Chapter 5: Duration Models

Joan Llull

Microeconometrics
IDEA Phd Program

Introduction

Duration analysis

Duration data: **how long** has an individual been in a state **when exiting** from it (e.g. weeks unemployed).

Examples: unemployment duration, marriage duration, life expectancy, firms exit from the market,...

We want to analyze:

- Why **durations differ** across individuals?
- How and why do exit probabilities **vary over time**?

Long tradition in **biometrics**: surviving probabilities, hazard functions,...

Duration data

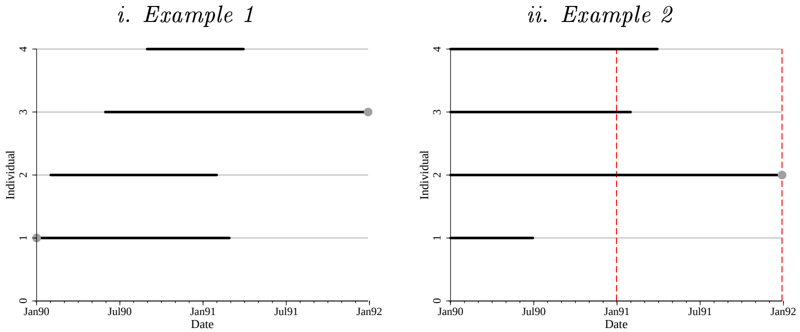
We **denote** these durations for the N observations as $t_1, t_2, ..., t_N$.

These data are typically **censored**. E.g.:

- Individuals may not find a job before the interview (we observe that $t > \bar{t}$, but not t).
- Individuals may find a job between selection and interview, and we cannot interview them (we observe $\underline{t} < t < \bar{t}$)

One of the main motivations for duration analysis: **dealing with censoring**.

Figure I. – Two examples of censored observations



Note: Black lines represent the time when the individual was unemployed. A dot indicates that the individual is still unemployed at that date, but we do not have further information about him/her. Vertical red dashed lines in Example 2 are interview dates.

The Hazard Function

The hazard function

Hazard function: probability (or density) of exiting at t conditional on being alive.

It can be **time-varying** (e.g. mortality) or **constant**.

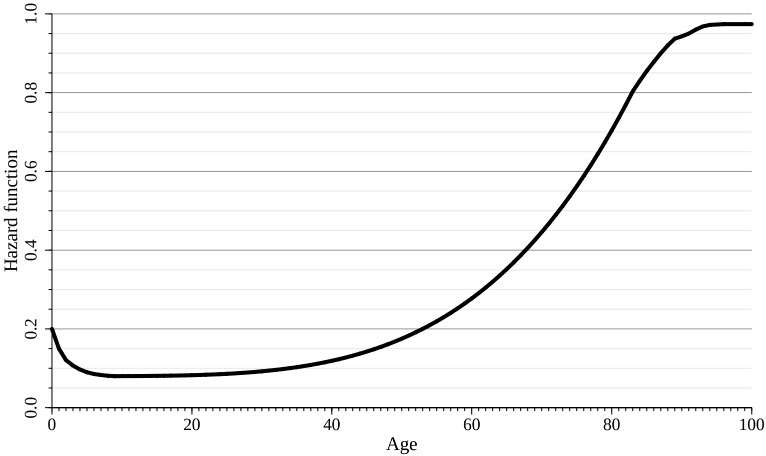
We analyze **discrete** and **continuous** hazard functions.

Why do we **care**?

- Theoretically appealing (e.g. job arrival rate).
- Empirically convenient (binomial discrete choice+censoring).

We start from **unconditional** hazards and then add regressors.

Figure II. – Mortality Hazard Rate



Note: The line depicts the hazard mortality rate, i.e. the probability of dying at age a conditional on survival until that age.

Hazard function for a discrete variable

Probability mass function for t: $p(\tau) = \Pr(t = \tau)$ for $\tau = 1, 2, ...$

Cdf for t: $F(t) = p(1) + p(2) + ... + p(t)$.

Hazard function for t:

$$h(\tau) = \Pr(t = \tau | t \geq \tau) = \frac{\Pr(t = \tau)}{\Pr(t \geq \tau)} = \frac{p(\tau)}{1 - F(\tau - 1)} = \frac{F(\tau) - F(\tau - 1)}{1 - F(\tau - 1)}.$$

We can **recover** $p(t)$ and $F(t)$ from $h(t)$ recursively:

$$1 - h(t) = \frac{1 - F(t)}{1 - F(t - 1)} \ \Rightarrow\ F(t) = 1 - \prod_{s=1}^{t} (1 - h(s)),$$

$$p(t) = (1 - F(t - 1))h(t) = h(t) \prod_{s=1}^{t-1} (1 - h(s)).$$

(Interpretation of these expressions)

Hazard function for a continuous variable

Pdf function for t: $f(\tau) = \lim_{dt \to 0} \frac{\Pr(\tau \leq t < \tau + dt)}{dt}$.

Cdf for t: $F(t) = \int_0^t f(u)du$.

Hazard function for t:

$$h(\tau) = \lim_{dt \to 0} \frac{\Pr(\tau \leq t < \tau + dt | t \geq \tau)}{dt} = \lim_{dt \to 0} \frac{\Pr(\tau \leq t < \tau + dt)}{\Pr(t \geq \tau)} \Big/ dt = \frac{f(\tau)}{1 - F(\tau)}.$$

We can **recover** $f(t)$ and $F(t)$ with the integrated hazard $H(t) = \int_0^t h(s)ds$:

$$H(t) = -\ln[1 - F(t)] \;\Rightarrow\; F(t) = 1 - \exp[-H(t)],$$

$$f(t) = \frac{\partial F(t)}{\partial t} = h(t)\exp[-H(t)] = h(t)(1 - F(t)).$$

(Comparison with **discrete** through $\ln[1 - F(t)]$)

Some frequently used hazard functions

In discrete duration, we can estimate **semi-parametric** hazards.

In continuous duration (and sometimes in discrete), we usually assume a given **parametric function**.

Two frequent examples:

- **Constant** hazard.
- **Weibull** distribution.

Constant hazard: the exponential distribution

The simplest hazard function is a **constant** hazard: $h(t) = \lambda$.

In the **continuous** case:

$$H(t) = \int_0^t \lambda du = \lambda t,$$

$$F(t) = 1 - e^{-\lambda t},$$

$$f(t) = \lambda e^{-\lambda t}.$$

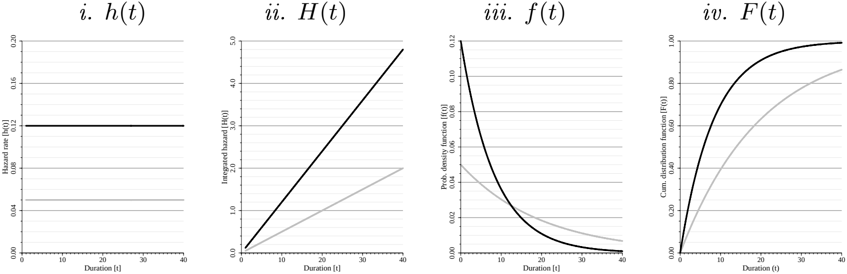
In the **discrete** case:

$$F(t) = 1 - (1 - \lambda)^t,$$

$$p(t) = \lambda(1 - \lambda)^{t-1}.$$

Having a constant hazard is the **memoryless property** of the exponential. Also $\mathbb{E}[T] = 1/\lambda$.

Figure III. – Examples of Hazard Functions (I): Constant



Note: Black: $\lambda = 0.12$; gray: $\lambda = 0.05$. These examples depict the hazard function (left), the integrated hazard (center-left), the pdf or probability mass function (center-right) and the cdf (right) of a constant hazard model with hazard equal to λ

The Weibull distribution

A **generalization** of the exponential distribution is:

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}},$$

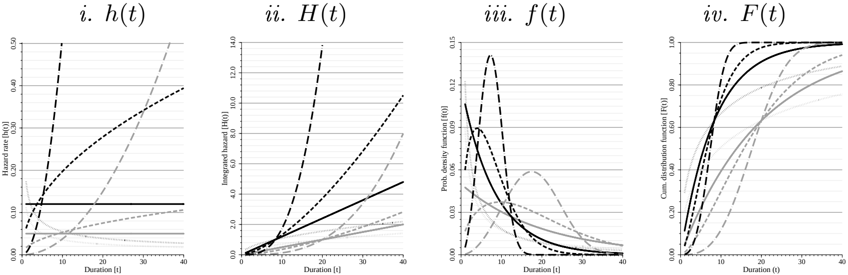
$$f(t) = \alpha \lambda^{\alpha} t^{\alpha - 1} e^{-(\lambda t)^{\alpha}}.$$

This gives the following **hazard** function:

$$h(t) = \alpha \lambda^{\alpha} t^{\alpha - 1}.$$

It is monotonically **increasing, decreasing, or constant** depending on whether $\alpha >\!<\!= 1$

FIGURE IV. – EXAMPLES OF HAZARD FUNCTIONS (II): WEIBULL



Note: Black: $\lambda = 0.12$; gray: $\lambda = 0.05$. Solid: $\alpha = 1$; dotted: $\alpha = 0.5$; dashed: $\alpha = 1.5$; dot-dashed: $\alpha = 3$. These examples depict the hazard function (left), the integrated hazard (center-left), the pdf or probability mass function (center-right) and the cdf (right) of a Weibull hazard model with parameters λ and α.

Conditional Hazard Functions

Including covariates to the hazard model

We want to see how duration is affected by **covariates $\boldsymbol{x}$**.

We are interested in the **conditional hazard**:

$$h(t, \boldsymbol{x}) = \frac{f(t|\boldsymbol{x})}{1 - F(t|\boldsymbol{x})}.$$

Proportional hazard model

It was introduced by **Cox** (1972). It is a.k.a. Cox regression or Cox model.

Main advantage: **simplicity**.

It is a **factorized** model:

$$h(t, \boldsymbol{x}) = \lambda(t) \exp(\boldsymbol{x}'\boldsymbol{\beta}).$$

$\lambda(t)$ can be either **constant** $\lambda(t) = 1$ or **Weibull** $\lambda(t) = \alpha t^{\alpha-1}$ (the scale cannot be separately identified from the constant)

$\lambda(t)$ is called the **baseline hazard**.

Discrete durations

Now conditional **hazard rates** are probabilities.

One approach: treat spells as **continuous** and use the PH model.

However, it is natural to use discrete choice models to specify the exit rate:

$$h(\tau, \boldsymbol{x}) = \Pr(t = \tau | t \geq \tau, \boldsymbol{x}) = G(\gamma_\tau + \boldsymbol{x}'\boldsymbol{\beta}_\tau),$$

where $G(.)$ is a cdf (e.g. normal or logistic).

Different ways to specify baseline hazard:

- Leave them **free**: $\gamma_t = \sum_{j=1}^{T^*} \gamma_j \, \mathbb{1}\{t = j\}$. (why T^*?)
- Specify them as **polynomials**: e.g. $\gamma_t = \gamma_0 + \gamma_1 \ln t + \gamma_2 (\ln t)^2$.
- ...

Assuming γ_t free delivers a **semi-parametric** estimation of $\lambda(t)$ that allows us to test parametric assumptions.

Likelihood functions

Complete durations

Assume that **we observe** $\{t_1, t_2, ..., t_N\}$

The **log-likelihood** function is:

$$\mathcal{L}_{\mathrm{N}} = \sum_{\mathrm{i=1}}^{\mathrm{N}} \ln f(t_i|\boldsymbol{x}_i),$$

where:

$$f(t|\boldsymbol{x}) = h(t, \boldsymbol{x}) \exp(-H(t, \boldsymbol{x})) = \lambda(t) \exp(\boldsymbol{x}'\boldsymbol{\beta}) \exp\left\{-\Lambda(t) \exp(\boldsymbol{x}'\boldsymbol{\beta})\right\}.$$

Different expressions for $\lambda(t)$ and $\Lambda(t)$ depending on **baseline hazards**.

Our two examples of censored durations

Example 1:

We observe t if $t < \bar{t}$; otherwise we only observe that $t > \bar{t}$

The **log-likelihood** function is:

$$\mathcal{L}_{\mathrm{N}} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \left\{ w_i \ln f(t_i|\boldsymbol{x}_i) + (1-w_i) \ln(1 - F(\bar{t}_i|\boldsymbol{x}_i)) \right\}, \quad w_i = \mathbb{1}\{t_i < \bar{t}_i\}.$$

Example 2:

We observe whether whether $t < \bar{t}^1$, $\bar{t}^1 < t < \bar{t}^2$, or $t > \bar{t}^2$.

The **log-likelihood** function is:

$$\begin{aligned}\mathcal{L}_{\mathrm{N}} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \Big\{ w_i^1 \ln F(\bar{t}_i^1|\boldsymbol{x}_i) + w_i^2 \ln \Big(F(\bar{t}_i^2|\boldsymbol{x}_i) \; & - F(\bar{t}_i^1|\boldsymbol{x}_i) \Big) \\ & + (1 - w_i^1 - w_i^2) \ln \Big(1 - F(\bar{t}_i^2|\boldsymbol{x}_i) \Big) \Big\},\end{aligned}$$

where $w_i^1 = \mathbb{1}\{t_i < \bar{t}_i^1\}$, and $w_i^2 = \mathbb{1}\{\bar{t}_i^1 < t_i < \bar{t}_i^2\}$.

An additional example

Example 2 modified:

Same as example 2, but with initial (known) durations $d_1, d_2, ..., d_N$.

The **log-likelihood** function is:

$$\mathcal{L}_{\mathrm{N}} = \sum_{\mathrm{i=1}}^{\mathrm{N}} \left\{ w_i^1 \ln \frac{F(d_i + \bar{t}_i^1 | \boldsymbol{x}_i) - F(d_i | \boldsymbol{x}_i)}{1 - F(d_i | \boldsymbol{x}_i)} + w_i^2 \ln \frac{F(d_i + \bar{t}_i^2 | \boldsymbol{x}_i) - F(d_i + \bar{t}_i^1 | \boldsymbol{x}_i)}{1 - F(d_i | \boldsymbol{x}_i)} \right.$$
$$\left. + (1 - w_i^1 - w_i^2) \ln \frac{1 - F(d_i + \bar{t}_i^2 | \boldsymbol{x}_i)}{1 - F(d_i | \boldsymbol{x}_i)} \right\},$$

where $w_i^1 = \mathbb{1}\{d_i < t_i < d_i + \bar{t}_i^1\}$, and $w_i^2 = \mathbb{1}\{d_i + \bar{t}_i^1 < d_i + t_i < \bar{t}_i^2\}$.

Discrete durations

The **log-likelihood** of a discrete duration model can be written as:

$$\mathcal{L}_{\mathrm{N}} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{\tau=1}^{T^*} w_{i\tau} \{ y_{i\tau} \ln G(\gamma_\tau + \boldsymbol{x}_i' \boldsymbol{\beta}_\tau) + (1 - y_{i\tau}) \ln(1 - G(\gamma_\tau + \boldsymbol{x}_i' \boldsymbol{\beta}_\tau)) \}.$$

where $y_{i\tau} = \mathbb{1}\{t_i = \tau\}$, and $w_{i\tau} = \mathbb{1}\{t_i \geq \tau\}$. **Two types** of contributions:

- Spells that end at time τ:

$$\ln \Pr(t = \tau | \boldsymbol{x}) = \ln h(\tau, \boldsymbol{x}) + \sum_{s=1}^{\tau-1} \ln(1 - h(s, \boldsymbol{x})).$$

- Spells that are incomplete at time T^*:

$$\ln \Pr(t > T^* | \boldsymbol{x}) = \sum_{s=1}^{T^*} \ln(1 - h(s, \boldsymbol{x})).$$

This problem is a **sequence of binary choice models**.

If **censoring** occurs at **different durations**: $T_i^* = \min\{\bar{t}_i, T^*\}$.

Unobserved Heterogeneity

State dependence vs unobserved heterogeneity

Consider a regressor $x = \{0, 1\}$, and a **constant proportional hazard** model such that:

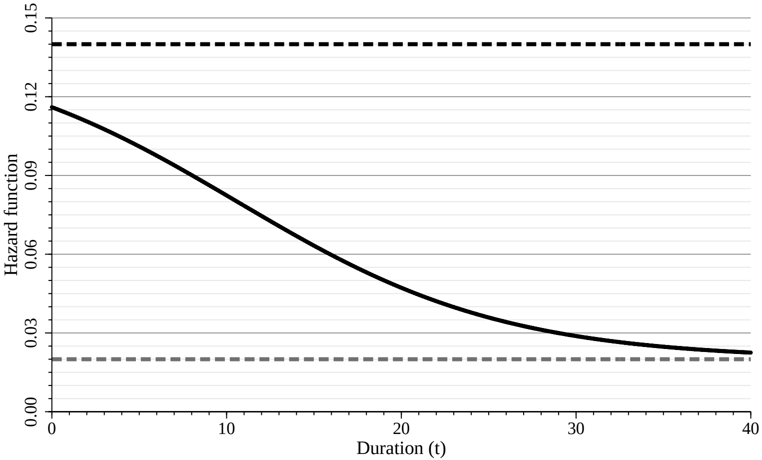
$$h(t, x = 0) = h_0 \qquad h(t, x = 1) = h_1 \qquad h_1 > h_0.$$

The **aggregate hazard** is:

$$h(\tau) = h_1 \Pr(x = 1 | t \geq \tau) + h_0 \Pr(x = 0 | t \geq \tau).$$

This aggregate hazard is **not constant** anymore given that $\Pr(x = 1 | t \geq \tau)$ decreases over time $\Rightarrow$ "spurious state dependence".

Figure V. – An Example with Unobserved Heterogeneity



Note: Black dashed: h_1 (hazard rate when $x = 1$); gray dashed: h_0 (hazard rate when $x = 0$); gray dashed: observed (unconditional) hazard.

The PH with unobserved heterogeneity

Lancaster (1979) includes unobserved heterogeneity as a **multiplicative random effect**:

$$h(t, \boldsymbol{x}, \nu) = \lambda(t) \exp(\boldsymbol{x}'\boldsymbol{\beta})\nu,$$

where ν is independent of $\boldsymbol{x}$ with $\mathbb{E}[\nu] = 1$ and with pdf $g(\nu)$ (e.g. gamma).

The cdf **conditional** on ν is:

$$G(t|\boldsymbol{x}, \nu) = 1 - \exp\left[-\int_0^t h(u, \boldsymbol{x}, \nu) du\right],$$

and the **unconditional** cdf is:

$$F(t|\boldsymbol{x}) = \int_0^\infty F(t|\boldsymbol{x}, v) g(v) dv,$$

Important remark: **identification** requires inclusion of exogenous regressors!

Unobserved heterogeneity with discrete data

Similarly to the continuous case:

$$\Pr(t = \tau|\boldsymbol{x}) = \int \Pr(t = \tau|\boldsymbol{x}, v)g(v)dv = \int h(\tau, \boldsymbol{x}, v) \prod_{s=1}^{\tau-1}[1-h(s, \boldsymbol{x}, v)]g(v)dv,$$

where, for instance:

$$h(t, \boldsymbol{x}, v) = G(\gamma_t + \boldsymbol{x}'\beta_t + v).$$

A **frequently used specification** for $g(v)$ is a discrete-support mass point distribution, i.e. $\{\xi_1, ..., \xi_m\}$ with probabilities $\{p_1, .., p_m\}$:

$$\Pr(t = \tau|\boldsymbol{x}) = \sum_{j=1}^{m} \left\{ h(\tau, \boldsymbol{x}, \xi_j) \prod_{s=1}^{\tau-1}[1 - h(s, \boldsymbol{x}, \xi_j)]p_j \right\},$$

and ξ_j and p_j are treated as **additional parameters**.

Multiple-exit Discrete Duration Models

Transition intensities vs conditional hazard rates

Let d_1 and d_2 be indicators for the **two potential alternatives** where to exit (e.g temporary or permanent job).

We can define the following **intensities of transition** to each state:

$$\phi_j(\tau) = \Pr(t = \tau, d_j = 1 | t \geq \tau), \quad j = 1, 2.$$

The **unconditional hazard rates** are:

$$h(\tau) = \Pr(t = \tau | t \geq \tau) = \phi_1(\tau) + \phi_2(\tau).$$

Also, we can define **conditional hazard rates** as:

$$h_j(\tau) = \Pr(y_{j\tau} = 1 | t \geq \tau, y_{k\tau} = 0), j \neq k \text{ where } y_{j\tau} = y_\tau \times d_j = \mathbb{1}\{t = \tau, d_j = 1\}.$$

The **mapping** between the two expressions is as follows:

$$h_j(\tau) = \frac{\Pr(y_{j\tau} = 1 | t \geq \tau)}{\Pr(y_{k\tau} = 0 | t \geq \tau)} = \frac{\phi_j(t)}{1 - \phi_k(t)}.$$

We can model **either of the two**. For instance, a MNL for ϕ's delivers a binary logit for h's with the same parameters.

Competing risk models

The models for conditional hazards $h_1(t)$ and $h_2(t)$ are often called **competing risk models**.

This name derives from considering two **latent variables** t_1^* and t_2^*, such that the observed duration is $t = \min\{t_1^*, t_2^*\}$.

If t_1^* and t_2^* are **independent**, $h_1(t)$ and $h_2(t)$ can be interpreted as **exit rates of latent durations**:

$$h_j(\tau) = \Pr(t_j^* = \tau | t_j^* \geq \tau).$$

This implies that when we analyze exits to alternative 1, we take exits to 2 as **censored observations**.

FIML estimation

The **log-likelihood** is analogous to the discrete case:

$$\mathcal{L}_{\mathrm{N}} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{\tau=1}^{T^*} w_{i\tau}\{y_{i\tau}(d_{1i} \ln \phi_1(\tau, \boldsymbol{x}_i) + d_{2i} \ln \phi_2(\tau, \boldsymbol{x}_i)) \\ +(1 - y_{i\tau}) \ln(1 - \phi_1(\tau, \boldsymbol{x}_i) - \phi_2(\tau, \boldsymbol{x}_i))\},$$

where $y_{i\tau} = \mathbb{1}\{t_i = \tau\}$, and $w_{i\tau} = \mathbb{1}\{t_i \geq \tau\}$.

Three types of **contributions**:

- Spells that end at time τ exiting to option 1:

$$\ln \Pr(t = \tau, d_1 = 1|\boldsymbol{x}) = \ln \phi_1(\tau, \boldsymbol{x}) + \sum_{s=1}^{\tau-1} \ln(1 - \phi_1(s, \boldsymbol{x}) - \phi_2(s, \boldsymbol{x})).$$

- Spells that end at time τ exiting to option 2:

$$\ln \Pr(t = \tau, d_2 = 1|\boldsymbol{x}) = \ln \phi_2(\tau, \boldsymbol{x}) + \sum_{s=1}^{\tau-1} \ln(1 - \phi_1(s, \boldsymbol{x}) - \phi_2(s, \boldsymbol{x})).$$

- Spells that are incomplete at time T^*:

$$\ln \Pr(t > T^*|\boldsymbol{x}) = \sum_{s=1}^{T^*} \ln(1 - \phi_1(s, \boldsymbol{x}) - \phi_2(s, \boldsymbol{x})).$$

LIML estimation

Alternatively, we can estimate the model **separately** for the two alternatives in a competing risk fashion.

In this case, the **conditional log-likelihood function** for alternative j is:

$$\mathcal{L}_{\mathrm{N}j} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{\tau=1}^{T^*} w_{i\tau}\{y_{ij\tau} \ln h_j(\tau, \boldsymbol{x}_i) + (1 - y_{ij\tau}) \ln(1 - h_j(\tau, \boldsymbol{x}_i))\},$$

or, equivalently:

$$\mathcal{L}_{\mathrm{N}j} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{\tau=1}^{T^*} w_{i\tau}\{y_{i\tau} d_j \ln h_j(\tau, \boldsymbol{x}_i) + (1 - y_{i\tau} d_j) \ln(1 - h_j(\tau, \boldsymbol{x}_i))\},$$

where $y_{ij\tau} = \mathbb{1}\{t_i = \tau, d_{ij} = 1\}$, $y_{i\tau} = \mathbb{1}\{t_i = \tau\}$, and $w_{i\tau} = \mathbb{1}\{t_i \geq \tau\}$, with $t_i = \min\{t^*_{1i}, t^*_{2i}\}$.