Queen's Economics Department Working Paper No. 1467

# Was Harold Zurcher Myopic After All? Replicating Rust's Engine Replacement Estimates

Christopher Ferrall

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

10-2021

# Was Harold Zurcher Myopic After All?
# Replicating Rust's Engine Replacement Estimates

**Christopher Ferrall**
Queen's University, Kingston, Canada
ferrallc@queensu.ca

October, 2021 [Current]

**Abstract**

Rust (1987) concludes that the data "clearly reject" the hypothesis that Harold Zurcher made bus replacement decisions using a monthly (and myopic) discount rate of 0 in favor of the factor 0.9999. The alternative model requires the nested fixed point algorithm developed in the paper which became the basis of an ongoing empirical literature. The p-value of the likelihood ratio test was 0.053. Recoding the preferred model and re-processing the raw data reveals two types of errors in the original analysis and a revised p-value of 0.078. This remains below .10, which can be inferred as the significance level that clear rejection was based on. Thus the myopic hypothesis is again rejected although for lower conventional significance levels it would not be.

## 1. Introduction

The Rust (1987) model of bus engine replacement is a founding application of empirical dynamic programming (DP). Using data from a city bus garage, the paper studies the decision to replace a bus's engine (or not) during monthly maintenance and inspection. The model continues to be used for teaching because it is simple and can be adapted and extended to other decisions.

Rust released documented Gauss code and data in the 1990s that is still available (Rust 2000). The use of that code is unknown. As far as I know, no replication of the MLE estimation estimates have been published using independent code. Augirreibiria and Mira (2002) use the same data and provide code for their pseudo MLE procedure but do not report replications of Rust's results.

This paper replicates selected sets of MLE estimates reported in Rust (1987) using the software package niqlow, a platform for designing, solving and estimating empirical DP models described in Ferrall (2020) and available at ferrall.github.io/niqlow. The replication code consists of high-level statements in niqlow rather than purpose-built code for this particular model. This report focuses on one specification emphasized in the original paper. The code replicates other original estimates as well.

Despite a 35 year gap in computing hardware and sofware, and a fundamentally different approach to coding the model, the solutions of the DP model appear to be identical for a given set of parameter values. However, MLE estimates do differ, and the cause appears to be a difference in the transformation of the original raw data. New code that reads the original data files results in different observations than the original sample. The new data coding agrees more closely with independent values available in the data archive Augirreibiria and Mira (2010) than the inferred values from the original text. The author was contacted but did not respond to questions surrounding the data discrepancy.

A main statistical inference reported in the paper is a likelihood ratio test of the null hypothesis of myopic behavior in replacing bus engines, which corresponds to estimates setting the discount factor at 0, versus a forward-looking model with the discount factor fixed at 0.9999. That specification requires the nested fixed point solution developed in the model. This replication attempts to reproduce both the MLE estimates of the parameter values and the results of these hypothesis tests.

# 2. Model and Its Solution

In the model, bus engine maintenance is a stationary environment in which a binary decision to replace an engine $(d = 1)$. This resets the state variable $x$, mileage on the engine, to 0. For estimation, the odometer readings obtained from the garage are converted to discrete values. The transition of $x$ next month takes on one of 3 values: 0, 1 and 2:

$$x' = j + x(1-d),\ Prob(j=i) = \theta_{3i},\ i = 0,1,2. \quad (1)$$

Other elements of the model are the monthly discount factor $\delta$, utility $U()$, and a vector, $\zeta$, of additive extreme value shocks to smooth choice probabilities. (The terminology here follows conventions in niqlow and differs from the original paper.)

Table 1 summarizes the model using the niqlow framework. A new class named Zurcher, is derived from the pre-defined Rust class. The state variable $x$ is an object of the Renewal class, following the term Rust used to describe the process.

**Table 1. Rust (1987) Model Summary**

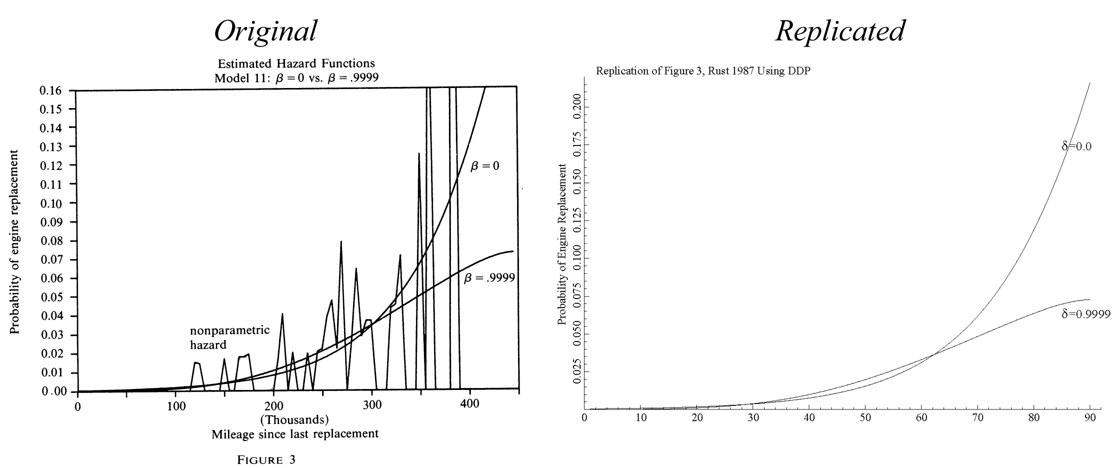
| Element | Value | Category | Params / Notes |
|---|---|---|---|
| Clock | $t$ | Ergodic | |
| CCP | $\zeta$ | ExtremeValue | |
| Actions | $\alpha = (d),$ | Binary Choice | $d = 1$ means replace engine. |
| States: | $\theta = (x)$ | Renewal | $N = 90$ or 175, probabilities $\theta_3$ |
| Utility | $U(\,) = \begin{pmatrix} -0.001\theta_1 x \\ -RC \end{pmatrix}$ | | Linear cost. |
| Parameters | $\psi = (RC, \theta_1, \theta_3)$ | | $\delta$ fixed at 0 and 0.9999 |

In niqlow the state vector is denoted $\theta$ which is unrelated to the subscripted parameters $\theta_1$ and $\theta_3$ in Rust's model.

## 2.1 Choice Probability Replication

Before turning to replicating the ML estimation procedure, consider the output at one of the estimated models in the original paper. Figure 3 in the paper compares predicted values of $P^{\star}(1;x)$, the probability of replacement given mileage, for discount factors of $\delta = 0$ and $\delta = 0.9999$. Figure 1 compares the original predictions and the replications, both using the Newton-Kantorovich algorithm which solves the fixed point in the value function as a system equations after an initial number of Bellman iterations. The additional information in the original figure is the empirical hazard rate. The replication is very close as can be determined when looking at the axis values at key points. For example, the curves both cross at about $x = 63$ and $P^{\star}$ at 0.03.

**Figure 1. Replicating Rust Figure 3**




# 3. Data and Estimates Replication

The original article estimates many different specifications on different bus engine types. The replication focuses on Column 2 of Table IX, a case emphasized in the original and the basis of the replicated Figure 3 above. This specification includes the linear cost of bus maintenance already shown. It is estimated on group 4 engines, setting the number of discretized odometer readings to 90, implying a width of 5000 miles per discrete bin. That is, if $m$ is the original mileage in thousands then $x$ equals the integer $k$ such that $5k \leq m < 5(k+1)$. With this coding the maximum change in $x$ from one month to the next is 2 (no more than 15,000 miles), which is why the innovation $j$ takes on values 0, 1 and 2.

The odometer was not reset when an engine was replaced so the raw data is not consistent with the model. Instead, the resetting and discretizing of mileage is done by the same Gauss program that estimated the model. Attempts to run the 20-year-old Gauss code and to access formatted files were unsuccessful. The replication code repeats the procedure described in the documentation and finds some discrepancies with reported statistics as discussed below.

The transition $j = x_{s+1} - x_s(1 - d_s)$ is constructed from the path of the discretized data. The transitions are IID and both $x$ and $d$ observed the transition probabilities $\theta_3$ form a simple multinomial problem not requiring optimal decisions (Rust's first stage estimation). Table 2 and Table 3 report the estimates of $\theta_3$ (fractions of transitions $j$) for bus groups 1-3 and group 4 from the original paper. The estimated transitions are compared with the values from the replication data.

The online archive related to Aguirregabiria and Mira (2010) provides files containing the Rust data in a flat format that was readable and is used as an independent check on the new code's output. The other columns of Tables 2 and 3 report the author's and AM (2010)'s data on $j$. Table 2 reports the 90 bin discretization; Table 3 reports the 175 bin version.

Looking first at group 4 (Panel B), which Figure 1 is based on, each month is 1/4292 = .00023299 of the group 4 sample. This equals 1/10th of the difference between the fractions. The discrepancy appears to be that 10 bus-month transitions are categorized as $j = 0$ in the replication that were originally coded as $j = 1$.. Converting the AM (2010) continuous mileage variable into $x$ and computing $j$, the observation counts match except for one observation beyond the maximum jump of 2 bins.

The multinomial likelihood differs by only .13% between the two codings of the discrete data. Since transition probabilities enter into the problem solved by the agent when determining conditional choices to replace an engine they have an effect on the solution.

In Groups 1-3 (Panel A) the discrepancy in the likelihood is somewhat larger, as summarized by the likelihood values and difference in cell percentages. There are 2 bus-month combinations categorized differently using the AM 2010 data. The replication code puts two observations in the $j = 1$ bin instead of $j = 0$, and again one observation is beyond $j = 2$ that is recoded in the replication.

**Table 2. Reconstructing Discretized Mileage Data (90 bins)**

**A. Column 1 (Groups 1-3)**

| | Replicated | | Original | | AM10 |
|---|---|---|---|---|---|
| **j** | **N** | **%** | **%** | **Δ** | **N** |
| 0 | 1182 | .3059 | .3010 | .0049 | 1184 |
| 1 | 2640 | .6832 | .6884 | -.0052 | 2638 |
| 2 | 42 | .0109 | .0106 | .0003 | 41 |
| 3+ | | | | | 1 |
| Obs | 3864 | | 3864 | | 3864 |
| lnL | -2595.63 | | -2575.88 | | |

**B. Column 2 (Group 4)**

| | Replicated | | Original | | AM10 |
|---|---|---|---|---|---|
| **j** | **N** | **%** | **%** | **Δ** | **N** |
| 0 | 1693 | 0.3945 | .3919 | .0023 | 1,693 |
| 1 | 2544 | 0.5927 | .5953 | -.0023 | 2,544 |
| 2 | 55 | 0.0128 | .0128 | .0000 | 54 |
| 3+ | 0 | | | | 1 |
| Obs | 4292 | | 4292 | | 4292 |
| lnL | -3145.11 | | -3140.57 | | |

Source of Original: Rust (1987), Table IX using MLE estimates of transition probabilities. Note: Table V shows unrestricted MLE estimates of these transition probabilities. Table IX are estimates joint with the dynamic programming model and are potentially different. However, Table V only shows 3 digits that agree with the Table IX whereas the coding discrepancy is large enough to affect the third digit. Replicated: Author's calculation from plain text data files included in the packaged documented in Rust (2000). AM: Author's calculation using continuous mileage data provided in the archive associated with Auguirregabiria and Mira (2010). One observation's mileage jump exceeded the maximum, as noted in the AM10 column. For the replication this observation was adjusted to the maximum mileage.

Table 3 starts with the same continuous data but uses 175 discrete bins instead of 90. Since the width of the bins is smaller, the maximum jump increases from 2 to 3 bins. However, a larger number of points end up past the maximum jump. In the AM10 direct calculation 3 and 11 observations go beyond the $j = 3$ jump in the two bus groups (Panel A and B). The replication code must adjust these values for consistent likelihood calculations. Comparing the replication to the original, the differences in likelihoods and cell percentages are larger than the corresponding values in Table 2. There is no equivalent to Table V for the 175 bin specification, and the tables of structural estimates differ slightly. In particular, Table IX reports the two free values $\theta_{30}$ and $\theta_{31}$ with the implied value of $\theta_{31}$ not shown. However, Table X reports four values that sum almost to 1.0 based on the four digits reported. Whether the estimated model included a fifth category with very small cell counts (consistent with the format of Table IX) or all four parameters were reported and sum to 1 cannot be determined from the available information.

**Table 3. Reconstructing Discretized Mileage Data (175 bins)**

**A. Column 1 (Groups 1-3)**

| | Replicated | | Original | | AM10 |
|---|---|---|---|---|---|
| **j** | **N** | **%** | **%** | **Δ** | **N** |
| 0 | 377 | .0976 | .0937 | .0039 | 377 |
| 1 | 1713 | .4433 | .4475 | -.0042 | 1715 |
| 2 | 1718 | .4446 | .4459 | -.0013 | 1721 |
| 3 | 56 | .0145 | .0129 | .0016 | 48 |
| 4+ | | | | | 3 |
| Obs | 3864 | | 3864 | | 3864 |
| lnL | -3900.72 | | -3856.01 | | |

**B. Column 2 (Group 4)**

| | Replicated | | Original | | AM10 |
|---|---|---|---|---|---|
| **j** | **N** | **%** | **%** | **Δ** | **N** |
| 0 | 519 | .1209 | .1191 | .0018 | 519 |
| 1 | 2453 | .5715 | .5762 | -.0047 | 2455 |
| 2 | 1228 | .2861 | .2868 | -.0007 | 1233 |
| 3 | 92 | .0214 | .0179 | .0035 | 75 |
| 4+ | | | | | 11 |
| Obs | 4292 | | 4292 | | |
| lnL | -4359.39 | | -4297.56 | | |

See notes to Table 2, except sources are Table VI and Table X in the original article. In the replication 11 observations were encountered in which the mileage exceeded the maximum possible, as shown in the AM10 column. These cases were adjusted to the maximum and subsequent mileage jumps for the next month was based on this.

## 3.1 Likelihood

Let a bus-month observation be denoted $Y \equiv (d \quad x)$. The full state and action vectors of the Rust model are both observed. Only the IID $\zeta$ shock is unobserved since it is integrated out to smooth choice probabilities. This data is (automatically) categorized in niqlow as a Full Information sample.

The estimated parameter vector is denoted $\hat{\psi}$. A sequence of outcomes for a single bus over $\hat{T}+1$ months is denoted $\{Y\}_{s=0}^{\hat{T}}$. The likelihood of a single sequence has the form:

$$L\left(\hat{\psi};\{Y\}\right) = \prod_{s=0}^{\hat{T}} \left\{P^{\star}(d_s;x_s)[P(x_{s+1};d_s,x_s)]^{I\{s<\hat{T}\}}\right\}. \tag{2}$$

This function is generated automatically by niqlow from the model and the data. The general form has been specialized to this model for clarity. As in all "structural estimation" the observed data are inserted into the theoretical probabilities produced by the fixed point in the value function. The first term is the optimal conditional choice probability based on the agent's information up to the smoothing shock $\zeta$. With Extreme Value shocks, a binary decision and a single state variable the form of $P^{\star}$ is the familiar:

$$P^{\star}(d;x) = \frac{e^{v(d,x)}}{e^{v(0;x)} + e^{v(1;x)}}. \tag{3}$$

Here $v(d,x) = U(d;x) + \delta EV(x';d,x)$ is the value of the observed action at the state $x_s$ net of the smoothing shocks which have been integrated out. The second term in (2) is the observed state transition which applies before the last observation on the path. In this model the transition is simply the trinomial probability of $j = x_{s+1} - x_s(1-d_s)$, the estimated parameter $\theta_{3j}$.

Equation (14.5) in Rust (1987) differs in two ways from (2). Translating terms and counting from 0 instead of 1, that equation would be written

$$L^{1987}\left(\hat{\psi};\{Y\}\right) = \prod_{s=1}^{\hat{T}} \{P^{\star}(d_s;x_s)\left[P\left(\theta_s;d_{s-1},x_{s-1}\right)\right]\}. \tag{4}$$

That is, (4) removes the first choice probability at $s=0$. By coincidence the effect is negligible. The estimated probability of not replacing an engine in most first observed months is very close to 1.0, because engines are never replaced until many months into service. Thus near their maxima the difference between the log-likelihoods (3) and (4) is $\approx log(1) = 0.0$.

The second difference is also inconsequential for this case. (3) multiplies the choice probability at $s$ by the transition to the current state $x_s$ not the transition from current state to the next state. This incorrect pairing is irrelevant since they are interchangeable in (3). However, it is standard now to estimate models with permanent or transitory unobserved states and actions. In this case the likelihood must add across unobserved states as well as multiply across time. Now the correct transition must multiply each conditional choice probability, which is what niqlow automatically does for such models.

## 3.2 Estimation and Inference

Having verified from Figure 1 that niqlow replicates the solution of the DP problem at estimated parameter values, and noting a small discrepancy between the original summary statistics and the replicated sample, now consider estimation. In niqlow this is accomplished using built-in tools applied to the model and the data that apply to any DP model of a general class.

Model 11 in Rust (1987) was re-estimated on Group 4 buses for the two discount factors of interest. The final values of the likelihood were collected and likelihood ratio statistics computed. Table 4 summarizes the reported and replicated values. The likelihood values are within 1% of the original values. This is consistent with the difference being explained by the coding discrepancy. And since the coding of the data appears to be consistent with other data, it appears that the original conversion code had some aspect that caused a shift of 10 $j$ values.

Note that the differences in log-likelihoods is approximately the same for $\delta = 0$ as for $\delta = 0.9999$. In the former case convergence to the fix point is immediate, so any differences in tolerances or other aspects of the Newton-Kantorovich implementation would have no effect on that case.

**Table 4. Replicating Empirical Results in Rust 1987**

**A. Estimates and standard errors for $\delta = .9999$.**

| Parameter | Original | Replicated |
|---|---|---|
| RC | 10.0750 | 10.075 |
| | (1.582) | (0.4942) |
| $\theta_1$ | 2.2930 | 2.293 |
| | (0.639) | (0.1200) |
| $\theta_{30}$ | 0.3919 | 0.3942 |
| | (.0075) | (0.0085) |
| $\theta_{31}$ | 0.5953 | 0.5930 |
| | (.0075) | (0.0090) |

**B. Likelihood ratio test for null of $\delta = 0$.**

| **δ** | **Original lnL** | **Replicated lnL** |
|---|---|---|
| .9999 | -3304.155 | -3308.4671 |
| 0 | -3306.028 | -3310.0168 |
| Like Ratio: | 3.746 | 3.0994 |
| p-value | 0.053 | .078 |

Source of Original: Rust (1987), Table IX Group 4 Estimates and author's calculation. Replicated standard errors using BHHH are reported in parentheses.

The original p-value goes from 0.053 to 0.078 using the reconstructed discrete mileage data. Thus the conclusion that Harold Zurcher was not myopic only survives under replication when 10% is the level of significance.

# 4. Conclusion

Despite some discrepancies in the original analysis, Harold Zurcher's decisions to replace bus engines do not appear to be myopic as modeled in Rust (1987). The claim of "clear rejection" made in the original has weaker support, because even using $\alpha = 5\%$ leads to a failure to reject the myopic hypothesis. For other bus groups and the 175 bin discretization of the mileage data, original conclusions may be reversed.

Perhaps more important is the fact that the original model and estimation procedure have been independently replicated within a modern open-source platform. This reduces the cost of replication existing empirical DP results, and it makes modifying existing models for both teaching and research purposes substantially easier.

# References


[1] Aguirregabiria, Victor and Pedro Mira 2002. "Swapping The Nested Fixed Point Algorithm: A Class Of Estimators For Discrete Markov Decision Models," *Econometrica 70*, 4, 1519-43.

[2] ---- 2010. "Dynamic Discrete Choice Structural Models: A Survey," *Journal of Econometrics*, 156, 1 (May), 38-67. Data archive: https://sites.google.com/view/victoraguirregabiriaswebsite/computer-code, (accessed November 2020).

[3] Ferrall, Christopher 2020. "Object Oriented (Dynamic) Programming: Replication, Innovation and 'Structural' Estimation ," Queen's Working Paper 1432, https://www.econ.queensu.ca/research/working-papers/1432.

[4] Rust, John 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", *Econometrica 55*, 5 (September), 999-1033.

[5] ----- 2000. "Nested Fixed Point Algorithm Documentation Manual," version 6, https://editorialexpress.com/jrust/nfxp.pdf (accessed November 2020).