$$

Furthermore,

$$E[-\frac{1}{n}I_p'H\Gamma|\tau_1, ..., \tau_n] = -\frac{1}{n}I_p'HP = O(K^{-2s}),\ Var(-\frac{1}{n}I_p'H\Gamma|\tau_1, ..., \tau_n) \le \frac{1}{n^2}I_p'HI_p = O(\frac{K^{-2s}}{n}).$$

Then by $2K^{-s}/\sqrt{n} \le 1/n + K^{-2s} \le \varepsilon_{pn}^2$ it follows that

$$\frac{1}{n}\sum_i I_p(p_i)(\hat{p}_i - p_i) - \frac{1}{n}\sum_i I_p(p_i)(\gamma_i - p_i) = O_p(\frac{K^{-s}}{\sqrt{n}} + K^{-2s}) = O_p(\varepsilon_{pn}^2).$$

Then by the triangle inequality

$$\hat{I} = \frac{1}{n}\sum_i I(\hat{p}_i) = \frac{1}{n}\sum_i [I(p_i) + I_p(p_i)(\gamma_i - p_i)] + O_p(\varepsilon_{pn}^2).$$

Note that for $\delta(I, \tau) = \sqrt{I/\tau}$,

$$\frac{\partial\delta(I,\tau)}{\partial I} = \frac{1}{2\delta(I,\tau)\tau},\ \frac{\partial\delta(I,\tau)}{\partial\tau} = -\frac{\delta(I,\tau)}{2\tau}.$$

The conclusion then follows by the usual delta method argument. **Q.E.D.**

Next for any $\alpha(\tau)$ define

$$\psi_i^\alpha = -\delta^{-1}\{E[\alpha(\tau_i)b(\tau_i)]\psi_i^\delta + \frac{\alpha(\tau_i)}{p(\tau_i)[1-p(\tau_i)]}(\gamma_i - p_i)\}.$$

The next result gives a rate of convergence for the boundary estimator $\hat{b}(t)$ and a uniform expansion for a mean square continuous linear functional of $\hat{b}(t)$

Lemma 6: If there is a constant C such that $\alpha(G^{-1}(g))$ is continuously differentiable of order s with $\left|d\alpha(G^{-1}(g))/dg\right| \leq C$ on $[0,1]$, then $\sup_t |\hat{b}(t) - b(t)| = O_p(\varepsilon_{pn})$ and

$$\int \alpha(\tau)\{\hat{b}(\tau) - b(\tau)\}F_0(d\tau) = \frac{1}{n}\sum_i \psi_i^\alpha + O_p(\varepsilon_{np}^2),$$

uniformly in α.

Proof: Note that for $b(\delta,p) = \delta^{-1}\ln(p/[1-p])$,

$$\frac{\partial b(\delta,p)}{\partial \delta} = \frac{-b(\delta,p)}{\delta},\ \frac{\partial b(\delta,p)}{\partial p} = \frac{1}{\delta p(1-p)}.$$

Then by Lemma 5, a delta method argument similar to that used in the proof of Lemma 5, and $\hat{\delta} = \delta + O_p(1/\sqrt{n})$ we have

$$\hat{b}\,(t) = b(t) - b(t)\frac{[\hat{\delta}-\delta]}{\delta} + \frac{1}{\delta p(t)[1-p(t)]}[\hat{p}(t) - p(t)] + \hat{R}(t),\ \ \sup_t \left|\hat{R}(t)\right| = O_p(\varepsilon_{pn}^2).$$

The first conclusion then follows by $b(t)$ bounded, which implies $p(t)$ is bounded away from zero and one, and by Lemma 5. To show the second conclusion note that for any bounded $a(t)$ it follows by the proof of Corollary 10 of (5) that

$$\int a(\tau)[\hat{p}(\tau) - p(\tau)]F_0(d\tau) = \frac{1}{n}\sum_i a(\tau_i)[\gamma_i - p_i] + O_p(\varepsilon_{pn}^2),$$

uniformly in $a(\tau)$ with uniformly bounded derivatives to order s. Let $a(\tau) = \alpha(\tau)/\{\delta p(t)[1-p(t)]\}$. By plugging in the above expansion for $\hat{b}(t)$ and using boundedness of $\alpha(\tau)$ we obtain

$$\begin{aligned}
&\int \alpha(\tau)\{\hat{b}(\tau) - b(\tau)\}F_0(d\tau) \\
=\ & -\delta^{-1}\{E[\alpha(\tau_i)b(\tau_i)](\hat{\delta}-\delta) + \int a(\tau)[\hat{p}(\tau) - p(\tau)]F_0(d\tau) + \int \alpha(\tau)\hat{R}(\tau)F_0(d\tau). \\
=\ & \frac{1}{n}\sum_i \psi_i^\alpha + O_p(\varepsilon_{np}^2) + \int \alpha(\tau)\hat{R}(\tau)F_0(d\tau) = \frac{1}{n}\sum_i \psi_i^\alpha + O_p(\varepsilon_{np}^2).\textbf{Q.E.D.}
\end{aligned}$$

5. Proof of Theorem 4

We first show that conditions i)-iv) of Lemma 2 are satisfied. Let

$$\begin{aligned}
h_{ji} &= m_{ji} - E[m_{ji}] + M_{\delta j}\psi_i^\tau + \alpha_{j0}(\tau_i)(\gamma_i - p_i), \\
\psi_i^\tau &= \frac{1}{2\delta E[\tau_i]}\{I(p_i) - I_0 - \delta^2(\tau_i - E[\tau_i])\}, \\
M_{\delta j} &= \sqrt{J}(D_{0\tau_{j+1}}^\delta - D_{0\tau_j}^\delta - \delta^{-1}E[\{\alpha_{0,\tau_{j+1}}(\tau_i) - \alpha_{0,\tau_j}(\tau_i)\}b(\tau_i)]) \\
\alpha_{j0}(\tau_i) &= M_{\delta j}\frac{1}{2E[\tau_i]\delta}I_p(p_i) + \frac{\sqrt{J}[\alpha_{0,\tau_{j+1}}(\tau_i) - \alpha_{0,\tau_j}(\tau_i)]}{\delta p_i[1-p_i]}.
\end{aligned}$$

Also let

$$\begin{aligned}
h_i &= (h_{i1},...,h_{iJ})' = m_i - E[m_i] + M_\delta\psi_i^\tau + \alpha_0(\tau_i)(\gamma_i - p_i), \\
M_\delta &= (M_{\delta 1},...,M_{\delta J})',\ \alpha_0(\tau) = (\alpha_{10}(\tau),...,\alpha_{J0}(\tau))', \\
\Omega &= E[h_i h_i'],\ V_1 = Var(m_i + M_\delta\psi_i^\tau),\ V_2 = E[\alpha_0(\tau_i)\alpha_0(\tau_i)'Var(\gamma_i|\tau_i)].
\end{aligned}$$

Note that $\Omega = V_1 + V_2$ by $E[\gamma_i|\tau_i] = p(\tau_i)$.

To show condition i) of Lemma 2 it suffices to show that $\lambda_{\min}(V_1) \geq C$, which we now proceed to show. Let

$$\tilde{m}_i = (\sqrt{J+1}\psi_i^\tau, m_i')'.$$

It follows in a straightforward way from Assumption 5 d) that

$$\lambda_{\min}(E[\tilde{m}_i\tilde{m}_i']) \geq C.$$

Also, for $B = [M_\delta, I]$ we have

$$V_1 = BE[\tilde{m}_i\tilde{m}_i']B'.$$

Therefore for any conformable vector λ with $\lambda'\lambda = 1$,

$$\lambda' V_1 \lambda = \frac{\lambda' B E[\tilde{m}_i \tilde{m}_i'] B' \lambda}{\lambda' B B' \lambda} \lambda' B B' \lambda \geq C \lambda' B B' \lambda \geq C \lambda_{\min}(BB') \geq C \lambda_{\min}(I) = C.$$

We next show that condition ii) of the Lemma 2 is satisfied. Recall that

$$m_{jJ}(t) = \sqrt{J} 1(\tau_{j,J} \leq t < \tau_{j+1,J}), \ (j = 1, ..., J).$$

Then taking epectations over the simulation,

$$\begin{aligned} E[m_{jS}(\delta, b)] &= \bar{m}_j(\delta, b) = \int m_{jJ}(\tau_s(\delta, b)) F_s(ds) \\ &= \sqrt{J}[F(\tau_{j+1,J}|\delta, b) - F(\tau_{j,J}|\delta, b)], \ (j = 1, ..., J). \end{aligned}$$

From Assumption 5 let

$$\hat{D}_j(\tilde{\delta}, \tilde{b}) = D(\tilde{\delta}, \tilde{b}; \hat{\delta}, \hat{b}, \tau_j), \ D_j(\tilde{\delta}, \tilde{b}) = D(\tilde{\delta}, \tilde{b}; \delta, b, \tau_j).$$

By Assumption 5 a) and Lemma 5,

$$\begin{aligned} \bar{m}_j(\hat{\delta}, \hat{b}) - \bar{m}_j(\delta, b) &= \sqrt{J}[D_{j+1}(\hat{\delta} - \delta, \hat{b} - b) - D_j(\hat{\delta} - \delta, \hat{b} - b)] + \hat{R}_j, \\ \left|\hat{R}_j\right| &\leq \sqrt{J} 2C[(\hat{\delta} - \delta)^2 + \sup_t |\hat{b}(t) - b(t)|^2] = O_p(\sqrt{J}\varepsilon_{pn}^2), \end{aligned}$$

uniformly in j. By Assumption 5 b) and Lemmas 5 and 6,

$$\begin{aligned} & \sqrt{J}[D_{j+1}(\hat{\delta} - \delta, \hat{b} - b) - D_j(\hat{\delta} - \delta, \hat{b} - b)] \\ = \ & \sqrt{J}[(D^{\delta}_{0\tau_{j+1}} - D^{\delta}_{0\tau_j})(\hat{\delta} - \delta) + \int \{\alpha_{0,\tau_{j+1}}(\tau) - \alpha_{0,\tau_j}(\tau)\}\{\hat{b}(\tau) - b(\tau)\} F_0(d\tau)] \\ = \ & \sqrt{J}[(D^{\delta}_{0\tau_{j+1}} - D^{\delta}_{0\tau_j})\{\frac{1}{n}\sum_i \psi_i^{\delta} + O_p(\varepsilon_{pn}^2)\}] \\ & -\sqrt{J}\delta^{-1} E[\{\alpha_{0,\tau_{j+1}}(\tau_i) - \alpha_{0,\tau_j}(\tau_i)\} b(\tau_i)]) \left(\frac{1}{n}\sum_i \psi_i^{\delta}\right) \\ & +\sqrt{J}\frac{1}{n}\sum_i \frac{[\alpha_{0,\tau_{j+1}}(\tau_i) - \alpha_{0,\tau_j}(\tau_i)]}{\delta p_i[1 - p_i]}(\gamma_i - p_i) + \sqrt{J} O_p(\varepsilon_{pn}^2) \\ = \ & \frac{1}{n}\sum_i h_{ji} + O_p(\sqrt{J}\varepsilon_{pn}^2) \end{aligned}$$

Then by $tr(\Omega)^{1/2} = O(J)$ we have

$$J^{-1/2}\sqrt{n} tr\,(\Omega)^{1/2} \left\|\hat{h} - \bar{h}\right\| \leq C J^{1/2}\sqrt{n}\left\|\hat{h} - \bar{h}\right\| \leq C\sqrt{n}\sqrt{J} O_p(\sqrt{J}\varepsilon_{pn}^2).$$

Hypothesis ii) of Lemma 2 then follows by $\sqrt{n}J\varepsilon_{pn}^2 \longrightarrow 0$, and by Lemma 3 and $nJ^3/S \longrightarrow 0$.

Next we verify hypothesis iii) of Lemma 2. Note that

$$\hat{M}_{\delta j} = \frac{\hat{m}_j(\hat{\delta} + \Delta, \hat{\beta}) - \hat{m}_j(\hat{\delta} - \Delta, \hat{\beta})}{2\Delta}$$

Let $\bar{m}_j(\delta, \beta) = \int m_j(\tau_s(\delta, \beta)) F(ds)$ and

$$\bar{M}_{\delta j} = \frac{\bar{m}_j\left(\hat{\delta} + \Delta, \hat{\beta}\right) - \bar{m}_j\left(\hat{\delta} - \Delta, \hat{\beta}\right)}{2\Delta}.$$

By the simulations i.i.d. given $\hat{\delta}, \hat{\beta}$ and $m_{jJ}(\tau) \leq C\sqrt{J}$,

$$E\left[\left(\hat{M}_{\delta j} - \bar{M}_{\delta j}\right)^2 \mid \hat{\delta}, \hat{\beta}\right] \leq \frac{CJ}{S\Delta^2}.$$

Then for $\bar{M}_{\delta} = (\bar{M}_{\delta 1}, ..., \bar{M}_{\delta J})'$ the Markov inequality gives

$$E\left[\left\|\hat{M}_{\delta} - \bar{M}_{\delta}\right\|^2\right] \leq \frac{CJ^2}{S\Delta^2}, \ \left\|\hat{M}_{\delta} - \bar{M}_{\delta}\right\| = O_p\left(\frac{J}{\sqrt{S}\Delta}\right).$$

Note that replacing $\hat{\delta}$ with $\hat{\delta}+\Delta$ in the boundary estimator $\hat{b}$ gives $[\hat{\delta}/(\hat{\delta}+\Delta)]\hat{b}$ and replacing $\hat{\delta}$ with $\hat{\delta}-\Delta$ gives $[\hat{\delta}/(\hat{\delta}-\Delta)]\hat{b}$. Also,

$$\frac{\hat{\delta}}{\hat{\delta}+\Delta}-1=\frac{-\Delta}{\hat{\delta}+\Delta},\ \frac{\hat{\delta}}{\hat{\delta}-\Delta}-1=\frac{\Delta}{\hat{\delta}-\Delta}$$

Let $\hat{D}_j(\delta,b)=D(\delta,b;\hat{\delta},\hat{b},j)$ and $D_j(\delta,b)=D(\delta,b;\delta_0,b_0,j)$ for true values δ_0 and b_0. Then by Assumption 5 a),

$$\begin{aligned}
\bar{M}_{\delta j} &= \frac{\bar{m}_j\left(\hat{\delta}+\Delta,\hat{\beta}\right)-\bar{m}_j(\hat{\delta},\hat{\beta})-[\bar{m}_j\left(\hat{\delta}-\Delta,\hat{\beta}\right)-\bar{m}_j(\hat{\delta},\hat{\beta})]}{2\Delta} \\
&= \frac{\sqrt{J}[\hat{D}_{j+1}(\Delta,\frac{-\Delta}{\hat{\delta}+\Delta}\hat{b})-\hat{D}_{j+1}(-\Delta,\frac{\Delta}{\hat{\delta}-\Delta}\hat{b};)]}{2\Delta}-\frac{\sqrt{J}[\hat{D}_j(\Delta,\frac{-\Delta}{\hat{\delta}+\Delta}\hat{b})-\hat{D}_j(-\Delta,\frac{\Delta}{\hat{\delta}-\Delta}\hat{b})]}{2\Delta}+\hat{R}_j \\
\left|\hat{R}_j\right| &\leq C\sqrt{J}\Delta^{-1}(\Delta^2+\left|\frac{\Delta}{\hat{\delta}+\Delta}\hat{b}\right|^2+\left|\frac{\Delta}{\hat{\delta}-\Delta}\hat{b}\right|^2)\leq C\sqrt{J}\Delta(1+\left|\hat{b}\right|^2).
\end{aligned}$$

We also have

$$\begin{aligned}
\sqrt{J}\frac{1}{\Delta}\hat{D}_{j+1}(\Delta,\frac{-\Delta}{\hat{\delta}+\Delta}\hat{b}) &= \sqrt{J}\hat{D}_{j+1}(1,\frac{-1}{\hat{\delta}+\Delta}\hat{b}), \\
\sqrt{J}|\hat{D}_{j+1}(1,\frac{-1}{\hat{\delta}+\Delta}\hat{b})-D_{j+1}(1,\frac{-1}{\hat{\delta}+\Delta}\hat{b})| &\leq C\sqrt{J}\left|\frac{\hat{b}}{\hat{\delta}+\Delta}\right|(|\hat{\delta}-\delta|+|\hat{b}-b|)\leq C\sqrt{J}O_p(\varepsilon_{pn}).
\end{aligned}$$

Also,

$$\begin{aligned}
&\sqrt{J}\left|D_{j+1}(1,\frac{-1}{\hat{\delta}+\Delta}\hat{b})-D^{\delta}_{0\tau_{j+1}}+\frac{1}{\delta}\int\alpha_{0,\tau_{j+1}}(\tau)b(\tau)F_0(d\tau)\right| \\
&\leq C\sqrt{J}(|\hat{\delta}-\delta|+|\hat{b}-b|)=\sqrt{J}O_p(\varepsilon_{pn}).
\end{aligned}$$

Applying an analogous set of inequalities to other terms and collecting remainders gives

$$\left|\bar{M}_{\delta j}-M_{\delta j}\right|\leq C\sqrt{J}(\Delta+O_p(\varepsilon_{pn})).$$

Combining results and stacking over j then give

$$\left\|\hat{M}_{\delta}-M_{\delta}\right\|=O_p(J(\frac{1}{\sqrt{S}\Delta}+\Delta+\varepsilon_{pn})).$$

Next, for $\hat{\psi}^{\tau}_i=\left(2\hat{\delta}\bar{\tau}\right)^{-1}[\hat{I}(\tau_i)-\bar{I}-\hat{\delta}^2\{\tau_i-\bar{\tau}\}]$ it follows straighforwardly that

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\psi}^{\tau}_i-\psi^{\tau}_i\right)^2=O_p(\varepsilon^2_{pn}).$$

Let $\tilde{V}_1=n^{-1}\sum_{i=1}^{n}\psi_{1i}\psi'_{1i}$ and $\psi_{1i}=m_i-E[m_i]+M_{\delta}\psi^{\tau}_i$. Note that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left\|\hat{\psi}_{1i}-\psi_{1i}\right\|^2 &\leq \|\bar{m}-E[m_i]\|^2+\left\|\hat{M}_{\delta}-M_{\delta}\right\|^2\frac{1}{n}\sum_{i=1}^{n}\left\|\hat{\psi}^{\tau}_i\right\|^2+\|M_{\delta}\|^2\frac{1}{n}\sum_{i=1}^{n}(\hat{\psi}_{1i}-\psi_{1i})^2 \\
&= O_p(\frac{J^2}{n})+O_p(J^2(\frac{1}{\sqrt{S}\Delta}+\Delta+\varepsilon_{pn})^2)+O_p(J^2\varepsilon^2_{pn}) \\
&= O_p(J^2(\frac{1}{\sqrt{S}\Delta}+\Delta+\varepsilon_{pn})^2).
\end{aligned}$$

Then by the Cauchy-Schwartz and triangle inequalities,

$$\begin{aligned}
\left\|\hat{V}_1-\tilde{V}_1\right\| &\leq \frac{1}{n}\sum_{i=1}^{n}\left\|\hat{\psi}_{1i}-\psi_{1i}\right\|^2+\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\hat{\psi}_{1i}-\psi_{1i}\right\|^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\|\psi_{1i}\|^2} \\
&= O_p(J^2(\frac{1}{\sqrt{S}\Delta}+\Delta+\varepsilon_{pn})).
\end{aligned}$$

It follows similarly that $\left\|\tilde{V}_1-V_1\right\|=O_p(J^{3/2}/\sqrt{n})$, so by the triangle inequality,

$$\left\|\hat{V}_1-V_1\right\|=O_p(J^2(\frac{1}{\sqrt{S}\Delta}+\Delta+\varepsilon_{pn})).$$

Next we derive a convergence rate for $\left\|\hat{V}_2 - V_2\right\|$. Let

$$
\begin{aligned}
D_\beta &= E[\alpha_0(\tau_i)q_i^{K\prime}],\ \Sigma = E[q_i^K q_i^{K\prime}],\ \alpha_K(\tau_i) = D_\beta \Sigma^{-1} q_i^K, \\
\Lambda &= E[q_i^K q_i^{K\prime}(\gamma_i - p_i)^2],\ \bar{V}_2 = D_\beta \Sigma^{-1}\Lambda\Sigma^{-1}D_\beta' = E[\alpha_K(\tau_i)\alpha_K(\tau_i)'(\gamma_i - p_i)^2].
\end{aligned}
$$

Note that by Assumption 5 b) and standard approximation properties of splines

$$
E[\{(\alpha_{0j}(\tau_i) - \alpha_{Kj}(\tau_i))(\gamma_i - p_i)\}^2] \le CE[\{\alpha_{0j}(\tau_i) - \alpha_{Kj}(\tau_i)\}^2] \le CK^{-2s_\alpha},
$$

for a constant C that does not epend on j. Then we have

$$
\begin{aligned}
\left\|\bar{V}_2 - V_2\right\|^2 &= \sum_{j,\ell=1}^{J}\{E[\alpha_{Kj}(\tau_i)\alpha_{K\ell}(\tau_i)(\gamma_i - p_i)^2] - E[\alpha_{0j}(\tau_i)\alpha_{0\ell}(\tau_i)(\gamma_i - p_i)^2]\}^2 \\
&= \sum_{j,\ell=1}^{J}\{E[\{\alpha_{Kj}(\tau_i) - \alpha_{0j}(\tau_i)\}\alpha_{K\ell}(\tau_i)(\gamma_i - p_i)^2] + E[\alpha_{0j}(\tau_i)\{\alpha_{K\ell}(\tau_i) - \alpha_{0\ell}(\tau_i)\}(\gamma_i - p_i)^2]\}^2 \\
&\le C\sum_{j,\ell=1}^{J}\{\sqrt{E[\{\alpha_{Kj}(\tau_i) - \alpha_{0j}(\tau_i)\}^2]}\sqrt{E[\alpha_{K\ell}(\tau_i)^2]} \\
&\quad + \sqrt{E[\{\alpha_{K\ell}(\tau_i) - \alpha_{0\ell}(\tau_i)\}^2]}\sqrt{E[\alpha_{0j}(\tau_i)^2]}\}^2 \\
&\le C\left(\sum_{j=1}^{J}E[\{\alpha_{Kj}(\tau_i) - \alpha_{0j}(\tau_i)\}^2]\right)\left(\sum_{j=1}^{J}\{E[\alpha_{0j}(\tau_i)^2] + E[\alpha_{K\ell}(\tau_i)^2]\}\right) \le CJ^2K^{-2s_\alpha}.
\end{aligned}
$$

Taking square roots we have

$$
\left\|\bar{V}_2 - V_2\right\| \le CJK^{-s_\alpha}.
$$

Define

$$
\bar{M}_{\beta jk} = \frac{\bar{m}_j\left(\hat{\delta}, \hat{\beta} + e_k\Delta\right) - \bar{m}_j\left(\hat{\delta}, \hat{\beta} - e_k\Delta\right)}{2\Delta}.
$$

It follows similarly to $\left\|\hat{M}_\delta - \bar{M}_\delta\right\| = \left\|\hat{M}_\delta - \bar{M}_\delta\right\| = O_p\left(J/\sqrt{S}\Delta\right)$ that

$$
\left\|\hat{M}_\beta - \bar{M}_\beta\right\| = O_p\left(J\sqrt{K}/\sqrt{S}\Delta\right).
$$

Next, let $\hat{p}_{\Delta k}(t) = \hat{p}(t) + \Delta q_{kK}(G(t))$ and $\hat{b}_{\Delta k}(t) = \hat{\delta}^{-1}\ln(\hat{p}_{\Delta k}(t)/[1 - \hat{p}_{\Delta k}(t)])$. By $\Delta\sqrt{K} \longrightarrow 0$ and $\sup_{G\in[0,1]}|q_{kK}(G)| \le C\sqrt{K}$ it follows that $\sup_t \Delta q_{kK}(G(t)) \longrightarrow 0$. Then w.p.a.1 we have

$$
\hat{b}_{\Delta k}(t) = \hat{b}(t) + \frac{\Delta q_{kK}(G(t))}{\hat{\delta}\hat{p}(t)[1 - \hat{p}(t)]} + \hat{R}_k(t,\Delta),\ \ \left|\hat{R}_k(t,\Delta)\right| \le C\Delta^2 K.
$$

Then we have

$$
\begin{aligned}
\bar{M}_{\beta jk} &= \frac{\bar{m}_j\left(\hat{\delta}, \hat{\beta} + e_k\Delta\right) - \bar{m}_j(\hat{\delta}, \hat{\beta}) - [\bar{m}_j\left(\hat{\delta}, \hat{\beta} - e_k\Delta\right) - \bar{m}_j(\hat{\delta}, \hat{\beta})]}{2\Delta} \\
&= \frac{\sqrt{J}[\hat{D}_{j+1}(0, \hat{b}_{\Delta k} - \hat{b}) - \hat{D}_{j+1}(0, \hat{b}_{-\Delta k} - \hat{b})]}{2\Delta} \\
&\quad - \frac{\sqrt{J}[\hat{D}_j(0, \hat{b}_{\Delta k} - \hat{b}) - \hat{D}_j(0, \hat{b}_{-\Delta k} - \hat{b})]}{2\Delta} + \hat{R}_{jk} \\
\left|\hat{R}_{jk}\right| &\le C\sqrt{J}\Delta^{-1}(\left|\hat{b}_{\Delta k} - \hat{b}\right|^2 + \left|\hat{b}_{-\Delta,k} - \hat{b}\right|^2) \le C\sqrt{J}\Delta K.
\end{aligned}
$$

We also have

$$
\begin{aligned}
\sqrt{J}\frac{1}{\Delta}\hat{D}_{j+1}(0, \hat{b}_{\Delta k} - \hat{b}) &= \sqrt{J}\hat{D}_{j+1}(0, \frac{\hat{b}_{\Delta k} - \hat{b}}{\Delta}), \\
\sqrt{J}|\hat{D}_{j+1}(0, \frac{\hat{b}_{\Delta k} - \hat{b}}{\Delta}) - D_{j+1}(0, \frac{\hat{b}_{\Delta k} - \hat{b}}{\Delta})| &\le C\sqrt{J}\left|\frac{\hat{b}_{\Delta k} - \hat{b}}{\Delta}\right|(|\hat{\delta} - \delta| + |\hat{b} - b|) \le C\sqrt{J}\sqrt{K}O_p(\varepsilon_{pn}).
\end{aligned}
$$

In addition

$$\sqrt{J}D_{j+1}(0,\frac{\hat{b}_{\Delta k}-\hat{b}}{\Delta};\delta,b,\tau_{j+1}) = \sqrt{J}D(0,\frac{q_{kK}(G(\cdot))}{\hat{\delta}\hat{p}(\cdot)[1-\hat{p}(\cdot)]};\delta,b,\tau_{j+1}) + \sqrt{J}\Delta D(0,\hat{R}_k(\cdot,\Delta);\delta,b,\tau_{j+1})$$
$$= \sqrt{J}D(0,\frac{q_{kK}(G(\cdot))}{\delta p(\cdot)[1-p(\cdot)]};\delta,b,\tau_{j+1}) + \hat{R}_{jk},$$
$$\left|\hat{R}_{jk}\right| \le \sqrt{J}\sqrt{K}O_p(\varepsilon_{pn}) + \sqrt{J}K\Delta.$$

Combining terms we have

$$\left\|\hat{M}_\beta - M_\beta\right\| = O_p(J\sqrt{K}/\sqrt{S}\Delta + JK\varepsilon_{pn} + JK^{3/2}\Delta)$$

Next, we have

$$\left\|\hat{M}_\delta \frac{1}{2\hat{\delta}\bar{\tau}n}\sum_{i=1}^n I_p(\hat{p}_i)q_i^{K\prime} - M_\delta \frac{1}{2\delta E[\tau_i]}E[I_p(p_i)q_i^{K\prime}]\right\|$$
$$\le \left\|\hat{M}_\delta - M_\delta\right\| \frac{1}{2\hat{\delta}\bar{\tau}}\left(\frac{1}{n}\sum_{i=1}^n I_p(\hat{p}_i)^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n q_i^{K\prime}q_i^K\right)^{1/2}$$
$$+\left\|M_\delta\right\|\left\|\frac{1}{2\hat{\delta}\bar{\tau}n}\sum_{i=1}^n I_p(\hat{p}_i)q_i^{K\prime} - \frac{1}{2\delta E[\tau_i]}E[I_p(p_i)q_i^{K\prime}]\right\|$$
$$= O_p(J\sqrt{K}(\frac{1}{\sqrt{S}\Delta}+\Delta+\varepsilon_{pn})) + O_p(JK\varepsilon_{pn}) = O_p(J\sqrt{K}(\frac{1}{\sqrt{S}\Delta}+\Delta+\sqrt{K}\varepsilon_{pn})).$$

Combining terms we then have

$$\left\|\hat{D}_\beta - D_\beta\right\| = O_p(J\sqrt{K}/\sqrt{S}\Delta + JK\varepsilon_{pn} + JK^{3/2}\Delta).$$

Next, for $\hat{\pi} = \hat{\Sigma}^{-1}\hat{D}_\beta$ and $\pi = \Sigma^{-1}D_\beta$ note that $\hat{V}_2 = \hat{\pi}'\hat{\Lambda}\hat{\pi}$ and $\bar{V}_2 = \pi'\Lambda\pi$. Also we have

$$\hat{V}_2 - \bar{V}_2 = (\hat{\pi}-\pi)'\hat{\Lambda}(\hat{\pi}-\pi) + 2\pi'\hat{\Lambda}(\hat{\pi}-\pi) + \pi'(\hat{\Lambda}-\Lambda)\pi.$$

By the law of large number for symmetric matrices, $\left\|\hat{\Sigma}-\Sigma\right\|_{op} = O_p\left(\sqrt{n^{-1}K\ln K}\right) = o_p(1)$, where $\|\cdot\|_{op}$ denotes the operator norm on symmetric matrices. Then by the eigenvalues of Σ bounded and bounded away from zero, $\lambda_{\max}(\hat{\Sigma}) = O_p(1)$ and $1/\lambda_{\min}(\hat{\Sigma}) = O_p(1)$. Let $\tilde{\Lambda} = \frac{1}{n}\sum_i q_i^K q_i^{K\prime}(\gamma_i - p_i)^2$. Note that

$$\hat{\Lambda} - \tilde{\Lambda} = \frac{1}{n}\sum_i q_i^K q_i^{K\prime}\left[(\gamma_i-\hat{p}_i)^2 - (\gamma_i-p_i)^2\right] \le \frac{1}{n}\sum_i q_i^K q_i^{K\prime}\left|(\gamma_i-\hat{p}_i)^2 - (\gamma_i-p_i)^2\right|$$
$$\le C\hat{\Sigma}\max_i|\hat{p}_i - p_i| = \hat{\Sigma}O_p(\varepsilon_{pn}),\ \hat{\Lambda}-\tilde{\Lambda} \ge -C\hat{\Sigma}O_p(\varepsilon_{pn}).$$

Also by the law of large numbers for symmetric matrices $\left\|\tilde{\Lambda}-\Lambda\right\|_{op} = O_p\left(\sqrt{n^{-1}K\ln K}\right)$. Therefore by the triangle inequality,

$$\left\|\hat{\Lambda}-\Lambda\right\|_{op} = O_p(\varepsilon_{pn}).$$

It follows that $\lambda_{\max}(\hat{\Lambda}) = O_p(1)$, $1/\lambda_{\min}(\hat{\Lambda}) = O_p(1)$, and for $\hat{\Upsilon} = \hat{\Lambda}-\Lambda$,

$$\left\|\hat{\Upsilon}\right\| = \sqrt{tr(\hat{\Upsilon}^2)} \le C\sqrt{J}\left\|\hat{\Lambda}-\Lambda\right\|_{op} = O_p(\sqrt{J}\varepsilon_{pn}).$$

Similarly we have $\left\|\hat{\Sigma}-\Sigma\right\| = O_p(K\sqrt{\ln(K)/n})$. We also have $\|D_\beta\| \le CJ\sqrt{K}$.Then it follows that for $\varepsilon_{Dn} = J\sqrt{K}/\sqrt{S}\Delta + JK\varepsilon_{pn} + JK^{3/2}\Delta$

$$\|\hat{\pi}-\pi\| \le \left\|(\hat{D}_\beta - D_\beta)'\hat{\Sigma}^{-1}\right\| + \left\|D_\beta{}'\hat{\Sigma}^{-1}(\Sigma-\hat{\Sigma})\Sigma^{-1}\right\| \le O_p(\varepsilon_{Dn}) + O_p(JK\sqrt{\ln(K)/n}) = O_p(\varepsilon_{Dn}).$$

It then follows by the triangle inequality that

$$\left\|\hat{V}_2 - \bar{V}_2\right\| \le O_p(1)(\|\hat{\pi}-\pi\|^2 + \|\pi\|\|\hat{\pi}-\pi\| + \|\pi\|^2\left\|\hat{\Lambda}-\Lambda\right\|)$$
$$= O_p(J\sqrt{K}\varepsilon_{Dn} + J^2K^2\sqrt{\ln(K)/n}) = O_p(J^2K/\sqrt{S}\Delta + J^2K^{3/2}\varepsilon_{pn} + J^2K\Delta).$$

By the triangle inequality we then have

$$\left\|\hat{\Omega}-\Omega\right\| = O_p(J^2K/\sqrt{S}\Delta + J^2K\Delta + J^2K^{3/2}\varepsilon_{pn} + JK^{-s_\alpha})$$

It then follows that Assumption iii) is satisfied by Assumption 5 e).

Finally, for Assumption iv) of Lemma A2, note that

$$\left(h_i'h_i\right)^2 = \left(\sum_{j=1}^{J} h_{ij}^2\right)^2 = \sum_{j=1}^{J}\sum_{k=1}^{K} h_{ij}^2 h_{ik}^2 \leq CJ\sum_{j=1}^{J} h_{ij}^4 \leq CJ^4,$$

so that

$$E\left[\left(h_i'h_i\right)^2\right]/nJ \leq CJ^3/n \longrightarrow 0.$$

Therefore condition iv) is satisfied. **Q.E.D.**

References

1. D Fudenberg, P Strack, T Strzalecki, Speed, accuracy, and the optimal timing of choices. *Am. Econ. Rev.* **108**, 3651–84 (2018).
2. WK Newey, The asymptotic variance of semiparametric estimators. *Econometrica*, 1349–1382 (1994).
3. WK Newey, F Windmeijer, Generalized method of moments with many weak moment conditions. *Econometrica* **77**, 687–719 (2009).
4. A Belloni, V Chernozhukov, D Chetverikov, K Kato, Some new asymptotic theory for least squares series: Pointwise and uniform results. *J. Econom.* **186**, 345–366 (2015).
5. H Ichimura, WK Newey, The influence function of semiparametric estimators. *mimeo* (2018).