Appendix to

The hydrogen transition challenge: Exploring co-evolutionary market dynamics between vehicle fleet and infrastructure

This document is under construction – so far feel free to use this version

Appendix 1 – random model

An indicator for coverage is the probability that a driver with average driving radius of r_d miles, has no stations nearby, given a concentration of stations S :[1]

$$p_0 \equiv \Pr\left(\text{distance to a station} \geq \mathrm{r_d} \,\middle|\, \mathrm{S}\right) \tag{0.1}$$

Next, a patch is defined as the area that can (physically) be occupied by at most one station, the probability that one patch is occupied can be written as:

$$p_n = \frac{S}{n}\frac{A_d}{A} \tag{0.2}$$

with n being the number of patches within an individual's driving area $A_d = \pi r_d^2$.

Then, for sufficiently small patch size, the expected number of stations per unit circle follows a Poisson distribution.

As a corollary, and transforming to continuous coordinates, the probability of having at least c stations within one's driving range equals (with $S \geq c$):

$$p_c = 1 - e^{-\mu} \sum_{\gamma=0}^{c-1} \frac{\mu^\gamma}{\gamma!} \tag{0.3}$$

where

$$\mu \equiv p_n * n = \frac{S * A_d}{A} \tag{0.4}$$

represents the mean number of stations per driving area A_d , and thus the mean coverage for S stations and driving range r_d .

The assumption of uncorrelated station entrance also allows determining a static equilibrium adoption profile – that is, one that is history independent. For doing so, a cumulative adoption fraction F_{ad} for a given station density is found by summing adopters A_c that experience coverage S_c multiplied by the probability of occurrence p_c over all potential levels of coverage:

$$F_{ad} = \sum_{c=0}^{s} A_c * p_c \tag{0.5}$$

---

[1] The study of biological population ecologies deals with questions that are conceptually similar. For instance, Pielou ('77) seeks the probability that the distance between of a plant (species) to a nearest neighbor is larger than some critical radius for survival, r_c , given a population density s:

$p_{d,n}^{r_c} \equiv \Pr\left(\text{distance to nearest plant of same species} \leq \mathrm{r_c} \middle| \mathrm{s}\right)$ and next, with some simplifying assumptions about population distributions, convenient expressions of the solutions are derived.

The population level adoption for a given coverage is derived by integrating over the product of a population density function for critical thresholds of adoption and the adoption fraction $a_{c,t}$ given coverage relative to a threshold t at:

$$A_c = \int_0^{\infty} \left(\rho_t * a_{c,t} \right) dt \tag{0.6}$$

In other words, in this model, the non-instantaneous adoption profile is determined by two sources. The first addresses non-homogeneous characteristics of station location or infrastructure: an individual's utility of adoption increases with the actual coverage.[2] Assuming a binary adoption threshold at the individual level, an adoption fraction for the aggregate population must yield a smoother curve, as specific environmental (such as geographical) circumstances differ from person to person, while it can be assumed that for very small (large) coverage relative to the threshold the fraction of adoption is very small (high), while the responsiveness is low for both. Thus, we can characterize the adoption fraction as a function that increases with coverage relative the threshold

$$a_{c,t} = f\left(\frac{c}{t}; \sigma_t \right) \tag{0.7}$$

Where the sensitivity parameter σ_t captures the combined effect of all the non-homogeneous factors. For instance, a larger (smaller) population corresponds to a more uniform (homogeneous) distribution of factors, approximated by a lower (higher) σ_t .

A second source for non-instantaneous adoption is a distribution of preferences that captures the heterogeneity within the population with respect to individuals' adoption threshold. This threshold depends on the demographic, socioeconomic, and other characteristics of the population. For instance urban travelers can be expected to have a lower threshold than average, while the poor, those in rural areas and those who drive long distances to multiple locations will have a higher threshold. Generally, we can assume a two parameter function with average μ_t and sensitivity σ_t :

$$\rho_t = f\left(\mu_t; \sigma_t \right) \tag{0.8}$$

The probability that an individual driver has no coverage requires that all patches are unoccupied, or:

$$p_{c0,s}^{r_c} = \left(1 - p_n\right)^n = \left(1 - \frac{S * A_d}{A} \frac{1}{n}\right)^n \tag{0.9}$$

---

[2] Note that the first moment results for adoption fractions are independent of the *geographical* distribution of the population, this as a result of the assumption on random station entrance.

When converting to a continuous formulation, in which the patch size becomes infinitely small ($\partial x \partial y$), $n = \frac{A}{\partial x \partial y} \to \infty$, the probability of not having coverage becomes[3]:

$$p^{r_c}_{c0,s} = e^{-\mu_c} \tag{0.10}$$

where

$$\mu_c \equiv p_n * n = \frac{S * A_d}{A} \tag{0.11}$$

represents the mean number of stations per driving area A_d , or the mean coverage.
To get a feel for the properties of this relation under growing station density or increasing driving range, a useful reference value is the fill-factor ϕ that is defined such that $\phi = 1$ corresponds with coverage, irrespective of location, for a minimum of stations($\equiv S_{\phi=1}$).
This is achieved under hexagonal tiling with $S = \frac{A}{A_d}$ and consequently:

$$\begin{aligned} p^{r_d}_{c0,s_\phi} &= e^{-1} \approx 0.37 \\ p^{r_d}_{c0,s} &= \left(p^{r_d}_{c0,s_\phi} \right)^{\phi} = e^{-\frac{S}{S_\phi}\frac{r_d^2}{r_\phi^2}} \end{aligned} \tag{0.12}$$

Thus, the probability of having no coverage decreases exponentially with S and r_d^2 .

While the zero-coverage threshold is indicative, it does not provide serious insight into the attractiveness and/or actual adoption dynamics. For this, higher coverage thresholds need to be explored. Luckily, the situation of (0.9) is easy to generalize. Probabilities for higher coverage are given by the binomial distribution and can be approximated by a simple analytical expression (e.g. Pielou ('77); Casella and Berger ('90))[4]:

$$\begin{aligned} p^{r_d}_{S_c,S} &= \binom{n}{S_c} \left(p_n \right)^{S_c} \left(1 - p_n\right)^{n-S_c} \sim \frac{\left(np_n\right)^{S_c}}{S_c\,!} \left(1 - p_n\right)^{n} \\ &= \frac{\left(\mu_c\right)^{S_c} \left(1 - \mu_c / n\right)^{n}}{S_c\,!} \to \mu_c^{\ S_c} \frac{e^{-\mu_c}}{S_c\,!} \end{aligned} \tag{0.13}$$

Finally, the probability $p^{r}_{>S_c,S}$ of having at least S_c stations within one's driving range equals:

$$p^{r_d}_{>S_c,S} = 1 - \sum_{S_c=0}^{S_c-1} p^{r_d}_{S_c,S} = 1 - \sum_{S_c=0}^{S_c-1} \mu_c^{\ S_c} \frac{e^{-\mu_c}}{S_c\,!} = 1 - e^{-\mu_c} \sum_{S_c=0}^{S_c-1} \frac{\mu_c^{\ S_c}}{S_c\,!} \tag{0.14}$$

---

[3] also required is a small enough p_n , which is clearly satisfied for this problem (as well in applications below)

[4] Note that expected coverage should equal the mean number of stations per driving area, which holds: $E\left[S_c \left| r_d, S, A \right.\right] = \sum_{S_c=0}^{S} S_c \mu_c^{\ s_c} \frac{e^{-\mu_c}}{S_c\,!} = e^{-\mu_c} \sum_{S_c=0}^{S} \frac{\mu_c \mu_c^{\ s_c - 1}}{\left(S_c - 1\right)!} \approx e^{-\mu_c} \mu_c e^{\mu_c} = \mu_c = S \frac{A}{\pi r_d^2}$

Here we cannot approximate the outcome by one analytical expression, as S_c is generally very small, compared to the total number of stations. Note further that both a critical and troublesome assumption behind this is that the station distribution is uncorrelated with drivers.

Figure 1 shows the fraction of the population covered (that is, at least the number of stations equal to a threshold within one's driving range) for increasing station density. For the reference number of stations the characteristics of California are taken, which has a land area of 160.000 square mile, and an equilibrium situation of 10.000 stations. Assumed is an average driving radius of 50 miles per driver, which yields an average of 200 stations per driver. However, the results can be scaled. The results are shown for different thresholds.

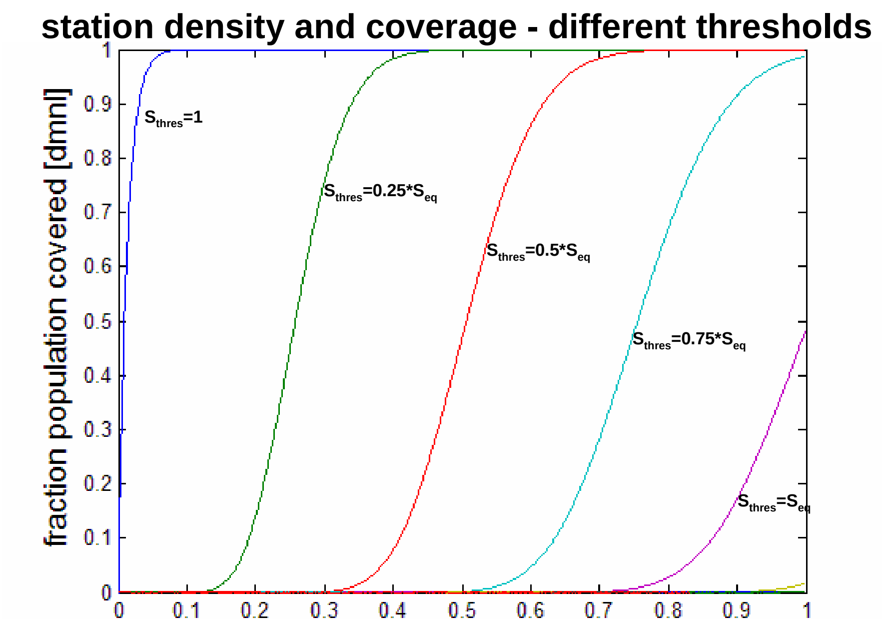

Figure 1 – Expected fraction population covered by at least the threshold for increasing station densities and different coverage thresholds.

To contrast this result from random emergence, perfect tiling (equidistant locating of all stations) would result in step functions that cross the random curves at 50% coverage (For this random distribution experiment both expected and median of coverage are 50%). Towards the other extreme, more clustering (higher mean-variance ratios) results in less steep curves

On the adoption curves

With $s_{\text{cov}}^{rel} \equiv \dfrac{s_{\text{cov}}}{s_{thresh}}$

$$
\begin{aligned}
p_{ad}\left(s_{\text{cov}}^{rel}\right)\Big|_{s_{\text{cov}}^{rel} \ll 1} &\approx 0 \qquad & p_{ad}{}'\left(s_{\text{cov}}^{rel}\right)\Big|_{s_{\text{cov}}^{rel} \ll 1} &\approx 0 \\
p_{ad}\left(s_{\text{cov}}^{rel}\right)\Big|_{s_{\text{cov}}^{rel} \gg 1} &\approx 1 \qquad & p_{ad}{}'\left(s_{\text{cov}}^{rel}\right)\Big|_{s_{\text{cov}}^{rel} \gg 1} &\approx 0
\end{aligned}
\tag{0.15}
$$

A useful approximation is for instance the logistic growth curve[5]:

$$
f_{ad}\left(s_{\text{cov}}, s_{thresh}\right) = \frac{\exp\left(4\sigma_{\text{cov}}\left(s_{\text{cov}}^{rel} - 1\right)\right)}{1 + \exp\left(4\sigma_{\text{cov}}\left(s_{\text{cov}}^{rel} - 1\right)\right)}
\tag{0.16}
$$

The foundation for this is the binomial choice model with the attractiveness of the alternative choice being equal to 1. A larger (smaller) elasticity to a change in coverage σ_{cov} indicates a more aggressive (smoother) adoption response to a change in coverage. The larger the population group considered, within a segment, the smoother the curve will be.

Density

For the distribution of the population preferences there is again a variety of choice for the density function. Here we choose one that is symmetric around the equilibrium value, such that the average threshold is at the equilibrium value (i.e., in the case of California this equals 100 stations within a disc with radius of 50 miles).

:

$$
\rho\left(s_{thresh} \,\middle|\, \sigma_{thresh}\right) = \frac{1}{Cum_{thresh}}\left(s_{\text{cov}}^{rel} * \left(2 - s_{\text{cov}}^{rel}\right)\right)^{\sigma_{thresh}}
\tag{0.17}
$$

with Cum_{thresh} the normalization term and parameter σ_{thresh} representing the heterogeneity in preferences. A higher (lower) value implies a larger (smaller) concentration at the mean/modus and thus more (less) homogeneity.

Adoption curve

The figure shows that the adoption curves have a slope of around 1, but more generally this slope can be shown to approximate:

$$
\frac{dF}{dS} \approx \widehat{2S_{thresh}} \qquad for \qquad 0 << S << \widehat{S_{thresh}}
\tag{0.18}
$$

That is, while not shown, shifting the average threshold only implies a linear shift in the adoption curve.

---

[5] There are many plausible alternatives (Weibull, Richards, Gompertz etc…). The choice of the function does not matter much for this coarse and conceptual level of analysis. The main argument to choose this form are its simplicity in shape (symmetric properties), such that the resulting patterns are easy to interpret, and, associated with this, the limited number of parameters (two). The cumulative normal distribution would also be an obvious choice; however no simple closed form solution exists for this.

Appendix 2 – tables

Table 1 – indices used for variables

Shorthand	Definition	Description / Note
∂	Household type	Different types can have different mobility patterns or different socio-economic characteristics.
l	Location	Location in the area A
t	Trip destination location	Location is relative to stating location of a household. In the model t is replaced by polar coordinates ($t = f(r,\theta)$)
u	Underway location	Arbitrary location between home and destination location for a trip. Used to identify "critcal location" at which service is required
s	Station location	relative to stating location of a household
blue	Observed parameters	
red	Unobserved parameters	

Table 2 - aggregate demand variables

Shorthand	Definition	Derivation
$H_a^{l,\partial}$	Total adopters of type ∂ in location l	$A_p h_a^{\partial}$
H_a	Total adopted population	$\int_{\partial,l} H_a^{l,\partial}$
$m^{\partial,t}$	Annual miles driven per type for a trip t by an individual of type ∂ (in location l)	$2d^t * f^{\partial,t}$
m^{∂}	Annual vehicle miles, for an individual of type ∂ in location l	$\int_t m^{\partial,t}$
$M^{l,\partial}$	Total annual miles driven by all individuals of type ∂ in location l	$A_p * \rho_h^{\partial} * m^{\partial}$
M	Total annual miles driven by the total population	$\int_{\partial,l} M^{l,\partial}$

Appendix 3 – Constant Elasticity of Substitution function

A useful non-linear weighted average function is the Constant Elasticity of Substitution (CES) function. As this form will be used throughout, its properties will be discussed here briefly. Its general shape is, in continuous form:

$$\bar{x} = \left[\int_y \rho(y) * \left(x(y) \right)^{\varepsilon} \right]^{1/\varepsilon} ; \qquad (0.19)$$

ε is the key parameter, that allows to incorporate assumptions on what. Higher (lower) values imply larger contributions from the bigger (smaller). In the extreme, $\varepsilon \to \infty(-\infty)$ implies the non-linear weighted average equals the maximum (minimum) contributions. Other special cases are $\varepsilon = 1$ (linear) and $\varepsilon = 0$ (unit elasticity of substitution, Cobb-Douglass). Further, ρ is the density function (thus, $\int_y \rho(y) = 1$), allowing for a non-uniform distribution over importance.

Potential applications

First, this could also be used to determine a non-linear weighting over different factors that. For instance, averaging over the underway sites of a trip t the effort entails:

$$e^{t,\bar{U}} = \frac{1}{d^t} \left[\int_{u=0}^{t} \left(e^{t,u} \right)^{\varepsilon_u} \right]^{1/\varepsilon_u} ; \qquad \varepsilon_u < 1 \qquad (0.20)$$

where $e^{t,u}$ is the effort, as perceived on one particular point on the trip d^t is the total trip distance.

Note further that one can use this expression to convert between two reciprocals, such as "trip effort" and "trip coverage" and can derive effective aggregates in a mathematically identical, but more intuitive way. When contribution to trip coverage is defined as the inverse of effort:

$$c_l \equiv \frac{1}{(e_l)^{\alpha}} \qquad (0.21)$$

then, for instance, the average effort can be found by integrating over coverage, weighted by any desired weight function:

$$\bar{e} = \left[\bar{c} \right]^{\frac{1}{-\alpha}} ; \qquad \bar{c} = \int_l w_l * c_l \qquad (0.22)$$

Where between brackets the effective coverage is derived. The different weighting of locations in terms of coverage is a much more intuitive exercise.

Appendix 4 – Parameters

Observable parameters

Parameter	ShN	Value	Implication
Average Population density	ρ_h	35	[people/square mile]
Driver adoption time, normal hazard rate	$\tau_a, 1/\lambda_n$	1	[year]
Typical Driving radius	d_n^s	30	[miles]
Average distance of a trip	$\widehat{r_f}$	0.6*rd	[miles] this implies that the variance parameters σ_{df} and σ_{df} are equal to 0.5 mode is 0.4*rd
Fuel economy	mpg	25	[miles/gallon equivalent]
Tank size	tf	4	[gallons equivalent]
Vehicle miles	Nvm	15.000	[miles/vehicle/year]
Fixed cost	FC	1,000,000	$/year
Reference profits per station	π_{ref}	0.1*Fixed Cost	
Normal station capacity	Nac	Na	Driving is only constraint by attractiveness from the supply side
Sensitivity of Expected Effective Supply to Distance	β_c	-0.6	Calibrated to correspond most local rational behavior – that is , probability of location, based on expected profits, corresponds with the relative attractiveness for that location, based on actual profits after entry. This is calibrated for station entry with existing grids with different distances.
Sensitivity of Expected Effective Demand to Distance	β_r	-0.2	
Normal fractional capture of potential market after entry	$g_{r,n}$	0.2	

Unobservable parameters

Model Parameter	ShN	Value	Implication
Sensitivity of Utility to drive to Effort	β_e	-0.5	At 50% adoption, if effort decreases by 1% , utility will increase by 0.5%.
Utility Inflection Point for Trip Effort	$e_{0.5}$	2	Average utility effect from effort equals 0.5 when effort is twice the normal effort, leading to 0.5*50% adoption fraction
weight distance effect on effort, weight of risk effect on effort, and weight of crowding on effort	w^{td} w^{ri} w^{tf}	0.5 0.5 0	
Fraction tank range critical	α_f	0.06	
Sensitivity of Market Attractiveness to Market Profitability	β_{en}	0.5	
Sensitivity of Exits to Profitability	β_{ex}	-0.1	
Sensitivity of location share to expected Profits to Expected Profitability s	$\beta_{en,s}$	2	
Sensitivity of Station share to effort	β_s	$-\infty$	People have full information; all trip trajectories are identical to the average. People select the best service station available.