www.simiode.org
SIMIODE A SYSTEMIC INITIATIVE FOR MODELING INVESTIGATIONS & OPPORTUNITIES WITH DIFFERENTIAL EQUATIONS

STUDENT VERSION
The flush toilet: slow and late start

Maila Brucal-Hallare and Charles Lamb
Department of Mathematics
Norfolk State University and Indiana University of Pennsylvania
Norfolk VA USA and Indiana PA USA

Abstract: This activity analyzes the spread of a technological innovation using the Bass Model from Economics. The equation is a first-order, two-parameter separable equation and the solution has a characteristic S-shaped curve or sigmoid curve. The students derive the solution to the model, use least squares method to estimate the parameters, and investigate another technological innovation using the method outlined here.

Keywords: Bass Model, rate of innovation, rate of imitation, diffusion of innovation

Tags: first-order differential equation, parameter-fitting, Julia, Excel

1 Scenario Description

To flush, or not to flush? That is not the question. The question is, how does one use differential equations to model the spread of toilet flushing as a technological innovation in the United States?

Various historical sources credit Sir John Harington, godson of Queen Elizabeth I, in designing the first flushable toilet in 1596. His design contained the two components of the modern toilet: a valve to flush out the water and a wash-down system to empty the bowl. His design, however, failed to eliminate the foul smell. More than two hundred years after, Alexander Cumming, a Scottish mathematician and inventor, improved Harington's design by inventing the S-trap that retained water permanently in the waste pipe and prevented the foul smell from escaping. The spread of Cumming's S-trap design was slow. When the Great Stink in Central London happened in the summer of 1858, the legislators of the Houses of Parliament pushed to require the S-trap as part of the sewage system [6]. The Houses of Parliament was located along the River Thames, which was practically an open sewer at that time. Indeed, government legislation has the power to accelerate the diffusion of technological innovations and in this case, it used its power to control the diffusion of the stink.

In America, flush toilets were introduced in the homes of the wealthy in the 1860's. Up until the 1980's, only 98 percent of American households have access to flush toilets. Table 1 below summarizes the data on how this technological innovation spread in the United States:

Year	Percentage
1860	1
1880	7
1890	13
1900	15
1920	20
1930	51
1940	60
1950	71
1960	87
1970	96
1980	98
1989	100

Table 1. The data on the innovation and diffusion of the flush toilet, OurWorldInData.org [5]. In the 1860s, only the American wealthy elites (1 percent of the population) have access to flush toilets.

In this activity, we will investigate the Bass Model, a mathematical model that economists use to model the diffusion of an innovation:

$$\frac{dF}{dt} = (p + qF)(1 - F) \tag{1}$$

where $F(t)$ is the fraction of the population who have adapted the technology, p is the rate of spontaneous innovation, and q is the rate of imitation, see [2]. The model does not take into account the social structure or even government intervention that significantly contributes to the spread of innovation. The second factor $1 - F$ captures the portion of the population that has not adapted the technology.

Observe that when $p = 0$, we obtain the classic logistic model, whose solution is shaped like an S, and when $q = 0$, we obtain the exponential model whose solution is asymptotic to the horizontal $F = 1$. The solution of the Bass Model for $p + q \neq 0$ has a characteristic S-shape, too, with the parameter p as the initial main driver of the spread, then the parameter q takes over, until the function $F(t)$ approaches 1. The first activity guides the students to a step-by-step approach to finding the closed-form of the solution to (1).

Next, the second activity uses the least-squares method to estimate the parameters when the Bass Model is applied to the data in Table 1. Special optimization packages, Optim in Julia and Solver in Excel, are used to approximate values for p and q to model the innovation and diffusion of the flush toilet in American homes.

A final activity gives the students the opportunity to fit data on another technological innovation: the household refrigerator. Like the flush toilet, only the wealthy had refrigerators in their homes during the first quarter of the 1900's. It was around the 1980's when finally, all American households owned a refrigerator.

Activity 1. Analyzing the mathematical model.

Consider the initial-value problem

$$\begin{cases} \dfrac{dF}{dt} = (p + qF)(1 - F) \\ F(0) = 0. \end{cases} \tag{2}$$

1. Using separation of variables, we obtain,

$$\frac{dF}{(p + qF)(1 - F)} = dt.$$

   The left-hand side can be integrated via the partial fraction decomposition technique, provided $p + q \neq 0$:

$$\frac{1}{(p + qF)(1 - F)} = \frac{A}{p + qF} + \frac{B}{1 - F}.$$

   Verify that $A = \dfrac{q}{p + q}$ and $B = \dfrac{1}{p + q}$.

2. With the given values for A and B, integrate the separated equation. Verify that we obtain:

$$t = \frac{1}{p + q} \ln \left| \frac{p + qF}{1 - F} \right| + C.$$

3. Next, we exponentiate the previous result and use the initial value $F(0) = 0$ to compute for C. Verify that the solution to the initial-value problem (2) is

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}} \tag{3}$$

Activity 2. Fitting the data using the model via the Optim package in Julia.

In order to approximate the parameter values p and q that measures rate of innovation and adoption, respectively, we will use the least squares approach.

Using the data in Table 1, denote the time by t_i and the data values in the second column by y_i. The goal is to fit the Bass model solution to the historical data. At a given time t_i, the error E between the model $F(t_i)$ and the actual data $y_i = y(t_i)$ is given by

$$E = \sum_{i=1}^{n} (F(t_i) - y_i)^2.$$

Simplifying, we get

$$E = \sum_{i=1}^{n} F(t_i)^2 - 2F(t_i)y_i + y_i^2,$$

which is a function that depends on the parameters p and q. The goal is to find values for p and q that minimizes the error E over all the given values t_i. This is a so-called *least-squares* approach, see [4] for more on intrinsically nonlinear models that cannot be linearized.

In order to accomplish this, we will use Julia, an open source programming software that has powerful mathematical capabilities, [1]. The Optim package [3] in Julia provides solutions to optimization problems, including unconstrained ones like what we have here.

In [1]:

```
# Enter the Collected Data
years = [1860, 1880, 1890, 1900, 1920, 1931, 1940, 1950, 1960, 1970, 1980, 1989];
adoption = [1, 7, 13, 15, 20, 51, 60, 71, 87, 96, 98, 100];
```

In [2]:

```
# Define Solution to Bass Model
# Note: * F has output in [0, 1]
#       * independent var. t; params p, q
function F(t, p, q)
    e = exp(-(p+q)*t)
    return (1 - e)/(1 + (q/p)*e)
end
```

In [3]:

```
# Transform Data
# Note: * The "." applies subtraction to each element of years
T = years .- 1860  # Shift data so 1860 is beginning of time
Y = adoption / 100 # Rescale adoption to values in [0, 1]
```

In [4]:

```
# Compute square error sqErr
# Note: * vectors: T, Fpq, Y
#       * scalars: p, q, E
#       * input: (p, q) is a tuple; note double set of parens, see appendix
function sqErr((p, q))
    Fpq = F.(T, p, q) # Apply F to each element t_i in T with params p, q
    E = sum(Fpq.*Fpq - 2*Fpq.*Y + Y.*Y)
    return E
end
```

In [5]:

```
# Minimize the square error
# Note: * optimize takes a function with vector input
#       * initial guess: p = .25, q = .5
using Optim
results = optimize(sqErr, [0.25, .5])
```

In [6]:

```
# Extract minimizing parameters from results object
(pMin, qMin) = Optim.minimizer(results)
```

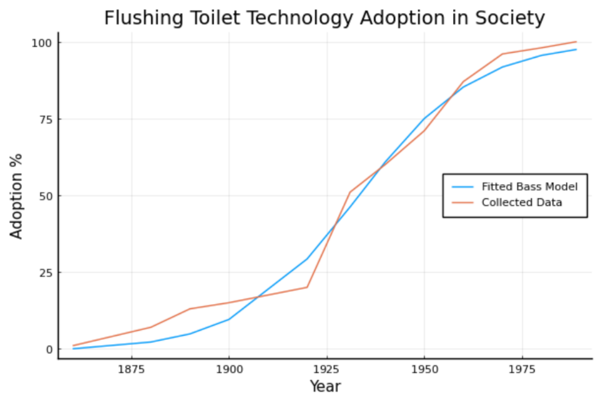


Figure 1. A visualization of the actual data and the fitted Bass model, with the optimal parameter values $(p, q) \approx (0.000544, 0.0649)$.

In [7]:

```
# Plot data and fitted model on same set of axes
# Note: * We mult F by 100 to scale to 0 - 100 interval
#       * plot! modifies current plot instead of overwriting
using Plots
plot(years, adoption, label="Collected Data")
Ffit = F.(T, pMin, qMin) * 100
plot!(years, Ffit, label="Fitted Bass Model",
    legend=:right, xlabel="Year", ylabel="Adoption %",
    title="Flushing Toilet Technology Adoption in Society")
```

Observe that in cell 3, time T is rescaled so that $T = 0$ corresponds to 1860. This modeling choice induces a small amount of error. Recall, the form of $F(t)$ in (3) was found under the assumption that $F(0) = 0$, which does not correspond to the historical data of 1% adoption in 1860. If however, $T = 0$ was chosen to correspond to 1596, we would skew the model and ultimately alter the values of the fitted parameters. There are trade-offs to either choice.

Observe in cell 5, the initial guess $[p_0, q_0] = [.25, .5]$ was chosen arbitrarily so that $p_0 \neq q_0$ and that both p_0 and q_0 are small. If the choice $p_0 = q_0$ is made, the symmetry in parameters might fail to be broken.

To complete this activity, either download Julia or complete the following steps:

1. Create an account in http://cocalc.com.
2. Select *Create New Project.* Create a title.
3. Select *Jupyter Notebook* and the kernel, *Julia.*
4. Copy the cells above and SHIFT-ENTER to run a cell.
5. What parameter values fit the model to the data? How does the graph look?

Activity 3. Fitting the data using the model via the Solver package in Excel.

Microsoft Office's spreadsheet application tool, Excel, can also be used to solve unconstrained optimization problems. Excel has its own optimization package called Solver. These are the steps to complete in order to estimate the parameters p and q:

1. Prepare the Excel spreadsheet before using Solver (see Figure 2):
   (a) Create one column each for the scaled data in Table 1: time t_i and percentage y_i.
   (b) Create a column, *Bass Model Fit*, that computes $F(t_i)$ using (3) with an initial guess for the parameters p and q.
   (c) Make a line graph to visualize the plots of (t_i, y_i) and $(t_i, F(t_i))$. Adjust the initial guess for the parameters p and q using this line graph.
   (d) Create columns that compute the error, $f(t_i) - y_i$, and the square of the error. Then compute the sum of squares of error (SSE); in Figure 2, SSE is recorded in cell G15.

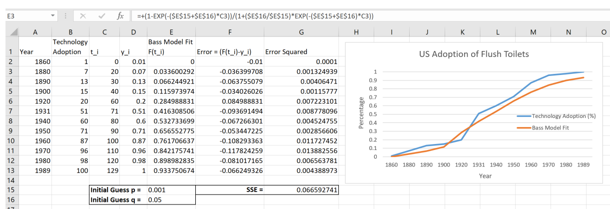


Figure 2. Setting up the data before using the Solver package in Excel.

2. Go to Data and click Solver. (If Solver is not available, then it must be added using the following steps: File $\rightarrow$ Options $\rightarrow$ Add-ins $\rightarrow$ Solver Add-in $\rightarrow$ OK.)
3. We are ready to compute the minimum of the sum of the squares of the errors. In Solver, indicate G15 in *Set Objective*. Click *minimum*. Indicate the location of the parameters p and q in *By Changing Variable Cells*. Hit Solve. See Figure 3.

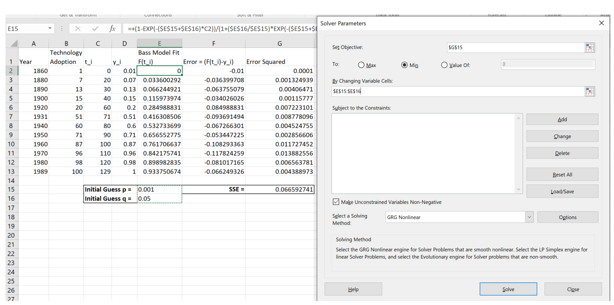
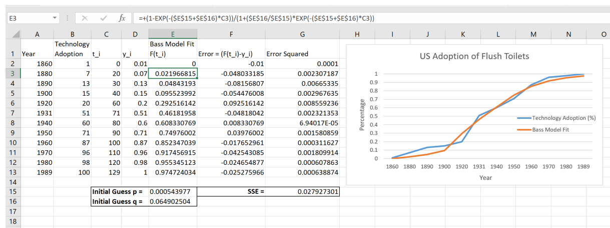


Figure 3. Using Solver in Excel.

4. Solver will update the parameter values that solves the problem. Compare the values obtained by Julia and by Excel. See Figure 4.



Year	Technology Adoption	t_i	y_i	Bass Model Fit F(t_i)	Error = (F(t_i)-y_i)	Error Squared
1860	1	0	0.01	0	-0.01	0.0001
1880	7	20	0.07	0.021966815	-0.048033185	0.002307187
1890	13	30	0.13	0.04843193	-0.08156807	0.00665335
1900	15	40	0.15	0.095523992	-0.054476008	0.002967635
1920	20	60	0.2	0.292516142	0.092516142	0.008559236
1931	51	71	0.51	0.46181958	-0.04818042	0.002321353
1940	60	80	0.6	0.608330769	0.008330769	6.94017E-05
1950	71	90	0.71	0.74976002	0.03976002	0.001580859
1960	87	100	0.87	0.852347039	-0.017652961	0.000311627
1970	96	110	0.96	0.917456915	-0.042543085	0.001809914
1980	98	120	0.98	0.955345123	-0.024654877	0.000607863
1989	100	129	1	0.974724034	-0.025275966	0.000638874
		Initial Guess p =		0.000543977	SSE =	0.027927301
		Initial Guess q =		0.064902504		

Figure 4. Estimate of the parameters using Solver in Excel.

Activity 4. Analyzing America's Adoption of a Refrigerator.

Having completed the previous that investigated the adoption of the flush toilet as a technological innovation in the United States, let us now look at actual data on the percentage of American households that own a refrigerator. See Figure 5.

1. Go to the website https://ourworldindata.org/technology-adoption. The website contains data on the US adoption of various technological innovations - charts, tables, and CSV files are available. Explore the website and be excited about the research possibilities that can arise from the information here!
2. Click on *Download the CSV file* and find data (t_i, y_i) on the adoption of the refrigerator (*not "household refrigerator"*). Evidently, there are two different refrigerator datasets.
3. Use either the Julia code or implement the Excel steps as described in the previous activities in order to estimate the parameters p and q.
4. What parameter values fit the model to the data? How does the graph look?

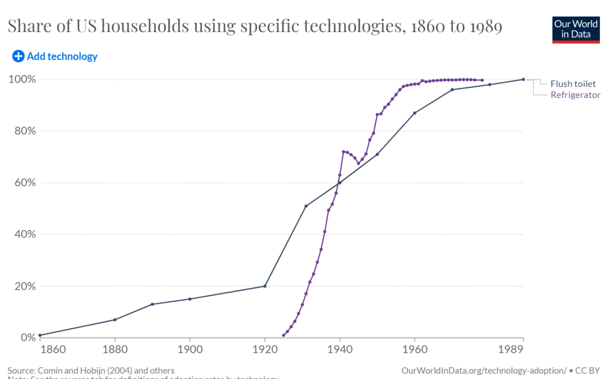


Figure 5. Percentage of American households that own a refrigerator vs having a flush toilet.

5. Compare the ratios q/p for the flush toilet data and the refrigerator data. Discuss the discrepancy. See Figure 5. Recall the interpretations of p and q, and you may wish to consider the speed of information spread in different eras as well as the costs of technological adoption.
6. Research the story behind the adoption of the household refrigerator. Why is there a dip in the data during the first few years of the 1940's?
7. There are various governmental, economic, and political factors that affect both the innovation and imitation parameters of the model. The introduction of this document discusses legislation in London in the summer of 1858 affecting the design and adoption of the flushing toilet. In the data, the landline is affected by an event in the early 1930s, and the previous question considered the refrigerator dip in the 1940's. What has not been discussed is the influence of other technologies. Examine the landline data. Describe what happens to the landline in the early 2000's and why the Bass model is not an appropriate model for this data.

Appendices

Julia Tutorials and Code Remarks

Julia is an open source mathematical language akin to MATLAB. Below are a few resources for further investigation:

- General Purpose Tutorials `https://julialang.org/learning/tutorials/`
- Symbolic DEs (SymPy) `https://docs.juliahub.com/SymPy/KzewI/`
- Numerical DEs (SCiML) `https://tutorials.sciml.ai/`

We remark on the included Julia code.

1. The vectorized "dot" operator applies the operator to each entry of the vector.
   (a) For example, `v .- 100` is equivalent to `v[1] = v[1] - 100`, `v[2] = v[2] - 100`, etc.
   (b) Functions can also be applied in a vectorized manner; such was done with `F.(T, p, q)`. The output is a vector consisting of the entries `F(`t_1`, p, q)`, `F(`t_2`, p, q)`, etc.
2. Argument destructuring was used in the definition of the square error: `function sqErr((p,q))`. A vector is passed into `sqErr`, length is expected to be two, and the first element is named `p` and the second is name `q`, see `https://docs.julialang.org/en/v1/manual/functions/#Argument-destructuring`.
3. Semicolons suppress output, like in cell 1.
4. External packages, such as `Optim` and `Plots`, need to be imported with `using`.
5. The bang `!` convention is used in Julia to denote a mutable operation, i.e. a function will update/modify its arguments. For example, `plot!` modifies the existing plot.

Single Julia Listing

```julia
# Enter the Collected Data
years = [1860, 1880, 1890, 1900, 1920, 1931, 1940, 1950, 1960, 1970, 1980, 1989];
adoption = [1, 7, 13, 15, 20, 51, 60, 71, 87, 96, 98, 100];

# Define Solution to Bass Model
function F(t, p, q)
    e = exp(-(p+q)*t)
    return (1 - e)/(1 + (q/p)*e)
end

# Transform Data
T = years .- 1860;
Y = adoption / 100;

# Compute square error sqErr
function sqErr((p, q))
    Fpq = F.(T, p, q)
    E = sum(Fpq.*Fpq - 2*Fpq.*Y + Y.*Y)
    return E
end

# Minimize the square error
using Optim
results = optimize(sqErr, [0.25, .5])
println(results)

# Extract minimizing parameters from results object
(pMin, qMin) = Optim.minimizer(results)
println("The minimizing parameter values are (p, q) = ($pMin, $qMin).")
println("The ratio q/p = $(qMin/pMin)")

# Plot data and fitted model on same set of axes
using Plots
plot(years, adoption, label="Collected Data")
Ffit = F.(T, pMin, qMin) * 100
plot!(years, Ffit, label="Fitted Bass Model",
    legend=:right, xlabel="Year", ylabel="Adoption %",
    title="Flushing Toilet Technology Adoption in Society")
```

Listing 1. Julia code to fit Bass model parameters p and q

REFERENCES

[1] Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. *Julia: A fresh approach to numerical computing*, SIAM Review, **59** (2017), 65-98, DOI https://doi.org/10.1137/141000671.

[2] Jackson, M. *Social and Economic Networks*, Princeton University Press, 2010.

[3] Mogensen et al. *Optim: A mathematical optimization package for Julia.* Journal of Open Source Software **3** (2018), no. 24, 615, DOI https://doi.org/10.21105/joss.00615.

[4] Mooney, D. & Swift, R. *A Course in Mathematical Modeling*, American Mathematical Society, 1999.

[5] Ritchie, H. & Roser, M. *Technology Adoption*, Our World In Data, 2017, https://ourworldindata.org/technology-adoption.

[6] Wikipedia Contributors, *Flush Toilet*, retrieved Oct, 15 2021, https://en.wikipedia.org/wiki/Flush_toilet.