Learning Treatment Effects under Covariate Dependent Left Truncation and Right Censoring

Yuyao Wang

Department of Mathematics, UC San Diego
yuw079@ucsd.edu

Joint work with:

Andrew Ying, Google

Ronghui (Lily) Xu, UC San Diego

Left truncation – selection due to delayed entry

- Example: aging studies.
  - Age is the time scale of interest.
  - Subjects enrolled at various ages instead of at the time origin (time at birth).

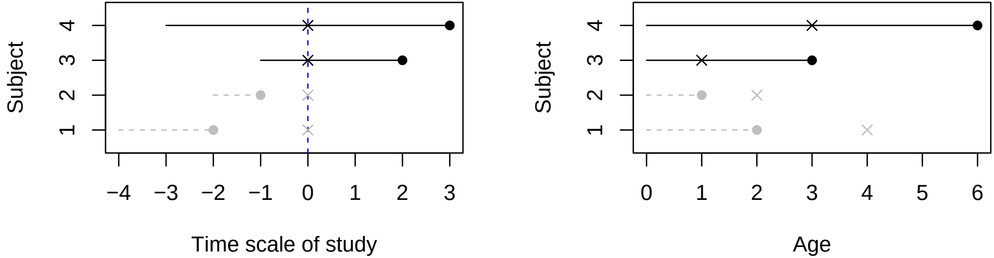


Figure: A toy example for aging study; '$\times$' - enrollment times; dots - times to events.

Left truncation – mathematical formulation

- Time-to-event: T^*
- Left truncation time: Q^* – usually the study enrollment time

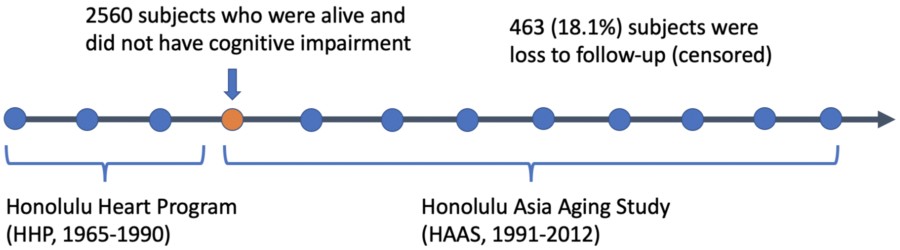
- T^* is **left truncated** by Q^* if only subjects with $T^* > Q^*$ are included in the data.

- Subjects with early event times tend not to be captured $\rightarrow$ **selection bias**

Examples:

- Aging studies – age is the time scale of interest
- Pregnancy studies
- Some cancer survivorship studies, e.g., SJLIFE.

HAAS data – Triple biases



- T^* - age to moderate cognitive impairment or death
  Q^* - age at entry of HAAS
  Only subjects with $T^* > Q^*$ are included.

- **Triple biases**:
  - **Selection bias from left truncation** – early event times are underrepresented.
  - Confounding in observational data.
  - Informative right censoring.

First focus on handling left truncation

Literature for handling left truncation

Under random left truncation / quasi-independence assumption

- Likelihood-based approaches.

! Strong assumption – may be violated in practice.

e.g., In HAAS data,

conditional Kendall's tau test for quasi-independence (Tsai, 1990): p-value = 0.0014.

Literature for handling left truncation

Under random left truncation / quasi-independence assumption

- Likelihood-based approaches.

! Strong assumption – may be violated in practice.

e.g., In HAAS data,

conditional Kendall's tau test for quasi-independence (Tsai, 1990): p-value = 0.0014.

Under covariate dependent left truncation

- In regression settings: Cox model with risk set adjustment.
- For marginal survival probability: inverse probability weighting (IPW) (Vakulenko-Lagun et al., 2022).

! Sensitive to misspecification of the truncation model; inefficient.

Our contributions

- Derive the efficient influence curve (EIC) for the expectation of a transformed event time.
- Construct EIC-based estimators that are shown to have favorable properties.
  - Model double robustness
  - Rate double robustness
  - Semiparametric efficiency
- Provide technical conditions for the asymptotic properties that appear to not have been carefully examined in the literature for time-to-event data.
- Our work represents the **first attempt** to construct doubly robust estimators in the presence of left truncation.
  - Does NOT fall under the established framework of coarsened data where doubly robust approaches were developed.

Y. Wang, A. Ying, and R. Xu (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.

Notation and estimand

- Q - left truncation time; T - event time; Z - covariates
- Variables with '$*$' – full data, i.e., if there were no left truncation;
- Variables without '$*$' – observed data: only contains subjects with $Q^* < T^*$.
  $O = (Q, T, Z)$.

Y. Wang, A. Ying, and R. Xu (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.

Notation and estimand

- Q - left truncation time; T - event time; Z - covariates
- Variables with '$*$' – full data, i.e., if there were no left truncation;
- Variables without '$*$' – observed data: only contains subjects with $Q^* < T^*$.
  $O = (Q, T, Z)$.

- Estimand: $\theta := \mathbb{E}\{\nu(T^*)\}$, where ν is a given function.
  - e.g., $\nu(t) = \mathbb{1}(t > t_0) \implies \theta = \mathbb{P}(T^* > t_0)$ (survival probability).
  - e.g., $\nu(t) = \min(t, t_0) \implies \theta = \mathbb{E}\{\min(T^*, t_0)\}$ (restricted mean survival time, RMST).

Y. Wang, A. Ying, and R. Xu (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.

Assumptions

1. **Conditional quasi-independence**:
   Q^* and T^* are conditionally independent given Z^* on the observed region $\{t > q\}$.
   - The dependence of Q^* and T^* can be explained by measured covariates.

2. **Positivity**: $\mathbb{P}(Q^* < T^* \mid T^*, Z^*) > 0$ a.s..

3. **Overlap**: There exist $0 < \tau_1 < \tau_2 < \infty$ and constants $\delta_1, \delta_2 > 0$ such that $T \geq \tau_1$ a.s. and $Q \leq \tau_2$ a.s. in the full data; $1 - F(\tau_2|Z) \geq \delta_1$ a.s., and $G(\tau_1|Z) \geq \delta_2$ a.s..

Y. Wang, A. Ying, and R. Xu (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.

Efficient influence curve and double robustness

- Efficient influence curve: $\varphi(O)$ = a constant factor times

$$U(\theta; F, G) = \underbrace{\frac{\nu(T) - \theta}{G(T|Z)}}_{\text{IPW}} - \underbrace{\int_0^\infty m_\nu(v, Z; F) \cdot \frac{F(v|Z)}{1 - F(v|Z)} \cdot \frac{d\bar{M}_Q(v; G)}{G(v|Z)}}_{\text{Augmentation}}$$

- F: the conditional CDF of $T^* \mid Z^*$.
- G: the conditional CDF of $Q^* \mid Z^*$.

(both can be nonparametrically identified from the observed data distribution)

- $m_\nu(v, z; F) = \mathbb{E}_F\{\nu(T^*) - \theta \mid T^* < v, Z^* = z\}$.

- The semiparametric efficiency bound : $\mathbb{E}(\varphi^2)$.

Efficient influence curve and double robustness

- Efficient influence curve: $\varphi(O)$ = a constant factor times

$$U(\theta; F, G) = \underbrace{\frac{\nu(T) - \theta}{G(T|Z)}}_{\text{IPW}} - \underbrace{\int_0^\infty m_\nu(v, Z; F) \cdot \frac{F(v|Z)}{1 - F(v|Z)} \cdot \frac{d\bar{M}_Q(v; G)}{G(v|Z)}}_{\text{Augmentation}}$$

- F: the conditional CDF of $T^* \mid Z^*$.
- G: the conditional CDF of $Q^* \mid Z^*$.

(both can be nonparametrically identified from the observed data distribution)

- $m_\nu(v, z; F) = \mathbb{E}_F\{\nu(T^*) - \theta \mid T^* < v, Z^* = z\}$.

- The semiparametric efficiency bound : $\mathbb{E}(\varphi^2)$.

Double robustness:

$\mathbb{E}\{U(\theta_0; F, G)\} = 0$ if either $F = F_0$ or $G = G_0$.

Estimation: model double robustness under asymptotic linearity

Let $\{O_i\}_{i=1}^n$ be an observed random sample of size n; $\quad O_i = (Q_i, T_i, Z_i)$.

- First estimate F and G
- Then solve $\sum_{i=1}^n U_i(\theta; \hat{F}, \hat{G}) = 0$ for $\theta \implies \hat{\theta}_{dr}$

Y. Wang, A. Ying, and R. Xu (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.

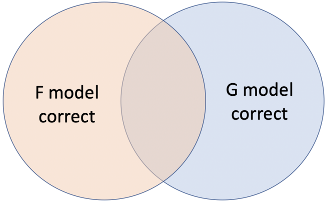
Estimation: model double robustness under asymptotic linearity

Let $\{O_i\}_{i=1}^n$ be an observed random sample of size n; $O_i = (Q_i, T_i, Z_i)$.

- First estimate F and G
- Then solve $\sum_{i=1}^n U_i(\theta; \hat{F}, \hat{G}) = 0$ for $\theta \implies \hat{\theta}_{dr}$

When (semi-)parametric models are used,

- $\hat{\theta}_{dr}$ is CAN if **either** the model for F or G is correctly specified.



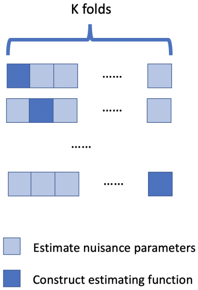
Furthermore, when both models are correctly specified,

- $\hat{\theta}_{dr}$ achieves the semiparametric efficiency bound;
- Consistent estimator for the asymptotic variance.

Y. Wang, A. Ying, and R. Xu (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.

Estimation: Rate double robustness with cross-fitting

K-fold cross-fitting



$$\Longrightarrow \hat{\theta}_{cf}$$

Y. Wang, A. Ying, and R. Xu (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.

Estimation: Rate double robustness with cross-fitting

Suppose

- the nuisance estimators are uniformly consistent;
- the integral product rate $\mathcal{D}(\hat{F}, \hat{G}; F_0, G_0) = o_p(n^{-1/2})$.

$$\mathcal{D}(F, G; F_0, G_0) := \mathbb{E}\left[\left|\int_{\tau_1}^{\tau_2} \{a(t, Z; F) - a(t, Z; F_0)\} \cdot \mathbb{1}(Q \leq t < T) \cdot d\left\{\frac{1}{G(t|Z)} - \frac{1}{G_0(t|Z)}\right\}\right|\right],$$

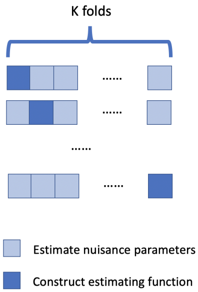
$$a(t, Z; F) = \int_0^t \{\nu(u) - \theta\} dF(u|Z)/\{1 - F(t|Z)\}.$$

Then

- $\hat{\theta}_{cf}$ is CAN; achieves the semiparametric efficiency bound;
- Consistent estimator for the asymptotic variance.

Nonparametric methods can be used to estimate F and G!

K-fold cross-fitting



$\Longrightarrow \hat{\theta}_{cf}$

Y. Wang, A. Ying, and R. Xu (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.

Literature for treatment effect estimation

For handling confounding bias

- Average treatment effect (ATE) $\leftarrow$ Augmented IPTW (AIPTW)
- Conditional average treatment effect (CATE) $\leftarrow$ Orthogonal learners (Foster and Syrgkanis, 2023)
  - e.g., R-learner (Nie and Wager, 2021); DR-learner (Kennedy, 2023).

Literature for treatment effect estimation

For handling confounding bias

- Average treatment effect (ATE) $\leftarrow$ Augmented IPTW (AIPTW)
- Conditional average treatment effect (CATE) $\leftarrow$ Orthogonal learners (Foster and Syrgkanis, 2023)
  - e.g., R-learner (Nie and Wager, 2021); DR-learner (Kennedy, 2023).

For right censored time-to-event data: confounding + censoring bias

- Doubly robust estimators for ATE (Zhang and Schaubel, 2012 ; Bai et al., 2017; Sjolander and Vansteelandt, 2017; Westling et al., 2023; Luo et al., 2023)
- Orthogonal learners for CATE (Morzywolek et al., 2023; Xu et al., 2024)

Literature for treatment effect estimation

For handling confounding bias

- Average treatment effect (ATE) $\leftarrow$ Augmented IPTW (AIPTW)
- Conditional average treatment effect (CATE) $\leftarrow$ Orthogonal learners (Foster and Syrgkanis, 2023)
  - e.g., R-learner (Nie and Wager, 2021); DR-learner (Kennedy, 2023).

For right censored time-to-event data: confounding + censoring bias

- Doubly robust estimators for ATE (Zhang and Schaubel, 2012 ; Bai et al., 2017; Sjolander and Vansteelandt, 2017; Westling et al., 2023; Luo et al., 2023)
- Orthogonal learners for CATE (Morzywolek et al., 2023; Xu et al., 2024)

For handling all three biases

- Methods based on regression models (Cheng and Wang, 2012; Cheng and Wang, 2015).
- IPW-based approaches

! Sensitive to model misspecifications; inefficient.

Our contributions

- Develop a general doubly robust framework for handling covariate dependent left truncation and right censoring.

- Construct model doubly robust and rate doubly robust estimators for ATE.
- Construct orthogonal and doubly robust learners for CATE that are shown to achieve oracle rate.

- Our work represents the **first attempt** to
  - develop doubly robust approaches that address all three sources of biases;
  - investigate CATE estimation for left truncated and right censored data.

Y. Wang, A. Ying, and R. Xu, (2024+). Learning treatment effects under covariate dependent left truncation and right censoring. *(Work in progress)*.

Notation

- Q: left truncation time; T: event time; Z: covariates.
- C: censoring time; $D = C - Q$: residual censoring time;
- A: binary treatment assignment.

- Variables with '$*$' – truncation-free data; without '$*$' – truncated data.
- $T^*(a)$ – potential event time under treatment a.

- Observe $O = (Q, X, \Delta, A, Z)$ only if $Q^* < T^*$.
- $X = \min(T, C)$, $\Delta = I(T < C)$.

Y. Wang, A. Ying, and R. Xu, (2024+). Learning treatment effects under covariate dependent left truncation and right censoring. *(Work in progress)*.

Assumptions

1. No unmeasured confounding, consistency.
2. Conditional independent truncation: $Q^* \perp\!\!\!\perp T^* \mid A^*, Z^*$.
3. Conditional noninformative residual censoring: $D \perp\!\!\!\perp (T - Q) \mid Q, A, Z$.
4. Positivity.
5. Overlap.

Doubly robust operator for left truncation

- $\zeta(T^*, A^*, Z^*; \theta)$ – an unbiased and non-degenerate estimating function for θ.
- The EIC for θ in truncated data: a constant factor times

$$\underbrace{\frac{\zeta(T, A, Z; \theta)}{G(T|A, Z)}}_{\text{IPW}} - \underbrace{\int_0^\infty m_\zeta(v, A, Z; \theta, F) \cdot \frac{F(v|A, Z)}{1 - F(v|A, Z)} \cdot \frac{d\bar{M}_Q(v; G)}{G(v|A, Z)}}_{\text{Augmentation}}$$

- F: the conditional CDF of $T^* \mid A^*, Z^*$.
- G: the conditional CDF of $Q^* \mid A^*, Z^*$.

- $m_\zeta(v, a, z; \theta, F) = \mathbb{E}\{\zeta(T^*, A^*, Z^*; \theta) \mid T^* \leq v, A^* = a, Z^* = z\}$.

- **truncAIPW**: generalize to any function ζ

Y. Wang, A. Ying, and R. Xu, (2024+). Learning treatment effects under covariate dependent left truncation and right censoring. *(Work in progress)*.

Doubly robust operator for censoring

- Adapt the **AIPCW** (Rotnitzky and Robins, 2005) to the **residual time scale**.
- For any function $\xi(Q, T, A, Z)$ in censoring-free data,

$$\underbrace{\frac{\Delta\, \xi(Q, X, A, Z)}{S_D(X - Q | Q, A, Z)}}_{\text{IPCW}} + \underbrace{\int_0^{\infty} \mathbb{E}_F\{\xi(Q, T, A, Z) | T - Q \geq u, Q, A, Z\} \cdot \frac{d M_D(u; S_D)}{S_D(u | Q, A, Z)}}_{\text{Augmentation}}$$

- Recall $D = C - Q$: the residual censoring time.

- S_D: the conditional survival function of $D \mid Q, A, Z$.
- Recall F: the conditional CDF of $T^* \mid A^*, Z^*$.

General framework and double robustness

- **General framework**:

Full data

↓ AIPTW(μ,π), orthogonal loss

Trunc-free & cen-free data

↓ truncAIPW(F,G)

Truncated & cen-free data

↓ AIPCW(F,S_D)

Observed data

- **Double robustness**: The expectation is maintained (up to a constant factor) if either F or (G, S_D) is the truth.

Y. Wang, A. Ying, and R. Xu, (2024+). Learning treatment effects under covariate dependent left truncation and right censoring. *(Work in progress)*.

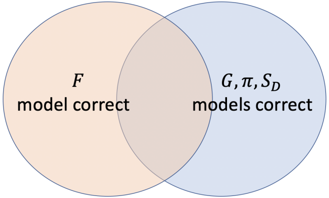
ATE estimation

- Estimand: $\theta_a = \mathbb{E}[\nu\{T^*(a)\}]$. Propensity score: $\pi(z) = \mathbb{P}(A^* = 1 | Z^* = z)$.
- AIPTW estimating function $\xrightarrow{\text{truncAIPW, AIPCW}}$ $U_a(\theta_a; F, G, \pi, S_D)$
- **Double robustness**: $\mathbb{E}\{U_a(\theta_a; F, G, \pi, S_D)\} = 0$ if either F or (G, π, S_D) is the truth.

ATE estimation

- Estimand: $\theta_a = \mathbb{E}[\nu\{T^*(a)\}]$. Propensity score: $\pi(z) = \mathbb{P}(A^* = 1 | Z^* = z)$.
- AIPTW estimating function $\xrightarrow{\text{truncAIPW, AIPCW}}$ $U_a(\theta_a; F, G, \pi, S_D)$
- **Double robustness**: $\mathbb{E}\{U_a(\theta_a; F, G, \pi, S_D)\} = 0$ if either F or (G, π, S_D) is the truth.
- Estimation:

Model double robustness.



Rate double robustness.

- Require the **product error rate** between $\hat{F}$ and $(\hat{G}, \hat{\pi}, \hat{S}_D)$ to be faster than $n^{-1/2}$.

Y. Wang, A. Ying, and R. Xu, (2024+). Learning treatment effects under covariate dependent left truncation and right censoring. *(Work in progress)*.

CATE estimation

- Estimand: $\tau(z) = \mathbb{E}[\nu\{T^*(1)\} - \nu\{T^*(0)\} \mid Z^* = z]$.

- Loss function $\ell(T^*, A^*, Z^*; \tau, F, \pi) \xrightarrow{\text{truncAIPW, AIPCW}} \tilde{\ell}(O; \tau, F, G, \pi, S_D)$
  - R-loss $\rightarrow$ ltrcR-loss
  - DR-loss $\rightarrow$ ltrcDR-loss

- **Neyman orthogonality** and **double robustness** are maintained.

CATE estimation

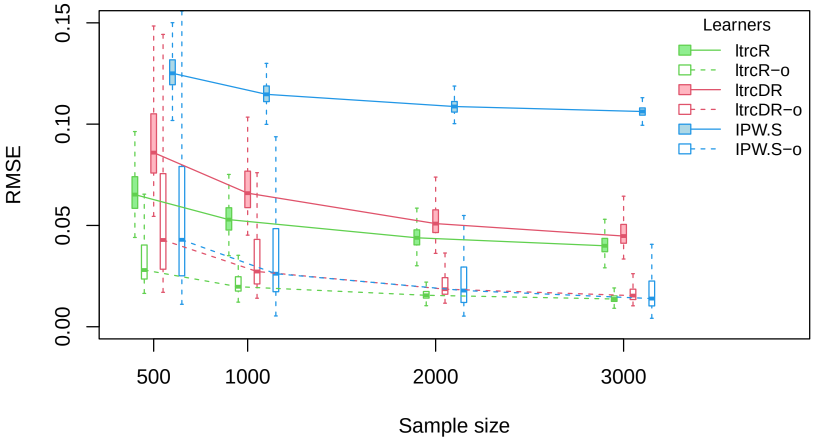
- Estimand: $\tau(z) = \mathbb{E}[\nu\{T^*(1)\} - \nu\{T^*(0)\} \mid Z^* = z]$.
- Loss function $\ell(T^*, A^*, Z^*; \tau, F, \pi) \xrightarrow{\text{truncAIPW, AIPCW}} \tilde{\ell}(O; \tau, F, G, \pi, S_D)$
  - R-loss $\rightarrow$ ltrcR-loss
  - DR-loss $\rightarrow$ ltrcDR-loss
- **Neyman orthogonality** and **double robustness** are maintained.
- $\hat{\tau} \quad \leftarrow \quad K$-fold cross-fitted empirical risk minimization.

CATE estimation

- Estimand: $\tau(z) = \mathbb{E}[\nu\{T^*(1)\} - \nu\{T^*(0)\} \mid Z^* = z]$.
- Loss function $\ell(T^*, A^*, Z^*; \tau, F, \pi) \xrightarrow{\text{truncAIPW, AIPCW}} \tilde{\ell}(O; \tau, F, G, \pi, S_D)$
  - R-loss $\rightarrow$ ltrcR-loss
  - DR-loss $\rightarrow$ ltrcDR-loss
- **Neyman orthogonality** and **double robustness** are maintained.
- $\hat{\tau} \quad \leftarrow \quad K$-fold cross-fitted empirical risk minimization.
- Estimation errors of the nuisance parameter only have **higher order impact** on $\hat{\tau} - \tau_0$.
  - If the nuisance are estimated at faster than $n^{-1/4}$ rate
    $\rightarrow \hat{\tau}$ achieve **oracle rate**, i.e., estimation error rate if the nuisance were known.

Simulation - CATE

- 500 simulated data sets. Estimand: $\tau(z) = \mathbb{E}\{\log T^*(1) - \log T^*(0) \mid Z^* = z\}$.
- Truncation rate: 31.3%; treatment rate: 50%; censoring rate 48.4%. $RMSE^2 = \frac{1}{n}\sum_{i=1}^n \{\hat{\tau}(V_i) - \tau_0(V_i)\}^2$.



Y. Wang, A. Ying, and R. Xu, (2024+). Learning treatment effects under covariate dependent left truncation and right censoring. *(Work in progress)*.

Summary:

- We have developed a general doubly robust framework for handling covariate dependent left truncation and right censoring.
  - e.g., estimate ATE, CATE.
  - Can also be applied to estimate parameters in marginal structure models.

Acknowledgment:

- We thank Dr. Kendrick Li on the helpful discussion on EIC derivation.

References and contact info:

- **Wang, Y.**, Ying, A. and Xu, R. (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.
- **Wang, Y.**, Ying, A. and Xu, R. (2024+). Learning treatment effects under covariate dependent left truncation and right censoring. *(Work in progress)*.
- R-package: `truncAIPW`. Email: yuw079@ucsd.edu

I'm on the job market for positions starting 2025 Fall!

Appendix

Assumptions

f, g, h: the densities of $T^*|Z^*$, $Q^*|Z^*$ and Z^*, respectively.

- **Conditional quasi-independence**: The observed data density for (Q, T, Z) satisfies

$$p_{Q,T,Z}(q,t,z) = \begin{cases} f(t|z)g(q|z)h(z)/\beta, & \text{if } t > q, \\ 0, & \text{otherwise}, \end{cases}$$

  where $\beta = \mathbb{P}(Q^* < T^*) = \int \mathbb{1}(q < t)f(t|z)g(q|z)h(z)\; dt\; dq\; dz$.

- **Positivity**: $G(T^*|Z^*) > 0$ a.s.

- **Overlap**: There exists $0 < \tau_1 < \tau_2 < \infty$ such that $T^* \geq \tau_1$ a.s., $Q^* \leq \tau_2$ a.s.; also $G(\tau_1|Z^*) \geq \delta_1$ a.s. and $F(\tau_2|Z^*) \leq 1 - \delta_2$ a.s. for some constants $\delta_1 > 0$ and $\delta_2 > 0$.