Testing the Drift-Diffusion Model

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

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

This manuscript was compiled on October 26, 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 consumption 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

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" (otherwise known as value-based tasks, or preferential 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. Choice Problems and Choice Processes

The agent is facing a binary *choice problem* c between action x and action y. In consumption tasks x and y are items the agent is choosing between. To allow for presentation effects, we view $c := (x, y)$ as an ordered pair, so $(x, y) \neq (y, x)$; in applications to laboratory data we let x denote the left-hand or top-most action. In perception tasks x and y are the two

*They ignore the issue of correlation between response times and 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: drewf@mit.edu

answers to the perceptual question; here x and y are held constant over all choice problems and d encodes the strength of the perceptual stimulus, e.g., the fraction of dots on the screen moving to the left. Let C denote the collection of choice problems observed by the analyst.

Let $t \in \mathbb{R}_+$ denote time. In each trial the analyst observes the action chosen and the decision time. 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 $F^c(t)$ the probability that the agent makes a choice by time t, and let $p^c(t)$ be the 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 $F(0) = 0$. We also assume that F has a strictly postive density $F' > 0$, and that $\lim_{t\to\infty} F(t) = 1$. [†] 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 also assume that each option is chosen with positive conditional probability at each time, so $0 < p^c(t) < 1$ for all t. We call (p^c, F^c) a *choice process.*

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

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

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

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

$\bar{T}^c$ to be the average decision time,

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

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

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

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

2. DDM representation

The drift diffusion model (DDM) is commonly used to explain choice processes in neuroscience and psychology. Througout, we call a function $b : \mathbb{R}_+ \to \mathbb{R}$ a *boundary* if it is continuous, non-negative, and eventually bounded.[‡] The two main ingredients of a DDM are the stimulus process Z and a boundary function b. 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 can be uniquely recovered from the data.

Definition 1 (DDM Representation)**.** Choice process (p^c, F^c) has a DDM representation if there exists a drift δ^c, a volatility parameter $\alpha^c > 0$ as well as a boundary $b^c : \mathbb{R}_+ \to \mathbb{R}_+$ such that for all $x, y \in X$ and $t \in \mathbb{R}$

$$p^c(t) = p^*\left(t, \delta^c, b^c, \alpha^c\right)$$
$$\text{and } F^c(t) = F^*\left(t, \delta^c, b^c, \alpha^c\right).$$

The original formulation of the DDM was for perception tasks where the drift δ^c is a function of the strength of the stimulus process in choice problem c. In consumption tasks researchers typically assume that the drift δ^c equals the difference between the utility of the two items, i.e., $\delta^c = u(x) - u(y)$ for all $c = (x, y)$, see, e.g., (16). Both formulations require that the boundary is the same for all decision problems. 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 choice processes admit a DDM representation. The following result follows immediately from rescaling δ and b.

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

We will thus without loss of generality normalize $\alpha = 1$. 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 choice process (p^c, F^c), define the *revealed drift*

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

The revealed drift is high when the agent makes very imbalanced choices or tends to decide quickly, and is low for choices that are closer to 50-50 or made more slowly.

[†] Many empirical applications of the DDM include an initial deterministic or stochastic "non-decision time" where no decision can be made. The assumption in the text allows for an arbitarily small probability of stopping on any finite time interval, which is observationally equivalent to 0 probaility on any finite data set.

[‡] That is, there exists $\bar{b}$ and $\bar{T}$ such that $b(t) \leq \bar{b}$ for all $t > T$. The model can be extended to allow the boundary to initially be infinite, which means that the agent never stops in an initial interval of time.

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

When $\tilde{\delta}^c$ is non zero and $(p^c(t) - 1/2)\tilde{\delta}^c > 0$ for all t, we define the *revealed boundary* as

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

The revealed boundary 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 where the revealed drift is 0, the revealed boundary is not defined, and our results do not apply. Similarly, for t such that $(p^c(t) - 1/2)\tilde{\delta}^c < 0$, $\tilde{b}^c(t) < 0$, and $\tilde{b}^c$ is not a well defined boundary.

We can extend the identification theorems below to accomodate a deterministic non-decision time by allowing the boundary to be infinite. However, if the non-decision time is stochastic, we conjecture that its distribution cannot be separately identified without restrictions on the shape of the boundary.

A. Characterization for a fixed decision problem. Our first result characterizes the DDM for a fixed decision problem $c \in C$ and the revealed drift and boundary will exactly match the true parameters. We rule out the knife edge case where the revealed drift equals zero to ensure that the revealed boundary is well defined.¶

Theorem 1. *For c with $\tilde{\delta}^c \neq 0$ the choice process (p^c, F^c) admits a DDM representation if and only if $\tilde{b}^c(t) \geq 0$ for all $t \geq 0$ and*

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

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

Thus, the choice process (p^c, F^c) is consistent with DDM whenever the observed distribution of stopping times F^c equals the distribution of hitting times generated by the revealed drift $\tilde{\delta}^c$ and revealed boundary $\tilde{b}^c$. Theorem 1 shows that for $\tilde{\delta}^c \neq 0$ 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 both the model and the estimated parameters. The estimated drift in the DDM model is a measure of how imbalanced and quick the agent's choices are, and the shape of the estimated boundary follows the imbalance of the agent's choices over time. This interpretation makes the empirical content of the parameters of DDM model more transparent and the model thus more useful. Moreover, as we show in Section 4, Theorem 1 allows us to test whether the true data generating process is indeed a DDM.

Note that this theorem shows that the distribution of stopping times contains additional information that is not captured by the mean. For example, a choice process where $p^c(t)$ and $\bar{T}^c$ are any two given constants is only consistent with one possible distribution of stopping times F^c. 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 will accept a constant boundary regardless of how the choice probability varies over time, thus leading to false positives.

¶ If the revealed drift equals zero, one needs to recover the boundary from the distribution of decision times F^c. This is an open problem in the mathematical literature. See Appendix A for further discussion.

B. Characterization for consumption tasks. Here X is the set of consumption alternatives, and each choice problem c consists of a pair of alternatives, so, in this section we index choice problems by superscript xy. For consumption tasks we assume that the order of the items does not matter. This is formally equivalent to a condition that we call *symmetry*:

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

Definition 2 (DDM Representation). A choice process $(p^{xy}, F^{xy})_{x,y \in X}$ has a choice-DDM representation if there exists a utility function $u : X \to \mathbb{R}$, and a boundary $b : \mathbb{R}_+ \to \mathbb{R}_+$ such that for all $x, y \in X$ and $t \in \mathbb{R}$

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

Theorem 2. *Suppose that the choice process $(p^{xy}, F^{xy})_{x,y \in X}$ has $\tilde{\delta}^{xy} \neq 0$ for all $x, y \in X$. It has a choice DDM representation iff*

(i) it is symmetric,

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

(iii) $\tilde{b}^{(x,y)}(t) = \tilde{b}^{(x,z)}(t)$ for all $x, y, z \in X$ and all $t \geq 0$.

(iv) $\tilde{\delta}^{(x,y)} + \tilde{\delta}^{(y,z)} = \tilde{\delta}^{(x,z)}$ 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 (iii) follows from our assumption that the agent uses the same stopping boundary in every menu. Condition (iv) comes from the assumption that the drift in a given menu depends on the difference of utilities, that is $\delta^{xy} = u(x) - u(y)$.‖

An analogous exercise could be done for perception tasks. Here condition (i) would be dropped and (iv) would be replaced with a different, perhaps more complicated condition that specifies the drift as a (potentially parametric) function of the stimulus in choice problem c.**

4. A Statistical Test for a Fixed Pair of Alternatives

The test we give is based on comparing model predictions with data estimates. We construct estimators of the drift and boundary for this test, that are of interest in their own right. Constructing these estimators is greatly aided by the explicit formulas for the drift and boundary given in Eq. (3) and Eq. (4). We estimate choice probabilities nonparametrically and plug them in the formulas, replacing expectations with sample averages, to estimate the revealed drift and boundary. We then

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

** Other exercises along these lines are possible. For instance, (36) models consumption-tasks by an accumulator model where the item-specific signals are correlated. This amounts to dropping conditions (iii) and (iv) since it is equivalent to DDM where both the drift and the boundary depend on x and y.

simulate many stopping times using the drift and boundary estimates. Simulation consistently estimates averages implied by the model, as in (37) and (38). We form a chi-squared test based on differences of the average over the simulations and over the sample of functions of the stopping time.

A. Estimation of drift and boundary. An essential ingredient for the drift and boundary estimators and for the test of the model is an estimator of the choice probability $p^c(t)$ conditional on decision occurring at time t. We focus on a linear probability estimator $\hat{p}(t)$ obtained as the predicted value from a linear regression of observations of the choice indicator data (a vector of zeros and ones) on functions of t. This estimator will be nonparametric by virtue of using flexible regressors that are designed to approximate any function. We consider both power series and piecewise linear functions for the regressors.

The regularity conditions we give assume that the boundary is bounded. An unbounded boundary would be needed to accomodate a deterministic non-decision time. Unboundedness is diffcult to allow for in regularity conditions involving nonparametric estimation.

To describe the estimators and the test, let the data consist of n observations $(\tau_1, \gamma_1), \dots, (\tau_n, \gamma_n)$ of the decision time τ_i and an indicator variable $\gamma_i \in \{0, 1\}$ that is equal to 1 if choice d is made and 0 otherwise, for $i = 1, ..., n$. We construct $\hat{p}(t)$ from a linear regression of γ_i on functions of $G(\tau_i)$, where $G(\tau)$ is a strictly increasing cumulative distribution function (CDF) that lies in the unit interval $[0, 1]$. Use of $G(\tau)$ allows for unbounded τ_i.[††] The resulting choice probability estimator $\hat{p}(t)$ is described in detail in an Appendix. Conditions for $\hat{p}(t)$ to be consistent and have other important large sample properties are given in Assumptions 2 and 3 to follow.

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

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

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

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. The test is based on comparing sample averages of functions of stopping times from the data with simulated averages implied by the estimators of the revealed drift and boundary. To describe the test let $m_J(\tau) = (m_{1J}(\tau), ..., m_{JJ}(\tau))'$ be a $J \times 1$ vector of functions of τ. Examples of $m_{jJ}(\tau)$ include indicator functions for intervals and low order powers of $G(\tau)$. A sample moment vector is $\bar{m} = \sum_{i=1}^{n} m_J(\tau_i)/n$.[‡‡] To describe the simulations 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 $\hat{m}_S = \sum_{s=1}^{S} m_J(\hat{\tau}_s)/S$. 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. The test statistic is

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

The model would be rejected if $\hat{A}$ exceeds the critical value of a $\chi^2(J)$ distribution.

If J is allowed to grow slowly with n and $m_J(\tau)$ 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. If $m_J(\tau)$ is chosen so that any function of τ can be approximated by a linear combination $c'm_J(\tau)$ as J grows then the test must reject as J grows when the DDM model is incorrect. An incorrect DDM model will imply $c'\bar{m}$ and $c'\hat{m}_S$ have different probability limits for some c and J large enough. Also, $\hat{A} \geq n\{c'[\bar{m} - \hat{m}_S]\}^2 / \left\{c'\hat{V}c\right\}$, so $\hat{A}$ grows as fast as n. Restricting J to grow slowly with n makes the test reject for large enough n.

It is straightforward to construct $\hat{V}$ using the bootstrap. Each bootstrap replication starts with a random sample $Z_n^j = (\tau_1^j, y_1^j), ..., (\tau_n^j, y_n^j)$ consisting of i.i.d. observations (τ_i^j, y_i^j), $(i = 1, ..., n)$, drawn at random with replacement from the data observations. Here j is a positive integer that denotes the bootstrap replication with $(j = 1, ..., B)$, so there are B replications. 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 $\hat{\tau}_s^b$, $(s = 1, ..., S)$, are constructed as described above, resimulating for each bootstrap replication, and $\hat{m}_S^j = \sum_{s=1}^{S} m_J(\hat{\tau}_s^j)/S$ calculated. 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 $\hat{V}_B$ is

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

In Section 3 of SI we give another estimator $\hat{V}_n$ based on asymptotic theory. In simulations of synthetic data 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.

The test statistic is based only on the distribution of decision times, and does not involve model choice probabilities and alternatives chosen in the data. This feature of the test does not affect its power to detect failures of the DDM model, because the choice probabilities for the estimated DDM model are equal to the nonparametric estimates $\hat{p}(t)$. To see this result note that there is a one-to-one relationship between

[††] In DDM models where b does not reach zero, decision times are not bounded, so it is important to allow for an unbounded regressor.

[‡‡] 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 (39).

the revealed boundary and the choice probabilities (given the revealed drift), with revealed choice probabilities given by

$$p^c(t) = \frac{\exp(2\tilde{\delta}^c \tilde{b}(t))}{\exp(2\tilde{\delta}^c \tilde{b}(t)) + 1}.$$

Plugging in the estimated drift $\hat{\delta}$ and boundary $\hat{b}(t)$ to this formula gives choice probability $p^c(t) = \hat{p}(t)$ equal to the nonparametric estimate. Thus, the choice probability implied by the estimated DDM model is unrestricted. The joint distribution of decision time and choice is completely characterized by the marginal distribution of decision times and the conditional choice probability. Nothing is lost in excluding the conditional choice probability from the test because it is not restricted by the estimated model.

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

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

The test based on these functions is based on comparing the 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 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 1. The data $(\tau_1, \gamma_1), \ldots, (\tau_n, \gamma_n)$ are i.i.d.

This is the basic statistical condition that leads to the data being more informative as the sample size n grows.

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 $s \ge 1$ 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. $\delta \neq 0$.

This assumption is clearly important for the revealed boundary formula in equation (revealed boundary formula). When $\delta = 0$ this formula does not hold, $p^c(t) = 1/2$ for all t, and the boundary need not be constant. Consequently the test given here would not be correct. Given this sensitivity of model characteristics to $\delta \neq 0$ it may make sense to test the null hypothesis that $\delta = 0$. This null hypothesis can be tested using the estimator $\hat{\delta}$ and the bootstrap standard error $SE_B(\hat{\delta}) = \{\sum_{j=1}^{B}(\hat{\delta}_j - \bar{\delta}_B)^2/B\}^{1/2}$. A t-statistic $\left|\hat{\delta}/SE_B(\hat{\delta})\right|$ that is substantially greater than the standard Gaussian critical value of 1.96 would provide evidence that $\delta \neq 0$.

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 C.

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 C are satisfied. Then for the $1-\alpha$ quantile $c(\alpha, J)$ of a chi-square distribution with J degrees of freedom*

$$\mathbb{P}\left[\hat{A} \ge c(\alpha, J)\right] \longrightarrow \alpha.$$

This test could be extended to multiple-alternatives settings along the lines of Theorem 2, but we do not do so here.[§§]

5. Examples for Synthetic Data

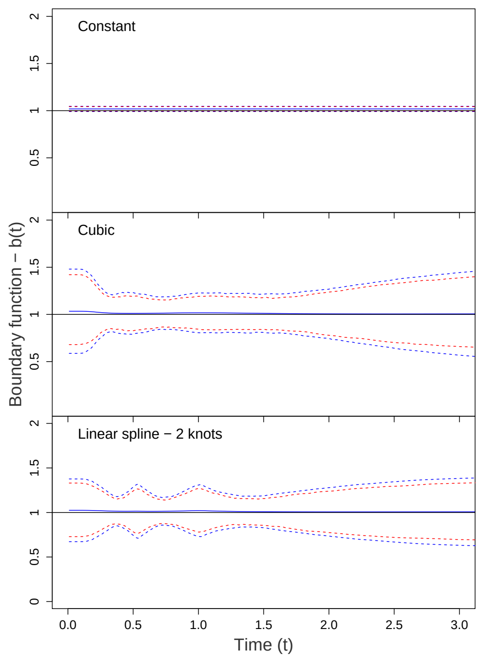
To consider how the estimators and test might work in practice we carry out some simulations where synthetic data was repeatedly generated from a DDM model. In the DDM model we set $\delta_0 = .5$ throughout and set the boundary to either be constant at -1 and 1. We set the sample size to be $n = 1000$ in each case. We consider three different boundary estimators: a constant boundary estimator where $\hat{p}(t)$ is the sample proportion that alternative 1 is chosen, a $\hat{p}(t)$ depending on cubic functions $(1, G, G^2, G^3)'$, and a continuous, piecewise linear function of G where the slope can change when G equals either .33 and .66. We repeat the generation of the simulated data and calculation of the estimators and test 500 times for each case.

Figure 2 plots the mean of and pointwise (inner) and uniform (outer) .025 and .975 quantile bands for the estimated boundary function. The quantile bands for the constant boundary are very small because the constant boundary is very precisely estimated relative to the boundaries with cubic and piecewise linear specifications. The quantile bands for cubic and piecewise linear boundaries seem large but are consistent with large sample approximations, as discussed in the Supplemental Information. In the Supplemental Information we find that $\hat{\delta}$ is a precise estimator of the drift parameter for sample size $n = 1000$.

Table 1 reports Monte Carlo rejection frequencies for the test statistic with bootstrap variance estimator. The $\hat{p}(t)$ is either does not depend on t or depends on piecewise linear functions of $G(t)$ with either no slope change, one slope change at $G = .5$, or two slope changes at $G = .33$ and .66. We consider the test statistic with bootstrap variance estimator $\hat{V}_B$ obtained from $B = 250$ bootstrap replications. We set

[§§] In allowing J to grow with sample size this result is like (40) and (41).

Figure 2: Boundary function estimation



$J = 5$ with only the middle three intervals included in the test statistic and $J = 8$ where only the middle six intervals are included. Rejection frequencies are given when critical values are chosen using the asymptotic chi-squared approximation with nominal rejection frequencies of 1, 5, 10, and 20 percent.

Table 1: Rejection Rates for Test Statistic

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

The acceptance regions for a test of level .10 that the rejection frequencies are equal their asymptotic values are $.010 \pm .006$, $.050 \pm .016$, $.100 \pm .022$, $.200 \pm .030$ for asymptotic levels .01, .05, .10, and .20 respectively. We find some tendency of the test statistic to reject too often when the number of intervals J is larger and the number of slope changes is larger. We found in additional simulations not reported here that for $\hat{p}(t)$ cubic in G or the analytic $\hat{V}$ the test statistic tended to overreject even more, especially for the analytic variance estimator. In the Appendix we give additional simulation results for $J = 5$ for a DDM model with an exponential boundary and for a Poisson model. There we find that the test has good power against the Poisson model, but shows little tendency to reject the DDM model with exponential boundary for $\hat{p}(t)$ piecewise linear in G with two slope changes. We also give rejection frequencies for the test for smaller sample sizes $n = 250$ and $n = 50\dot{0}$. There we find that the large sample approximation remains quite accurate for the smaller sample sizes for a constant and linear boundary specification, but the approximation is considerably worse than for $n = 1000$ when slope changes are included.

The tendency displayed in Table 1 to overreject for larger J and/or more flexible boundary specifications indicates some difficulty in reliably testing the implications of the DDM model with 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 1, where more parsimonious specifications tend to overeject less often. One potentially useful nonparametric restriction is monotonicity of the boundary, which could permit inference using the approach of (42). This seems potentially fruitful but is beyond the scope of this paper.

Appendix

A. Choice Problems with Zero Drift

When the drift in the DDM model is 0, $p(t) = 1/2$ for all $t \geq 0$, due to the symmetry of the problem. This implies the following extension of Theorem 1:

Theorem 4. *For c with $\tilde{\delta}^c = 0$ the choice process (p^c, F^c) admits a DDM representation if and only if $p^c \equiv 1/2$ and there exists $\tilde{b}^c$ such that for all $t \geq 0$*

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

In this case the boundary is not revealed by the choice probability. The question of how to recover the boundary from the distribution of stopping times is known as the "inverse first-passage time problem". The existence and uniqueness of the boundary remains an open problem even in the simpler case of a one-sided boundary and a Brownian motion with drift (see the introduction in (43)). Most closely related to our work is (44) whose Theorem 3.1 (under some regularity conditions) connects the boundary and the distribution over choice times in our model through a non-linear volterra integral equation, but does not prove that this equation admits a unique solution.

B. The Choice Probability Estimator

The choice probability estimator $\hat{p}\,(t)$ considered here is the predicted value from from a linear regression of γ_i on functions of $G(\tau_i)$. To describe $\hat{p}(t)$ let a $K \times 1$ vector of functions with domain $[0, 1]$ be

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

For example $q^K(G)$ could consist of powers of G or be piecewise linear functions of the form 1, G, and $1(G > \ell_{k-2})(G - \ell_{k-2})$, $(k = 3, ..., K)$. The $\hat{p}(t)$ we consider is

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

The transformation $G(\tau)$ to the unit interval helps $\hat{p}(t)$ be a good estimator with unbounded τ. It is helpful for this purpose to have $G(\tau_i)$ be quite evenly distributed over the unit interval, as near to uniform as possible. One possible choice of $G(\tau)$ is the cumulative distribution function of 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 $\delta_0 = .5$ and a constant boundary of -1 and 1.

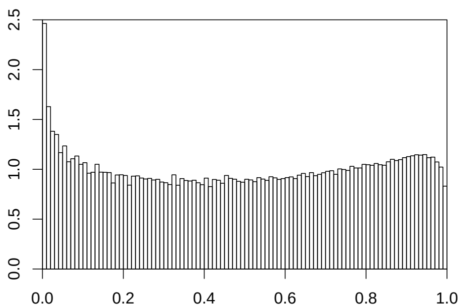


Fig. 1. Density of transformed FPT (τ)

The histogram is bounded well away from zero and infinity over most of its range so that we expect the linear probability estimator based on this $G(\tau)$ should work well. 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.

C. 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)

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

b) for each t there is a constant D^{δ}_{0t} and function $\alpha_{0t}(t)$ such that $|\alpha_{0t}(\tau_i)| \le C$, $\left|D^{\delta}_{0t}\right| \le C$, $|d^s\alpha_{0t}(t)/dt^s| \le C$ 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)

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

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,

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

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.

D. Additional Tests on Synthetic Data:

Table 2 gives rejection frequencies for the test on synthetic data from a DDM model with constant boundary, an exponential boundary $b(t) = 1/2 + 2\exp(-3t/2)$, and a Poisson process. The Poisson process has $p(t) = e^a/(e^a + e^b)$ and $F^*(t) = 1 - e^{-\lambda t}$ for $\lambda = e^a + e^b$, with a and b chosen to that $p(t)$ and $E[\tau]$ match those of DDM model with drift $1/2$ and $b(t) = 1$. Table 2 differs from Table 1 in one boundary slope changing at the sample median of $G(\tau_1), ..., G(\tau_n)$ rather than at .5 and two slopes changing at the .33 and .66 quantiles rather than at the values .33 and .66. Results in Table 2 are for $J = 5$ only. We continue to use $B = 250$ bootstrap replications and report results for 500 sythetic data set replications.

Table 2: Rejection Rates for Test Statistic

Model	Boundary Estimate	20%	10%	5%	1%
Constant Boundary	Constant	.182	.096	.048	.014
	Linear	.220	.128	.060	.012
	1 Slope Change	.186	.106	.060	.024
	2 Slope Changes	.236	.166	.106	.056
Exponential Boundary	Constant	1.00	1.00	1.00	1.00
	Linear	.354	.218	.140	.050
	1 Slope Change	.262	.164	.104	.036
	2 Slope Changes	.270	.152	.094	.028
Poisson	Constant	1.00	1.00	1.00	1.00
	Linear	.994	.988	.980	.904
	1 Slope Change	.862	.798	.696	.512
	2 Slope Changes	.522	.378	.282	.156

We find that for the DDM model with a constant boundary the test rejection frequencies increase as the specification of the boundary becomes richer, as in Table 1. Remarkably, for a DDM model with exponential boundary and a piecewise linear estimator with two slope changes, the rejection frequencies are similar to those where the boundary was constant. Thus, in this example specifying an incorrect piecewise linear boundary does not make the asymptotic approximation worse. We also find that the test has good power against a Poisson model, with the rejection frequencies being much larger when the data is generated by a Poisson model than when the data is generated by a DDM model.

To see the effect of smaller samples on the large sample approximation we also carried out simulations for $n = 250$ and $n = 500$ for the DDM model with constant boundary and $J = 5$. These results are reported in Table 3.

Table 3: Rejection Rates for Smaller Sample Size

n	Boundary Estimate	20%	10%	5%	1%
250	Constant	.216	.102	.040	.010
	Linear	.206	.116	.060	.020
	1 Slope Change	.256	.178	.136	.078
	2 Slope Changes	.320	.210	.168	.098
500	Constant	.200	.084	.038	.010
	Linear	.180	.090	.048	.018
	1 Slope Change	.224	.122	.072	.040
	2 Slope Changes	.294	.198	.144	.064

We find that the large sample approximation remains quite accurate for the smaller sample sizes for a constant and linear boundary specification but the approximation is considerably worse than for $n = 1000$ when slope changes are included.

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

1. Ratcliff R, McKoon G (2008) The diffusion decision model: Theory and data for two-choice decision tasks. *Neural computation* 20(4):873–922.
2. Shadlen MN, Kiani R (2013) Decision making as a window on cognition. *Neuron* 80(3):791–806.
3. Fehr E, Rangel A (2011) Neuroeconomic foundations of economic choice—recent advances. *The Journal of Economic Perspectives* 25(4):3–30.
4. Roe RM, Busemeyer JR, Townsend JT (2001) Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological review* 108(2):370.
5. Clithero JA (2018) Improving out-of-sample predictions using response times and a model of the decision process. *Journal of Economic Behavior & Organization* 148:344–375.
6. Krajbich I, Armel C, Rangel A (2010) Visual fixations and the computation and comparison of value in simple choice. *Nature neuroscience* 13(10):1292–1298.
7. Krajbich I, Rangel A (2011) 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.
8. Krajbich I, Lu D, Camerer C, Rangel A (2012) The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in psychology* 3:193.
9. Milosavljevic M, Malmaud J, Huth A, Koch C, Rangel A (2010) The drift diffusion model can account for value-based choice response times under high and low time pressure. *Judgement & Decision Making* 5:437–449.
10. Krajbich I, Bartling B, Hare T, Fehr E (2015) Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications* 6(7455):700.
11. Reutskaja E, Nagel R, Camerer CF, Rangel A (2011) Search dynamics in consumer choice under time pressure: An eye-tracking study. *The American Economic Review* 101(2):900–926.
12. Wald A (1947) *Sequential analysis.* (John Wiley & Sons).
13. Stone M (1960) Models for choice-reaction time. *Psychometrika* 25(3):251–260.
14. Edwards W (1965) Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology* 2(2):312–329.
15. Ratcliff R (1978) A theory of memory retrieval. *Psychological review* 85(2):59.
16. Milosavljevic M, Malmaud J, Huth A, Koch C, Rangel A (2010) The drift diffusion model can account for value-based choice response times under high and low time pressure. *Judgement & Decision Making* 5:437–449.
17. Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, Pouget A (2012) The cost of accumulating evidence in perceptual decision making. *The Journal of Neuroscience* 32(11):3612–3628.
18. Fudenberg D, Strack P, Strzalecki T (2018) Speed, accuracy, and the optimal timing of choices. *American Economic Review* 108(12):3651–84.
19. Tajima S, Drugowitsch J, Patel N, Pouget A (2019) Optimal policy for multi-alternative decisions. *Nature neuroscience* 22(9):1503–1511.
20. Hawkins GE, Forstmann BU, Wagenmakers EJ, Ratcliff R, Brown SD (2015) Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience* 35(6):2476–2484.
21. Chiong K, Shum M, Webb R, Chen R (2018) Split-second decision-making in the field: Response times in mobile advertising. *Available at SSRN.*
22. Ratcliff R (2002) 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.
23. Baldassi C, Cerreia-Vioglio S, Maccheroni F, Marinacci M (2018) An axiomatization of the drift diffusion model and its extension to multi-alternative choice. *mimeo.*
24. Busemeyer JR, Townsend JT (1992) Fundamental derivations from decision field theory. *Mathematical Social Sciences* 23(3):255–282.
25. Busemeyer JR, Townsend JT (1993) Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review* 100(3):432.
26. Busemeyer JR, Johnson JG (2004) Computational models of decision making. *Blackwell handbook of judgment and decision making* pp. 133–154.
27. Alós-Ferrer C, Fehr E, Netzer N (2018) Time will tell: recovering preferences when choices are noisy. *University of Zurich, Department of Economics, Working Paper* (306).
28. Echenique F, Saito K (2017) Response time and utility. *Journal of Economic Behavior & Organization* 139:49–59.
29. Hebert B, Woodford M (2019) Rational inattention when decisions take time. *mimeo.*
30. Woodford M (2014) An optimizing neuroeconomic model of discrete choice. *Columbia University working paper.*
31. Che YK, Mierendorff K (2019) Optimal dynamic allocation of attention. *American Economic Review* 109(8):2993–3029.
32. Liang A, Mu X, Syrgkanis V (2012) Dynamically aggregating diverse information. *mimeo.*
33. Liang A, Mu X (2020) Complementary information and learning traps. *The Quarterly Journal of Economics* 135(1):389–448.
34. Zhong W (2019) Optimal dynamic information acquisition. *mimeo.*
35. Aczél J (1966) *Lectures on functional equations and their applications.* (Academic press) Vol. 19.
36. Natenzon P (2019) Random choice and learning. *Journal of Political Economy* 127(1):419–457.
37. McFadden D (1989) A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society* pp. 995–1026.
38. Pakes A, Pollard D (1989) Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society* pp. 1027–1057.
39. Newey WK (1994) The asymptotic variance of semiparametric estimators. *Econometrica* pp. 1349–1382.
40. De Jong R, Bierens HJ (1994) On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity. *Econometric Theory* 10(1):70–90.
41. Hong Y, White H (1995) Consistent specification testing via nonparametric series regression. *Econometrica: Journal of the Econometric Society* pp. 1133–1159.
42. Chernozhukov V, Newey WK, Santos A (2015) Constrained conditional moment restriction models. *arXiv preprint arXiv:1509.06311.*
43. Zucca C, Sacerdote L (2009) On the inverse first-passage-time problem for a wiener process. *The Annals of Applied Probability* pp. 1319–1346.
44. Buonocore A, Giorno V, Nobile A, Ricciardi L (1990) On the two-boundary first-crossing-time problem for diffusion processes. *Journal of Applied Probability* 27(1):102–114.