**JASA ACS Reproducibility Initiative - Author Contributions Checklist Form**

The purpose of the Author Contributions Checklist (ACC) Form is to document the code and data supporting a manuscript, and describe how to reproduce its main results.

As of Sept. 1, 2016, the ACC Form must be included with all new submissions to JASA ACS.

This document is the initial version of the template that will be provided to authors. The JASA Associate Editors for Reproducibility will update this document with more detailed instructions and information about best practices for many of the listed requirements over time.

## Data

**Abstract (Mandatory)**

We include three types of data in our submission. First, links to original raw data (public) as well as the scripts needed to download and process the data are included. Second, we include the results of this processing as R objects with accompanying documentation as part of our package. Third, because estimation requires a fair amount of compute time, we include the results of parameter estimation as an R object.

**Availability (Mandatory)**

Original data is publicly available (see below). Intermediate inputs (the second type above) are included with the submission and wrapped into an R package (currently on github).

**Description (Mandatory if data available)**

Permissions: Original raw data is available from the CHARM Mazurka Project (<http://charm.rhul.ac.uk/projects/p2_3.html>). We have cited this as the source of our data in the manuscript as required by CHARM.

Licensing information: None.

Link to data:

* Scripts to grab and process raw data: <https://github.com/dajmcdon/dpf/tree/master/data-raw>
* Processed data for analysis: <https://github.com/dajmcdon/dpf/tree/master/data>
* Estimated parameters: <https://github.com/dajmcdon/dpf/blob/master/extras/mazurkaResults.Rdata>

Data provenance: Original (raw) data is scraped from <http://mazurka.org.uk/auto/earis/mazurka68-3/> using the scripts mentioned above.

File format: Processed data and estimated parameters are R data frames.

Metadata: Processed data is fully documented within the package in standard style following Wickham <http://r-pkgs.had.co.nz/data.html>

Version information: none

**Optional Information (complete as necessary)**

Unique identifier / DOI

## Code

**Abstract (Mandatory)**

We are including 3 types of code. The main functions are contained in a fully documented R package. We use this package to produce our results, but the functions are usable for other switching state space models. We also provide 3 R scripts that use the package to perform all analyses. Finally, we provide an R markdown file which regenerates all figures and data tables in the main paper as well as the supplementary material. The supplement itself is the knitted pdf generated by this Rmd file.

**Description (Mandatory)**

How delivered: R package, Collection of 3 R scripts, Rmarkdown file. Each was submitted separately as a zip file upon manuscript submission. Direct links are given below as requested.

Licensing information: MIT License

Link to code/repository:

* R package: <https://github.com/dajmcdon/dpf>
* R scripts for parameter estimation: <https://github.com/dajmcdon/dpf/tree/master/extras> (the resulting estimates discussed in the data section above are also in this directory).
* Rmarkdown for supplement and figures/tables: File “music-supplement.Rmd” at <https://github.com/dajmcdon/dpf/tree/master/manuscript/>
* Version information (e.g., for a Git repository, the number or branch+commit)

**Optional Information (complete as necessary)**

Hardware:

To run the estimation procedure, we use the “batchtools” R package on a cluster. This package must be setup for the estimation to reproduce correctly. However, this package is cluster independent (it will run our procedure unchanged on a single machine if configured for a single machine). Given a correct setup of that package, our code should run on any configuration. We used version 0.9.11 of batchtools.

## Instructions for Use

**Reproducibility (Mandatory)**

To reproduce all tables, figures, and the entire Supplementary Material:

1. Clone the github repo and launch R from the root directory of the repository.
2. Using `devtools::install\_github(‘dajmcdon/dpf’)` install the R package.
3. Using R with `knitr` installed (e.g. with Rstudio) knit “manuscript/music-supplement.Rmd” Note: some tex style files are not included, changing these as appropriate will not affect figure generation.

To re-estimate all the parameters using package data (see the README in the “extras” directory) :

1. Clone the github repo and launch R from the “extras” directory of the repository.
2. Using `devtools::install\_github(‘dajmcdon/dpf’)` install the R package.
3. Install and configure “batchtools” via CRAN (see also <https://mllg.github.io/batchtools/articles/batchtools.html>)
4. Run the script “cluster\_analysis\_my\_model.R”. (as noted in that script, this takes ~5 hours per performance).
5. Upon completion, run the script “collect\_cluster\_results.R”. This produces the file “mazurkaResults.Rdata” which is the only input to the tables, figures, and Supplementary Material above.

**Replication (Optional)**

The “dpf” package contains a vignette. If the repo is simply cloned, it is available in the “vignettes/” directory. If installed, devtools 2.0.0 seems to require installing with

`devtools::install\_github(‘dajmcdon/dpf’, build\_opts = c("--no-resave-data", "--no-manual"))`.

Once installed, `browseVignettes(“dpf”)` will display the pdf and source.