Online Appendix for "A structural meta-analysis of welfare reform experiments and their impacts on children"

Joseph Mullins

October 11, 2024

A Details of Model Solution

The additive representation can be shown by taking the form of the solution as given, and showing that the recursion preserves this relationship. Note that in the terminal period, the value function is simply $V_T(\theta) = (1-\beta)^{-1}\alpha_{\theta,k}\log(\theta)$ and so this form holds at T. We have:

$$V_{k,t}(\chi,\theta) = \mathbb{E}_{\epsilon}\max_{C,I,j}\left\{U_{kj}(C,\theta) + \epsilon_j + \beta\mathbb{E}_{\chi'|\chi,\theta}[\tilde{V}_{k,t+1}(\chi')] + \beta\alpha_{V,t+1,k}(\delta_I\log(I) + \delta_\theta\log(\theta))\right\}$$

where we have written θ' already in terms of I and θ. Fixing the discrete choice j fixes net income, $Y_{kj}(\chi)$ and conditional on this choice the optimal investment problem has a simple solution:

$$p_{H_j+F_j}I = \frac{\beta\delta_I\alpha_{V,t+1,k}}{1+\beta\delta_I\alpha_{V,t+1,k}}Y_{kj}(\chi)$$

which can be substituted into the problem above to get:

$$V_{k,t}(\chi,\theta) = \mathbb{E}_{\epsilon}\max_{j}\left\{\tilde{U}_{kj}(\chi) + \epsilon_j + \beta\mathbb{E}_{\chi'|\chi}\tilde{V}_{k,t+1}(\chi')\right\} + \alpha_{V,t,k}\log(\theta)$$

where $\alpha_{V,t,k} = \alpha_{\theta,k} + \beta\delta_\theta\alpha_{V,t+1,k}$ as in the main text and indirect utility $\tilde{U}$ takes the form in the main text. Note that $\alpha_{V,t}$ is scaled uniformly by $\alpha_{\theta,k}$ and so it can be rewritten as:

$$\alpha_{V,t,k} = \alpha_{\theta,k}\tilde{\Gamma}_t, \qquad \tilde{\Gamma}_t = 1 + \beta\delta_\theta\tilde{\Gamma}_{t+1}.$$

Accordingly, Γ_t from the main text is defined as $\Gamma_t = \delta_I\beta\tilde{\Gamma}_{t+1}$.

B Identification

This section derives the expressions that comprise the linear system in equation (2). The system is four dimensional, corresponding to five choice comparisons: (1) using paid care vs not, conditional on participation and work; (2) working vs not, conditional on participation; (3) participating in welfare vs not, assuming no time limit; (4) participating in welfare vs not, assuming a time limit; and (5) participating in food stamps vs no participation.

In what follows, it is useful to recall that the following fact about nested logit choice probabilities. Consider the log-odds ratio for choosing nest B_1 vs nest B_2:

$$\log\left(\mathbb{P}(B_1)/\mathbb{P}(B_2)\right) = \sigma_l^{-1}\left(\mathcal{V}_1 - \mathcal{V}_2\right)$$

where $\mathcal{V}_b$ is the inclusive value of choosing this nest and σ_l is the dispersion parameter for choices across nests in this layer, l. The inclusive value $\mathcal{V}_b$ can likewise be written as:

$$\mathcal{V}_b = \mathcal{V}_j - \sigma_{l+1}\log(\mathbb{P}(j \in B_b))$$

where $\mathcal{V}_j$ is the value of making choice j. This could be a final value if j is a single choice, or an inclusive value if j is another nest.

It will also be useful to define the concept of *finite dependence* (Arcidiacono and Miller 2011). A model exhibits finite dependence if for any two choices there exists a sequence of future choices that returns the individual back to the same distribution of state variables. When this is the case, the formula above can be applied iteratively until the period at which the states are in the same distribution, at which point the difference in the values of the two options differences out, and we are left with a sequence of utilities and choice probabilities.

Finite dependence is simple here as the only choice affecting future states is participation. Consider the comparison of welfare participation $P_1 = 2$ vs not $P_1 = 1$ at time t. Without time limits, individuals differ at $t+1$ in terms of lagged participation, but will return to the same state at $t+2$ with $P_1' = 1$ and $P_2' = 1$. In a world with time limits, we require a two-period sequence: $(P_1' = 1, P_1'' = 1)$ and $(P_2' = 2, P_2'' = 1)$ which will return individuals to the same state at $t+3$. These sequences will be applied below when deriving ratios of participation choice probabilities.

Child Care Choices

This first equation for child care choices is derived in the main text:

$$\log(\tilde{p}_k(F, X, Z)) = \sigma_3^{-1}\left[(1+\alpha_{\theta,k}\Gamma_t)\log\left(\tilde{Y}_{k,F}(P, X, Z)\right) + \alpha_{F,k} + \alpha_{\theta,k}\Gamma_t g_2\right] \tag{1}$$

and identification of the pair $\sigma_3^{-1}(1+\alpha_{\theta,k}\Gamma_t)$ and $(\alpha_{F,k} + \alpha_{\theta,k}\Gamma_t g_2$ can be derived from variation in Z alone or from X also. For example, differencing across Z and Z' gives:

$$\log\left(\frac{\tilde{p}_{k,F}(P, X, Z')}{\tilde{p}_{k,F}(P, X, Z)}\right) = \sigma_3^{-1}(1+\alpha_{\theta,k}\Gamma_t)\log\left(\frac{\tilde{Y}_{k,F}(P, X, Z')}{\tilde{Y}_{k,F}(P, X, Z))}\right)$$

With this slope term in hand, the age-specific term $\alpha_{F,k} + \alpha_{\theta,k}\Gamma_t$ pins down the level.

Labor Supply

As above, define

$$\tilde{p}_{k,H}(P, X, Z) = \frac{p_k(H=1|P,\chi)}{p_k(H=0|P,\chi)}$$

as the relative choice probability of working in state $\chi = (X, Z)$, conditional on participation P. It will be useful to also define the proportional return to working as:

$$\tilde{Y}_{k,H}(P, X, Z) = \frac{Y_k(H = 1, P, X, Z)}{Y_k(H = 0, P, X, Z)}.$$

For convenience let us also define:

$$p_{k,F}(P, X, Z) = \mathbb{P}[F = 1|H = 1, P, X, Z]$$

With these expressions, the nested conditional choice probability for the work choice can be written as:

$$\log\left(\tilde{p}_{k,H}(P, X, Z)\right) = \sigma_2^{-1}\left(\alpha_{H,k} + \alpha_{\theta,k}\Gamma_t g_1\right) + \\ \sigma_2^{-1}\left[(1 + \alpha_{\theta,k}\Gamma_t)\log\left(\tilde{Y}_{k,H}(P, X, Z)\right) + R(X, Z)\alpha_R A - \sigma_3 \log(1 - p_{k,F}(P, X, Z))\right] \quad (2)$$

which forms the second row of the linear system in (2).

Notice that three parameters now dictate the responsiveness of labor supply to different components of the treatment. The term $\sigma_2^{-1}(1 + \alpha_{\theta,k}\Gamma_t)$ is an age-specific semi-elasticity of labor supply with respect to financial incentives. The term $\sigma_2^{-1}\alpha_R$ determines the response to work requirements through non-pecuniary motives, and $\sigma_2^{-1}\sigma_3$ determines the responsiveness of labor supply to childcare costs. All three terms are identified as long as there is rank-independent variation in returns to work, the existence of work requirements ($R(X, Z)$), and childcare prices. In practice this is delivered by all sources of variation in X and Z, conditional on type, but in principle one could use only variation in Z, as long as there are at least three treatments that vary returns to work, work requirements, and childcare subsidies in a rank-independent way. This paper's application has four treatments that do satisfy these requirements.

Combining these slope terms with the slope terms identified in the previous stage is sufficient to invert out values for σ_1, σ_2, $\alpha_{\theta,k}\Gamma_t$, $\alpha_{F,k}$ and g_2. Returning now to (2), the parameters g_1 and $\alpha_{H,k}$ remain to pin down the level of work choice probabilities at each age and in each state.

It may help to note here that if there exists a combination of (X, Z, Z') such that the only difference in policies is the existence of mandatory services, this expression implies:

$$\log\left(\frac{\tilde{p}_{k,H}(1, X, Z')}{\tilde{p}_{k,H}(1, X, Z)}\right) = \sigma_2^{-1}\alpha_R.$$

This is exactly the case for MFIP in which treatment arms (1) and (2) differ only in the imposition of work requirements and hence will be an important source of identification for α_R.

Program Participation without Time Limits

Because participation decisions dynamically influence future states, more elaborate comparisons are required. In a slight abuse of notation, let the state χ be decomposed into (X, Z, A_{-1}) so that we can indicate values of A_{-1} exclusively. Define:

$$\tilde{p}_{k,A}(X, Z, A_{-1}) = \frac{p_k[P = 2|X, Z, A_{-1}]}{p_k[P = 1|X, Z, A_{-1}]}$$

as the probability of participating in welfare relative to participating only in food stamps. The term

$$\tilde{Y}_{k,A}(X,Z) = \frac{Y_k(H=0,P=2,X,Z)}{Y_k(H=0,P=1,X,Z)}$$

is the proportional return in net income to participating in welfare assuming no work. Finally, define:

$$\tilde{u}_{k,A}(X,Z,A_{-1}) = (1+\alpha_{\theta,k}\Gamma_t)\log(\tilde{Y}_{k,A}(X,Z)) - R(X,Z)\alpha_R \\ - \alpha_P(1-A_{-1}) - \sigma_2 \log\left(\frac{p_k(H=0|P=2,X,Z)}{p_k(H=0|P=1,X,Z)}\right) \quad (3)$$

as the difference in static payoffs from choosing $P=2$ over $P=1$. Putting these expressions together gives:

$$\log(\tilde{p}_{k,A}(X,Z,A_{-1})) = \sigma_1^{-1}\left[\tilde{u}_{k,A}(X,Z,A_{-1}) + \beta\mathbb{E}_{X'|X}(V(X',Z,A_{-1}=1) - V(X',Z,A_{-1}=0))\right].$$

A simple comparison of this log odds ratio for different values of A_{-1} identifies the application cost α_P up to scale:

$$\log\left(\frac{\tilde{p}_{k,A}(X,Z,1)}{\tilde{p}_{k,A}(X,Z,0)}\right) = \sigma_1^{-1}\alpha_P.$$

Then in order to difference out the dynamic values, we can exploit the model's finite dependence to get:

$$\log(\tilde{p}_{k,A}(X,Z,A_{-1})) = \sigma_1^{-1}\tilde{u}_{k,A}(X,Z,A_{-1}) - \beta\left[\mathbb{E}_{X'|X}\log\left(\frac{p_k(P=1|X',Z,1)}{p_k(P=1|X',Z,0)}\right)\right].$$

This forms the third row of the linear system in (2). Following identical logic to the previous steps, the response in program participation to benefit generosity is determined by σ_1 and a comparison between Z and Z' fixing the other variables will identify it. The fourth two is the relatively simpler comparison of food stamp participation with no participation, $p_{k,S}(X,Z) = p_k[P=1|X,Z]/p_k[P=0|X,Z]$:

$$\log(\tilde{p}_{k,S}(X,Z) = \sigma_1^{-1}\alpha_{S,k} + \sigma_1^{-1}(1+\alpha_{\theta,k}\Gamma_t)\log(\tilde{Y}_{k,S}(Z,A))$$

Program Participation with Time Limits

Following the same notational convention from the previous section, we can derive an expression for $\tilde{p}_{k,A}$ in the presence of time limits. First we must further decompose the state χ into: (X,Z,A_{-1},ω) where ω is cumulative welfare use. Then we get:

$$\log(\tilde{p}_{k,A}(X,Z,A_{-1},\omega)) = \sigma_1^{-1}\tilde{u}_{k,A}(X,Z,A_{-1}) + \sigma^{-1}\left[\beta\mathbb{E}_{X'|X}(V(X',Z,1,\omega+1) - V(X',Z,0,\omega))\right].$$

In this case, a two-period sequence of decisions places the agent back into the same state space, giving:

$$\log(\tilde{p}_{k,A}(X,Z,A_{-1},\omega)) = \sigma_1^{-1}\Bigg\{\tilde{u}_{k,A}(X,Z,A_{-1}) + \beta\mathbb{E}_{X'|X}\Bigg[(1+\alpha_{\theta,k}\Gamma_t)\log(\tilde{Y}_{k,A}(X',Z)) \\ -\sigma_1\log\left(\frac{p_k(P=1|X',Z,1,\omega+1)}{p_k(P=2|X',Z,0,\omega)}\right) - \beta\sigma_1\mathbb{E}_{X''|X'}\log\left(\frac{p_k(P=1|X'',Z,0,\omega+1)}{p_k(P=1|X'',Z,1,\omega+1)}\right)\Bigg]\Bigg\}. \quad (4)$$

which forms the 6th and final row of the linear system. To see how experimental variation can help identify β using this equation, note that this expression holds also for the case where time limits do not exist. If time limits were the only difference in the policy environment between Z and Z', we would get:

$$\log\left(\frac{\tilde{p}_{k,A}(X,Z',A_{-1},\omega}{\tilde{p}_{k,A}(X,Z,A_{-1})}\right) = \beta\mathbb{E}_{X'|X}\left[\log\left(\frac{p_k(P=1|X',Z,1)}{p_k(P=2|X',Z,0)}\right) - \log\left(\frac{p_k(P=1|X',Z',1,\omega+1)}{p_k(P=2|X',Z',0,\omega)}\right)\right]$$
$$+\beta^2\mathbb{E}_{X''|X}\left[\log\left(\frac{p_k(P=1|X'',Z,0)}{p_k(P=1|X'',Z,1)}\right) - \log\left(\frac{p_k(P=1|X'',Z',0,\omega+1)}{p_k(P=1|X'',Z',1,\omega+1)}\right)\right] \quad (5)$$

which is an expression that identifies β exclusively through the effect of time limits on participation choices.

B.1 A Minimum Distance Estimator

Having derived the system of equations that relates choice probabilities to parameters, this section briefly describes a minimum distance estimator that exclusively uses random assignment to identify Φ_3. In practice, there is one system like (2) for each type k and child age t. For ease of exposition, consider the estimator for just one pair (k,t) with the understanding that the approach easily generalizes by stacking such equations.

Abusing notation slightly, let Z now jointly indicate experimental site as well as treatment group. Define:

$$\boldsymbol{h}_l(Z) = \int \boldsymbol{h}_l(X,Z;\boldsymbol{p})dF(X|Z) \qquad l\in\{0,1\}.$$

This gives a linear system:

$$\boldsymbol{h}_0(Z) = \boldsymbol{\kappa}_0(\Phi_2) + \boldsymbol{\kappa}_1(\Phi)_2\boldsymbol{h}_1(Z)$$

with a natural estimator being:

$$\widehat{\Phi_2,\Phi_1} = \arg\min(\hat{\boldsymbol{h}}_0(Z) - \boldsymbol{\kappa}_0(\Phi_2) - \boldsymbol{\kappa}_1(\Phi)_2\hat{\boldsymbol{h}}_1(Z))^T\mathbf{W}(\hat{\boldsymbol{h}}_0(Z) - \boldsymbol{\kappa}_0(\Phi_2) - \boldsymbol{\kappa}_1(\Phi)_2\hat{\boldsymbol{h}}_1(Z))$$

Which is consistent and asymptotically normal for any positive definite $\mathbf{W}$ under standard regularity conditions.

C Taxes and Transfers

Total transfers are defined as the sum of taxes, welfare (if participating) and food stamps (if participating). Let n be the number of children, g the state in which the program takes place, y the year, and ω the current usage of time limits. E is monthly earnings.

C.1 Taxes

Taxes consist of a federal and a state computation. When earned income is sufficiently low, taxes will arrive in the form of a net payment (when income tax obligations are exceeded by the EITC). In theory, the relevant parameters to compute taxes include those that define the federal and state EITC programs, state and federal deductions and exemptions, and the marginal income tax rate with their corresponding brackets for state and federal income tax. In practice, I use the TAXSIM model of Feenberg and Coutts 1993, to approximate the tax function. Given the relevant year, state, and family size, TAXSIM computes net payments/obligations, $T^T_{mt}(e)$. This function is called for a fine grid of earnings between \$0 and \$100,000, then approximated using a polynomial spline function with year, state and family size-specific coefficients.

C.2 Control Groups

AFDC payments follow the formula:

$$\text{AFDC}(g,n,y,E) = \max\{B(g,n,y) - (1-0.33)\max\{E-120,0\},0\}$$

where $B(g,n,y)$ is the benefit standard. Food stamp payments are:

$$\text{SNAP}(g,n,y,E) = G(g,n,y) - 0.3(0.8E + \text{AFDC} - 134)$$

C.3 MFIP

: Both arms of MFIP change benefit formulae to:

$$\text{MFIP}(g,n,y,E) = \max\{\min\{1.2(B(g,n,y)+G(g,n,y)) - (1-0.38)E, B(g,n,y)+G(g,n,y)\},0\}$$

where B and G are the benefit standards and maximum food stamp payment in Minnesota for each number of kids (n) and year (y). Hence, food stamps and welfare are disregarded at the same rate rather than the double-deduction that occurs when AFDC payments are deducted from food stamps.

C.4 CT-Jobs First

The welfare payment for CTJF also folds in food stamps and features a "cliff" in benefits at the eligibility cutoff:

$$\text{CTJF}(g,n,y,E) = \mathbf{1}\{E < PG(n,y)\}(B(g,n,y)+G(g,n,y))$$

where $PG(n,y)$ is the poverty guidelines for a family of size n in year y.

C.5 FTP

The benefit formula for FTP is:

$$\text{FTP}(g,n,y,E) = \max\{B(g,n,y) - 0.5\max\{E-200,0\},0\}$$

Food stamps are the same as the control group.

D Details of Estimation

D.1 First Stage

We collect parameters into three blocks:

$$\Theta_1 = (\alpha_A, \alpha_S, \alpha_H, \alpha_F, \alpha_\theta, \alpha_R, \alpha_P, \beta_\Gamma, \sigma, \beta, y, \mu_W, \beta_W, \sigma_\eta, \mu_q, \beta_q, \lambda_0, \lambda_1, \delta, \mu_o, \sigma_o, \lambda_R)$$

$$\Theta_2 = (\sigma_W, \sigma_q)$$

$$\Theta_3 = (\beta_\tau, \pi_{\eta,0})$$

Recall that $\boldsymbol{y}_m = \{y_{m,t}\}_{t=1}^{T_m}$ is the panel of observed outcomes for mother m in period t. Let $W_m = \{W_{m,t}\}_{t=1}^{T_m}$ be observable components of the state space for mother m in each period t. As in the main text, $\chi_{m,t}$ is the vector of state variables that determine the distribution of outcomes. For simplicity, let dependence of $\chi_{m,t}$ on η and $W_{m,t}$ be implied whenever necessary. The likelihood is:

$$\mathcal{L} = \sum_m \log\left(\sum_{k=1}^{K}\sum_{t=1}^{T_m}\sum_{\eta_t=0}^{K_\eta} \mathbb{P}[y_{m,t}|\chi_{m,t},\Theta_1,\Theta_2]\mathbb{P}_k[\eta_{m,t+1}|\eta_{m,t},\Theta_1]\pi_{\eta,0}[\eta_1|k,X_{\tau,m}]\mathbb{P}[k|X_{\tau,m},\Theta_3]\right)$$

The expectation maximization routine proceeds as follows, fixing a current guess of parameters, Θ^l.

E-Step Using Θ^l, construct posterior probabilities for the latent state (k, η_t) in each time period:

$$q_{mt}(\eta,k) = \mathbb{P}[k,\eta|\boldsymbol{y}_m, X_m, \Theta^l]$$

$$q_{mt}(\eta_{t+1},\eta_t,k) = \mathbb{P}[\eta_{t+1},\eta_t,k|\boldsymbol{y}_m, X_m, \Theta^l]$$

This problem is made tractable by exploiting the Markov structure of the outcome probabilities using the forward-back algorithm.

M-Step Using the weights q_m, update Θ^l by moving up the likelihood:

$$\mathcal{L}(\Theta) = \sum_m\sum_k\Bigg[\log(\mathbb{P}[\eta_1,k|\Theta_3])q_{m,1}(\eta,k) + \sum_t\sum_\eta \log\left(\mathbb{P}[y_{m,t}|\chi_{m,t},\Theta_1,\Theta_2]\right)q_{m,t}(\eta,k) + \sum_{\eta'}\mathbb{P}_k[\eta'|\eta,\Theta_1]q_{mt}(\eta',\eta,k)\Bigg] \quad (6)$$

This maximization step is additively separable in each of the three blocks. One iteration of the step involves:

(a) Updating Θ_1^l with 5 iterations of the LBFGS algorithm using `Optim.jl` (Mogensen and Riseth 2018) with automatic differentiation using `ForwardDiff.jl` (Revels, Lubin, and Papamarkou 2016). This is completed in a number of separate block steps. Each evaluation of this piece of the likelihood requires solution of the model via backward induction.

(b) Update Θ_2^l with the new maximizer, which is given as a pair of weighted standard deviations for predicted wage and price residuals.

(c) Update the coefficients that determine type probabilities, β_τ, by maximizing the weighted initial type likelihood using LBFGS, and updating $\pi_{\eta,0}$ with the new maximizer which is given by a weighted frequency estimator.

This process is repeated until Θ^l and Θ^{l+1} satisfy a convergence criterion. Standard errors are estimated using the covariance of the score equation for each panel observation.

D.2 Second Stage - Production

Let X_m be a vector of location dummies and Z_m be a full set of location-treatment interaction dummies. The moment condition (4) implies a system of equations:

$$\boldsymbol{I}_m = X_m\Pi_1 + Z_m\Pi_2 + \eta_m$$

$$S_m = X_m A + \boldsymbol{I}_m B(\delta, g) + \xi_m$$

Where the terms η_m and ξ_m are expectation errors with unknown distribution. Their distribution is approximated as a joint normal, allowing for a likelihood to be written for the factor score S and inputs $\boldsymbol{I}$ conditional on X_m and Z_m. The priors are:

- $\delta_I \sim U[0, 2]$
- $\delta_\theta \sim U[0, 1]$
- $g_1, g_2 \sim U[-2, 2]$

To see that these priors are loose, the upper bound on δ_I implies that a log point increase in net household income would yield a 2 standard deviation increase in skills. A flat (i.e. improper) prior is specified for the parameters $(\Pi 1, \Pi_2, A)$. Estimates and figures reported in the main text are derived from 10,000 draws from a No-U-Turn Hamiltonian Monte-Carlo chain (Hoffman, Gelman, et al. 2014) implemented in `julia` using `Turing.jl` (Ge, Xu, and Ghahramani 2018).

Rearranging the second moment condition gives:

$$S_m = \boldsymbol{I}B(\delta, g) + \sum_{k,\eta} q_{m,1}(k, \eta)\mu(k, \eta, \boldsymbol{X}) + \eta_m$$

where $q_{m,1}$ is the posterior weight over latent states defined in Appendix D.1. A lower dimensional approximation for $\mu(k, \eta, \boldsymbol{X})$ appears to suffice, with higher dimensional interaction terms seeming to make little difference to estimates. Approximating the expectation error η_m as normal once again delivers a likelihood for S conditional on $\boldsymbol{I}, \boldsymbol{y}, \boldsymbol{X}$. The same methods and priors are used to sample from the posterior distribution.

References

Arcidiacono, Peter, and Robert A. Miller. 2011. "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity." *Econometrica* 79 (6): 1823–1867. https://doi.org/10.3982/ECTA7743. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA7743. https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA7743.

Feenberg, Daniel Richard, and Elizabeth Coutts. 1993. "An Introduction to the TAXSIM Model." *Journal of Policy Analysis and Management* 12 (1): 189–194.

Ge, Hong, Kai Xu, and Zoubin Ghahramani. 2018. "Turing: a language for flexible probabilistic inference." In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain,* 1682–1690. http://proceedings.mlr.press/v84/ge18b.html.

Hoffman, Matthew D, Andrew Gelman, et al. 2014. "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." *J. Mach. Learn. Res.* 15 (1): 1593–1623.

Mogensen, Patrick Kofod, and Asbjørn Nilsen Riseth. 2018. "Optim: A mathematical optimization package for Julia." *Journal of Open Source Software* 3 (24): 615. https://doi.org/10.21105/joss.00615.

Revels, J., M. Lubin, and T. Papamarkou. 2016. "Forward-Mode Automatic Differentiation in Julia." *arXiv:1607.07892 [cs.MS],* https://arxiv.org/abs/1607.07892.