Received: 5 October 2021 | Revised: 24 April 2023 | Accepted: 13 May 2023

DOI: 10.1002/jae.3001

WILEY Journal of APPLIED ECONOMETRICS

REPLICATION

Was Harold Zurcher myopic after all? Replicating Rust's engine replacement estimates

Christopher Ferrall

Queen's University, Kingston, Ontario, Canada

Correspondence
Christopher Ferrall, Queen's University, Kingston, ON, Canada.
Email: ferrallc@queensu.ca

Summary

Rust (1987) studies the dynamic decision making under uncertainty made by Harold Zurcher to replace bus engines. In the decades since, the model has been applied, extended, and used as an example multiple times. This paper resolves some discrepancies in how data were transformed in the original and subsequent archives. Using a package that standardizes computation of estimated dynamic programming, it replicates the 12 original maximum likelihood estimates and the six main hypothesis tests of whether Zurcher's decisions were myopic or not. The discrepancy in the data processing results in modest differences in estimates and log-likelihoods, but the *p*-values are essentially the same because the differences are very similar across values of the discount factor in the tests.

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.

Gauss code and data released in the 1990s are still available (Rust, 2000), and several papers have replicated elements of the original analysis. In particular, Larsen et al. (2012) report replications of some original results using Matlab code while testing the validity of assuming the extreme value distribution for choice errors. Aguirregabiria and Mira (2002) use the same data and provide code for their pseudo MLE procedure but do not report replications of Rust's results. Su and Judd (2012) conduct Monte Carlo experiments on their method using Rust's point estimates but do not replicate the estimates nor use the original data. The software package `ruspy` (Blesch, 2019) implements both the original and Su and Judd solution techniques. Müller and Reich (2021) use the original data and estimate the original model but use a different technique and do not report agreement with the original maximum likelihood estimation.

This paper's modest goal is to replicate MLE estimates in Rust (1987) in order to demonstrate the software package `niqlow Documentation` (n.d.) described in Ferrall (2022). The replication is based not on bespoke code but rather high-level statements in `niqlow`. In particular, 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 reproduces both the MLE estimates of the parameter values and the results of these hypothesis tests.

Despite a 35-year gap in computing hardware and software, and a fundamentally different approach to coding the model, the solutions of the DP model are identical for a given set of parameter values. The MLE estimates do differ with the original values because of the difference in the transformation of the original raw data mentioned above. Using the data files in Larsen et al. (2012), an error in the discretization of the data was resolved. As with their analysis, re-estimation with the original data encoding results in identical estimates and test statistics. Using the data produced by the corrected encoding procedure results in different estimates. However, the shifts in likelihood are similar in the myopic and forward-looking versions of the models, so test statistics are essentially the same.

2 | MODEL AND ITS SOLUTION

The terminology here follows conventions in `niqlow` and differs from the original paper. Table 1 shows the difference in notation, which is followed, so details can be compared directly with other work implemented in `niqlow`. In the model, bus maintenance is a stationary environment with a binary decision to replace an engine ($d = 1$) each month. This resets mileage x on the engine to $x = 0$. For estimation, the odometer readings are reduced to two different number of discrete possibilities, $N = 90$ and $N = 175$. The transition next month takes on one of J values:

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

The number of feasible jumps required to explain the data was $J = 3$ for $N = 90$ and $J = 5$ for the $N = 175$. Other elements of the model are the monthly discount factor δ, utility $U()$, and an implicit vector of additive extreme value shocks to smooth conditional choice probability (CCP).

Panel B of 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. This objected-oriented approach makes it possible to build specialized and extended versions of existing models replacing only the code or features that differ.

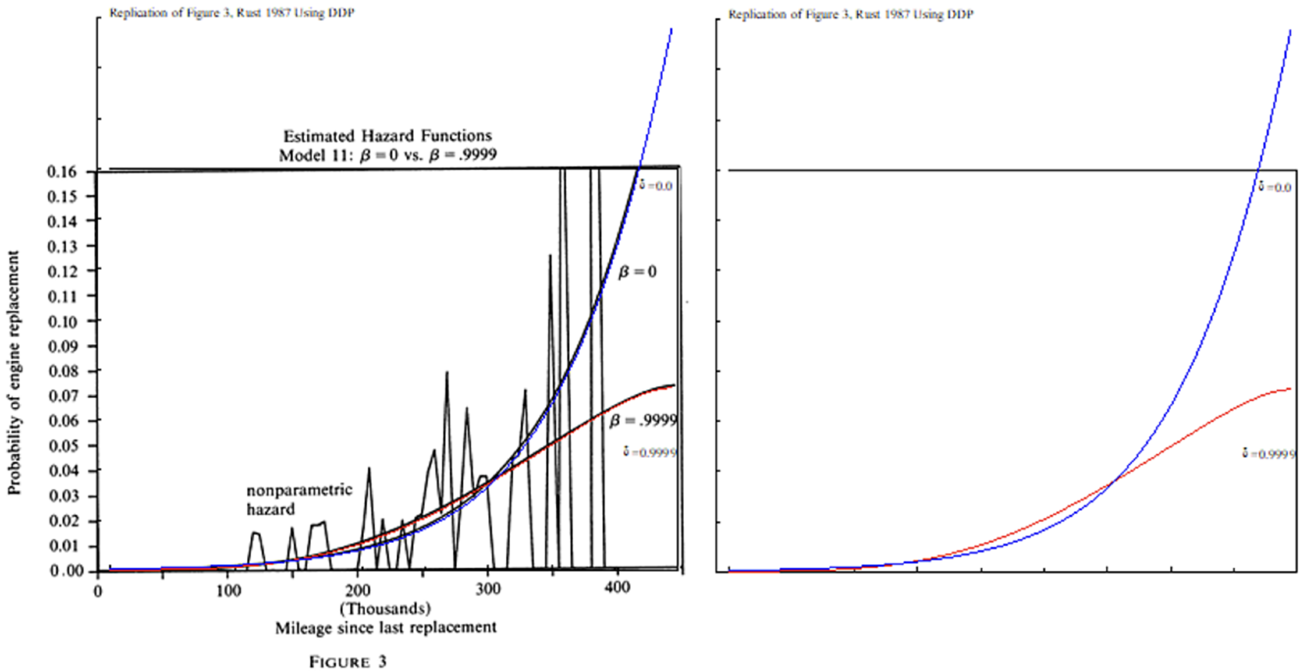
2.1 | Choice probability replication

Before turning to the ML estimation procedure, consider output at one of the estimated models in the original paper. Fig. 3 in the paper compares $P^\star(1;x)$, the replacement probability, for discount factors of $\delta = 0$ and $\delta = 0.9999$. This does not involve the data. It is a check on the recoding of the Newton-Kantorovich algorithm, described in Rust (1988), to solve the fixed point in the value function along with the logit specification for the CCPs. Figure 1 displays the original figure

TABLE 1 Rust (1987) model summary.

A. Crosswalk with original notation			
Term	**Original**	**`niqlow`**	**Notes**
Discount factor	β	δ	$\delta\epsilon\{0, 0.9999\}$
Mileage bin	x	x	$N = 90$ or 175
Decision	i	d	Replace or not
Value shock	$\varepsilon(i)$	ζ_α	Additive extreme value
Innovation	j	j	$x' = (1-d)x + j.$
Parameter vector	θ	ψ	
Time index	t	s	For observed outcomes
CCP	$P(i\|x,\theta)$	$P^\star(\alpha;\theta)$	
Transition	$p(x_t+1\|x_t,i,\theta_3)$	$P(\theta';\alpha,\theta)$	State-to-state with optimal choice

B. `niqlow` Definition of the model		
Element	**Value**	**Category/details**
Clock	t	Ergodic
Value shock	ζ	Extreme value
Actions	$\alpha = (d)$	Binary choice
States	$\theta = (x)$	Renewal with probabilities θ_3
Utility	$U(\alpha,\theta) = -\begin{pmatrix} \theta_1 x/1000 \\ RC \end{pmatrix}$	
Parameters	$\psi = (\delta, RC, \theta_1, \theta_3)$	



Left: original printed Figure 3 digitized with the replicated replacement probability curves overlayed (after matching the axes scales). Right: the replication shown alone.

FIGURE 1 Replicating rust fig. 3.

and the replication. The additional information in the original figure is the empirical hazard rate. The replication figure is shown alone on the right and is overlayed on the original scanned figure demonstrating they coincide exactly given the scan's resolution.

3 | DATA AND ESTIMATES REPLICATION

3.1 | Observed mileage transitions

The original article estimates the linear cost specification on the two mileage bin counts described above and three different engine groups. The odometer was not reset when an engine was replaced so the raw data are not consistent with the model. For example, an odometer reading of 320,500 miles in month s after a bus replacement at 310,300 miles would be changed to 10,200 miles. This is converted to a discrete value, which depends on N and the maximum mileage of 450,000. For $N = 90$, it would be changed to $x_s = 2$ because it is the third discretized bin of size 5000 miles. Resetting and discretizing 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, so it was not possible to look at the coded mileage data directly. However, the code and generated data files from Larsen et al. (2012) replicate the original code and agree with it in terms of summary statistics. Another modern source of the continuous and discretized data is the Rust Data Repository (n.d.) at OpenSourceEconomics. All versions use the recorded mileage at replacement to determine the month it occurred. The month of replacement is separately recorded, and in some cases, there is a 1-month discrepancy between this information.

The transition or innovation $j = x_{s+1} - x_s(1 - d_s)$ is constructed from the path of the discretized data x_s. The transitions are IID. Both x and d are observed, so the transition probabilities θ_3 form a multinomial problem not involving optimal replacement decisions (Rust's first-stage estimation). Tables 2 and 3 report the estimates of θ_3 (fractions of transitions j) for bus groups 1–3 and group 4 from the original paper for 90 and 175 bins, respectively. The estimated transitions are compared with the values from the replication data.

It is in these transitions that a small discrepancy occurs. Both the original Gauss code and the Matlab code used in Larsen et al. (2012) compute the discrete x_s using the ceiling function when it should be the floor function. For months

TABLE 2 Reconstructing discretized mileage data ($N = 90$) bins.

	Replicated			Original			
j	***N***		**%**	***N***		**%**	**Δ**
A. Column 1 (Groups 1–3)							
0	1189		0.3077	1162		0.3007	+27
1	2635		0.6819	2662		0.6889	−27
2	40		0.0104	40		0.0104	0
Obs		3864			3864		0
lnL		−2592.90			−2570.97		21.93
B. Column 2 (Group 4)							
0	1715		0.3996	1682		0.3919	+33
1	2522		0.5876	2555		0.5953	−33
2	55		0.0128	55		0.0128	0
Obs		4292			4292		0
lnL		−3153.83			−3140.57		13.26

Note: The third bus group pools A and B and is not reported here.*Source*: (Original) Author's calculation of Rust (1987) tab. IX using data from Larsen et al. (2012), which match the original. (Replicated) Author's calculation from the same data but correcting j after replacements.

TABLE 3 Reconstructing discretized mileage data ($N = 175$ bins).

	Replicated			Original			
j	***N***		**%**	***N***		**%**	**Δ**
A. Column 1 (Groups 1–3)							
0	385		0.0976	36		0.0937	+23
1	1710		0.4425	1729		0.4475	−19
2	1719		0.4449	1723		0.4459	−4
3	49		0.0127	49		0.0127	0
4+	1		0.0003	1		0.0003	0
Obs		3864			3864		0
lnL		−3896.51			−3861.00		35.51
B. Column 2 (Group 4)							
0	539		0.1256	511		0.1191	+28
1	2449		0.5706	2472		0.5760	−23
2	1225		0.2854	1230		0.2866	−5
3	70		0.0163	70		0.0163	0
4+	9		0.0021	9		0.0021	0
Obs		4292			4292		0
lnL		−4371.93			−4331.72		40.21

Note: See notes to Table 2, except sources are tab. VI and tab. X in the original article.

without a bus replacement, this has no effect because when $d_s = 0$, the transition $j = x_{s+1} - x_s$ is the same regardless if both values are increased by one. However, in replacement months, this results in $j = x_{s+1}$, and this had minimum value 1 instead of 0. (It is not surprising that in months when an engine is replaced, the added mileage is within the lowest mileage category which never occurred in the incorrect coding.) The first-stage transition probabilities are therefore incorrect, which affects the second state estimates through θ_3.

These differences are reported in final column labeled Δ in Tables 2 and 3. In the four different samples, the multinomial likelihood on the transition probabilities differs by roughly 1% 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, these differences will shift all MLE estimates.

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 two to five bins. Comparing the replication to the original, the differences in likelihoods and cell percentages are somewhat larger than the corresponding values in Table 2. Multiplying the percentage differences by the number observations yields roughly 23 bus-month values are misclassified in each panel.

3.2 | Likelihood

The estimated parameter vector is denoted $\hat{\psi}$. Let a bus-month observation be denoted $Y \equiv \begin{pmatrix} d & x \end{pmatrix}$. The full state and action vectors of the Rust model are both observed. Only the ζ shock is unobserved since it is integrated out to smooth choice probabilities. In `niqlow`, panel data are read in and matched to the states and actions present in the model. This kind of data is automatically categorized as a Full Information because no actions or states are hidden.

A sequence of outcomes for a single bus over $\hat{T}+1$ months is denoted $\{Y\}_{s=0}^{\hat{T}}$, where $Y_s = \{\alpha_s, \theta_s\}$ is simply the decision and state vectors combined. In `niqlow`, these are in general vectors, but in this model, they reduce to two scalars (d_s, x_s). The likelihood of a single sequence has the form:

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

That is, the contribution in observed month s is the CCP for the observed decision times the probability of the observed transition to the state next period. The transition is not observed in the last observed month for a bus. 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. In this model, the transition probability is simply the multinomial parameter θ_{3j}.

Eq. (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

$$L^{1987}(\hat{\psi}; \{Y\}) = \prod_{s=1}^{\hat{T}} \{P^{\star}(\alpha_s; \theta_s)[P(\theta_s; \alpha_{s-1}, \theta_{s-1})]\}. \tag{3}$$

That is, (3) removes the first choice probability at $s = 0$. The effect is coincidentally negligible. The estimated probability of not replacing an engine in most first observed months is very close to 1.0, because engines are only replaced many months into service. Thus, near the maximum, the difference between the log-likelihoods (2) and (3) is $\approx log(1) = 0.0$.

TABLE 4 Replicating Table IX of Rust (1987) (third-stage FIML).

	Column 1 Groups 1–3		Column 2 Group 4		Column 3 Groups 1–4	
A. $\delta = 0.9999$.						
Param.	Orig.	Repl.	Orig.	Repl.	Orig.	Repl.
θ_{30}	0.30100	0.3077	0.39190	0.3996	0.3489	0.3561
	(0.0075)	(0.0068)	(0.0075)	(0.0089)	(0.0052)	(0.0040)
θ_{31}	0.6884	0.6819	0.5953	0.5876	0.6394	0.6323
	(0.0074)	(0.0066)	(0.0075)	(0.0095)	(0.0053)	(0.0041)
RC	11.727	11.7270	10.075	10.0750	9.7558	9.7558
	(2.6020)	(2.7377)	(1.5820)	(1.6512)	(1.2270)	(1.2361)
θ_1	4.8259	4.8259	2.293	2.2930	2.6275	2.6275
	(1.7920)	(1.8380)	(0.6390)	(0.6594)	(0.6180)	(0.5782)
ln L	−2708.37	−2725.28	−3304.16	−3317.41	−6055.25	−6086.07
B. $\delta = 0$.						
Param.	Orig.	Repl.	Orig.	Repl.	Orig.	Repl.
θ_{30}	0.30100	0.3077	0.3919	0.3996	0.3489	0.3561
	(0.0074)	(0.0068)	(0.0075)	(0.0089)	(0.0052)	(0.0040)
θ_{31}	0.68840	0.6819	0.5953	0.5876	0.6394	0.6323
	(0.0075)	(0.0066)	(0.0075)	(0.0095)	(0.0053)	(0.0041)
RC	8.2985	8.2985	7.6538	7.6538	7.3055	7.3055
	(1.0417)	(1.1257)	(0.7197)	(0.8080)	(0.5067)	(0.5459)
θ_1	109.903	109.9031	71.5133	71.5133	70.2769	70.2769
	(26.163)	(27.2148)	(13.778)	(15.0571)	(10.750)	(10.5554)
ln L	−2710.75	−2727.66	−3306.03	−3319.31	−6061.64	−6092.54
C. Likelihood ratio test for null of $\delta = 0$. (Myopia test)						
LR	4.76	4.76	3.74	3.79	12.78	12.94
p-value	0.029	0.029	0.053	0.052	0.0004	0.0003

Note: Standard errors reported in parentheses.*Source*: (Original) Rust (1987), tab. IX and author's calculation using Newton-Kantorovich and Newton/BHHH optimization.

The second difference is also inconsequential for this case. Equation (2) 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 (2). However, for models with permanent or transitory unobserved states and actions, they are not interchangeable. The likelihood should add across unobserved states as well as multiply across time, so the conditional choice must be multiplied by the correct transition before summing as in (2), which `niqlow` does automatically for such models.

3.3 | Estimation

Having verified from 1 that `niqlow` replicates the solution of the DP problem at estimated parameter values, and noting the discrepancy between the original and replicated samples, now consider second-stage estimation of the structural parameters. 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. Originally, the model was estimated in two stages to reduce computation. The replication code does the same using `niqlow`'s ability to mark estimated parameters as either "transition only" or "utility only." However, only the third-stage estimates, when all parameters are estimated jointly, are reported below.

The ML estimates were reported in original tab. IX ($N = 90$) and tab. X ($N = 175$). These models share the linear cost of bus maintenance already shown, and three different bus type samples (Groups 1–3, Group 4, and Groups 1–4 combined). The original estimates and the replications for the first two groups are reported in Tables 4 and 5. The values

TABLE 5 Replicating Table X of Rust (1987) (third-stage FIML).

	Column 1 Groups 1–3		Column 2 Group 4		Column 3 Groups 1–4	
A. $\delta = 0.9999$.						
Param.	Orig.	Repl.	Orig.	Repl.	Orig.	Repl.
θ_{30}	0.0937	0.0996	0.1191	0.1256	0.1071	0.1133
	(0.005)	(0.0025)	(0.005)	(0.0055)	(0.0034)	(0.0024)
θ_{31}	0.4475	0.4426	0.5762	0.5706	0.5152	0.5099
	(0.008)	(0.0074)	(0.008)	(0.0072)	(0.0055)	(0.0036)
θ_{32}	0.4459	0.4449	0.2868	0.2854	0.3621	0.3609
	(0.008)	(0.0077)	(0.007)	(0.0059)	(0.0053)	(0.0028)
θ_{33}	0.0127	0.0127	0.0158	0.0163	0.0143	0.0146
	(0.002)	(0.0018)	(0.002)	(0.0017)	(0.0013)	(0.0012)
RC	11.7257	11.7361	10.896	10.0982	9.7687	9.7802
	(2.597)	(2.7486)	(1.581)	(1.6645)	(1.226)	(1.2251)
θ_1	2.4569	2.4520	1.1732	1.1702	1.3428	1.3410
	(0.912)	(0.9459)	(0.327)	(0.3355)	(0.315)	(0.2933)
In L	−3993.99	−4029.11	−4495.14	−4535.65	−8607.89	−8683.84
B. $\delta = 0$.						
Param.	Orig.	Repl.	Orig.	Repl.	Orig.	Repl.
θ_{30}	0.0937	0.0996	0.1191	0.1256	0.1071	0.1133
	(0.005)	(0.0026)	(0.005)	(0.0055)	(0.003)	(0.0024)
θ_{31}	0.4475	0.4425	0.5762	0.5706	0.5152	0.5099
	(0.008)	(0.0074)	(0.008)	(0.0073)	(0.006)	(0.0036)
θ_{32}	0.4459	0.4449	0.2868	0.2854	0.3621	0.3610
	(0.008)	(0.0077)	(0.007)	(0.0059)	(0.005)	(0.0028)
θ_{33}	0.0127	0.0127	0.0158	0.0163	0.0143	0.0146
	(0.002)	(0.0018)	(0.002)	(0.0017)	(0.014)	(0.0012)
RC	8.2969	8.3071	7.6423	7.6483	7.3113	7.3213
	(1.048)	(1.1364)	(0.720)	(0.8451)	(0.507)	(0.5537)
θ_1	56.1657	56.2896	36.6692	36.7245	36.0175	36.1160
	(13.421)	(14.0936)	(7.068)	(7.9807)	(5.515)	(5.4576)
In L	−3996.35	−4031.50	−4497.00	−4537.52	−8614.24	−8690.27
C. Likelihood ratio test for null of $\delta = 0$. (Myopia test)						
LR	4.72	4.77	3.72	3.76	12.7	12.84
p-value	0.030	0.029	0.054	0.053	0.0004	0.0003

Note: See Table 4 except source of original is Rust (1987), tab. X.

reported are "third-stage" estimates in which all parameters are estimated including the first-stage transition probabilities. Standard errors are also reported. The original procedure used BHHH and semi-closed forms for the derivatives of the value function with respect to structural parameters (derived in Rust, 1988). The replication uses BHHH and numerical central step derivatives. For $\delta = 0.9999$, the convergence in the stage 3 estimates is weak: The norms of gradients remain large but progress ends. This combines large differences in the value function (which enter exponents) and precision in the convergence in the Newton-Kantorovich.

The likelihood values are within 1% of the original values. This magnitude is consistent with the exact match with choice probabilities for the same parameter values in Figure 1 and the coding discrepancy in the values of j. Also note that the differences in log-likelihoods are 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 have no effect on that case. Standard errors are all similar in magnitude between the original and replicated values.

3.4 | Inference: replicating the myopia test

One way to aggregate the overall effect of the data discrepancies and their effect on the estimates is to recompute the likelihood ratio test of the myopia restriction (panel B versus panel A in each table, one degree of freedom). The results are reported in panel C of the tables. The p-values are essentially the same between the original and replicated results. The myopic and forward-looking versions share the same transition data (j) and the effects on the distribution of j is almost a constant shift in log-likelihood resulting in nearly identical LR statistics. This is likely because mileage x and subsequent costs make modest changes month-to-month relative to the large reset when replacement occurs. Shifting the distribution of j has little effect on the replacement dynamics.

4 | CONCLUSION

Despite some discrepancies in the original analysis, a complete re-analysis of the main estimates from Rust (1987) is replicated. The claim of "clear rejection" of myopic decision making in the original paper survives the effect of these changes

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.

ACKNOWLEDGMENTS

I thank Brad Larsen for providing his code and data files, John Rust for his responses to many inquiries, and Victor Aguirreibiria for suggestions on an earlier version.

OPEN RESEARCH BADGES

This article has been awarded Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results.

DATA AVAILABILITY STATEMENT

Online appendix is found at https://doi.org/10.15456/jae.2023160.1758358723.

REFERENCES

Aguirregabiria, V., & Mira, P. (2002). Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models. *Econometrica*, *70*(4), 1519–43.

Blesch, M. (2019). Ruspy software package. Retrieved January 2022 from https://ruspy.readthedocs.io/en/latest/index.html

Ferrall, C. (2022). Object oriented (dynamic) programming: Closing the 'structural' estimation coding gap. *Comput. Econ.* https://doi.org/10.1007/s10614-022-10280-4

Larsen, B. J., Oswald, F., Reich, G., & Wunderli, D. (2012). A test of the extreme value type I assumption in the bus engine replacement model. *Econ. Lett.*, *116*(2), 213–216.

Müller, P., & Reich, G. (2021). Harold Zurcher and Paul Volcker—The minutes of a meeting that never was (October 21, 2021). Available at SSRN: http://dx.doi.org/10.2139/ssrn.3303999

niqlow Documentation (n.d.). Retrieved March 2022 from https://ferrall.github.io/niqlow/

Rust Data Repository (n.d.). Open source economics. Retrieved March 2022 from https://github.com/OpenSourceEconomics/rust-data

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, *55*, 999–1033.

Rust, J. (1988). Maximum likelihood estimation of discrete control processes. *SIAM J. Control Optim.*, *26*(5), 1006–24.

Rust, J. (2000). Nested fixed point algorithm documentation manual. version 6. Retrieved November 2020 from https://editorialexpress.com/jrust/nfxp.pdf

Su, C.-L., & Judd, K. L. (2012). Constrained optimization approaches to estimation of structural models. *Econometrica*, *80*(5), 2213–2230.

How to cite this article: Ferrall, C. (2023). Was Harold Zurcher myopic after all? Replicating Rust's engine replacement estimates. *Journal of Applied Econometrics*, 1–8. https://doi.org/10.1002/jae.3001