Optimal Recall Period Length in Consumer Payment Surveys

Marcin Hitczenko

Consumer Payment Research Center
Federal Reserve Bank of Boston

September 4, 2014

FEDERAL RESERVE BANK OF BOSTON™

Motivation

CPRC estimates frequency of use of payment instruments.

- Average # of cash payments per week among U.S. adults.
- Also credit, debit, check, . . .

Reliable and comprehensive records for individuals . . .

- may not exist (cash).
- may pose significant respondent burden (privacy and credit card statements).
- are relatively expensive to obtain.

Instead, we rely on consumer surveys.

- Ask respondent for # of payments made.
- Involve inherent cognitive biases.

Survey Design

Virtually every aspect of the survey will affect responses [3, 8, 13, 15].

- Survey mode: **web**, telephone, in-person [5, 7, 14].
- Type of recall: [1, 4]
  - **Specific**: How many payments made in last week?
  - Typical: How many payments made in typical week?
- **Recall period**: day, week, month, year? [6, 9, 10, 12].

In this work, we focus on recall period.

Q1 Which recall period gives optimal results in estimating population means for cash, credit, debit, and check use?

Q2 Can we improve estimates by assigning different recall periods to different respondents?

Consider a hypothetical researcher ...

- Interested in population parameter ω. **Ex:** weekly average.
- Selects N individuals and asks for # payments made in last ℓ days.
- If no recall error, collects $A_\ell = \{A_{1\ell}, \ldots, A_{N\ell}\}$, where $A_{i\ell}$ is *actual* # of payments by respondent i.
- $\hat{\omega}(A_\ell)$ is estimate of ω. **Ex:**

Weekly data ($\ell = 7$)

$$\hat{\omega}(A_7) = N^{-1} \sum_{i=1}^{N} A_{i7}$$

Yearly data ($\ell = 365$)

$$\hat{\omega}(A_{365}) = N^{-1} \sum_{i=1}^{N} \frac{7A_{i,365}}{365}$$

- Wants sampling design so that estimator is unbiased:

$$\mathrm{E}\left[\hat{\omega}\left(A_\ell\right)\right] = \omega.$$

What if recall data is used in unbiased estimator instead of actual data?

- $R_\ell = \{R_{1\ell}, \ldots, R_{N\ell}\}$ represents *reported* data.
- Evaluate $\hat{\omega}(R_\ell)$ through mean-squared error:

$$\begin{aligned} \mathrm{MSE}(\hat{\omega}) &= \mathrm{E}\left[\hat{\omega} - \omega\right]^2 \\ &= \mathrm{Var}(\hat{\omega}) + \mathrm{Bias}^2(\hat{\omega}). \end{aligned}$$

- Generally, $\lim_{N\to\infty} \mathrm{MSE}(\hat{\omega}) = \mathrm{Bias}^2(\hat{\omega})$.

Focus is on population estimate, *not* individual recall.

- Perfect recall $\not\Longrightarrow$ perfect estimates. **Ex:** Perfect recall for year, but interested in Thanksgiving week.
- Imperfect recall $\not\Longrightarrow$ poor estimates. **Ex:** Regression to mean.

$$A_{i\ell} \sim F(\mathrm{mean} = \mu_{i\ell}) \text{ and } R_{i\ell} = pA_{i\ell} + (1-p)\mu_{i\ell} \implies \mathrm{E}[R_{i\ell}] = \mu_{i\ell}.$$

Q1: Which recall period gives optimal results in estimating population means for cash, credit, debit, and check use?

We rely on two datasets:

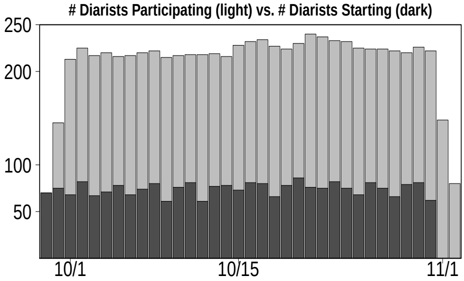
- *2012 Diary of Consumer Payment Choice (DCPC)*
  - 2,547 individuals from RAND's American Life Panel (ALP).
  - Track payment activity for three consecutive days in October 2012.
  - Provides direct insight into ω.
  - Patterns in data help define reasonable estimator forms.
- *2011-2012 Payment Recall Survey (PRS)*
  - 3,369 individuals from RAND's American Life Panel (ALP).
  - About 1,850 individuals participated in both surveys.
  - Fielded in five phases between May 2011 and September 2012.
  - Recall the # of payments made for day, week, month, and year for all four major payment instruments.
  - Provides insight into quality of recall for different recall periods.

DCPC Data

Day	$ Value	PI Used	Other Information
10/1	13.39	cash	10:15AM, grocery store, ...
10/1	45.00	credit	4:00PM, restaurant,...
⋮	⋮	⋮	⋮
10/3	200.00	credit	12:30PM, automobile, ...

Table: Data for one individual.

- 3-day periods randomly distributed in month.
- Provides the daily number of payments made with each payment instrument.



What does DCPC data look like?

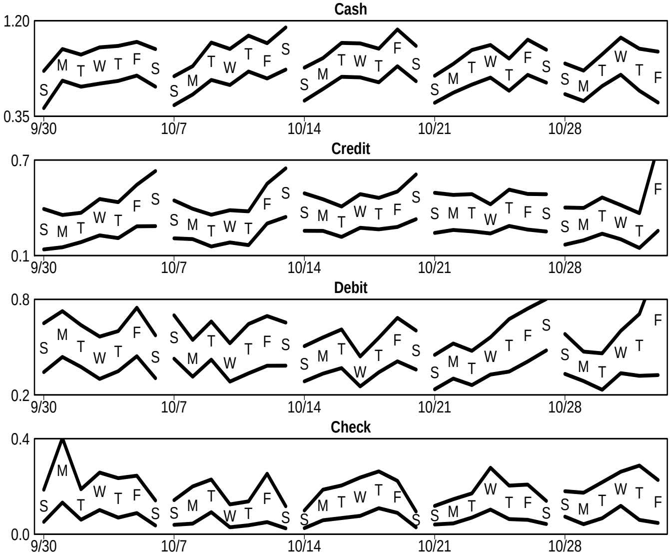


Figure: Means and 95 percent confidence intervals for mean daily use.

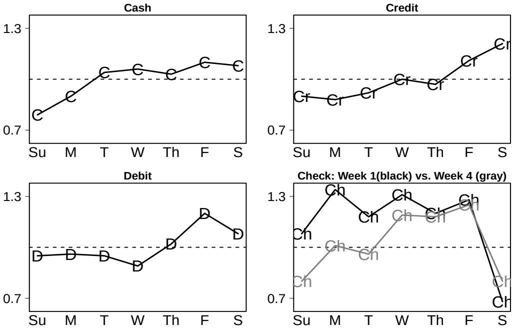
We fit a mixed-effects log-linear model for each instrument:

$$\# \text{ payments on day } t \sim \mathrm{Poisson}(\mu_{it}).$$

- $\log(\mu_{it}) = \mu_i + f(t)$
- μ_i: random effect corresponding to individual.
- $f(t)$: fixed effects corresponding to day-of-week or day-of-month.

Comparison of models finds

- Strong day-of-week effects for all four instruments.
- Evidence of monthly cycle for checks.



Back to our hypothetical researcher ...

- $\omega =$ mean # payments per week in October 2012.
- SRS among target population (defined by ALP).
- Seemingly reasonable linear estimators:

$$\hat{\omega}_\ell = \sum_{i=1}^{N} w_{i\ell} R_{i\ell}$$

  - $\ell = 1$: $w_{i1} = (N_d)^{-1}$, $N_d =$ # reporting for day-of-week d.
  - $\ell > 7$: $w_{i\ell} = \frac{7}{N\ell}$.
- Possible limitations:
  - Monthly ($\ell = 30$) and yearly ($\ell = 365$) recall is not quite right; intervals of 30 and 365 days do not have equal representation of each day of week.
  - Yearly recall ($\ell = 365$) extends to periods outside of October 2012.

PRS Data

- Respondents participate in 1-3 phases (3-9 months between surveys).
- In each phase of survey:
  - Sequence of payment instruments is randomized.
  - Order of day, week, and month is randomized; year is always last.
  - Day is randomly assigned within past week.

Data for one individual (in each phase of survey)

	Day in Last Week	**Past Week**	**Past Month**	**Past Year**
Cash	2	8	30	350
Credit	1	7	25	200
Debit	0	2	10	90
Check	0	0	0	0

Table: Reported # of payments.

Example of timing of PRS and DCPC:

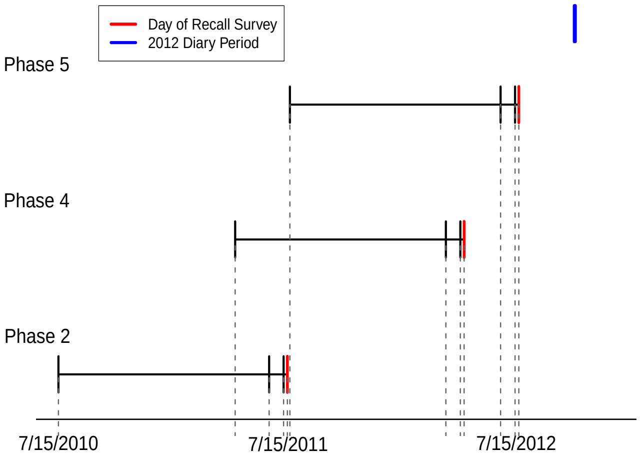


Figure: Timing for individual who took DCPC on October 15th, 2012.

We want to estimate bias of estimator based on recall period length ℓ:

$$\mathrm{Bias}(\hat{\omega}_\ell) \quad = \quad \mathrm{E}[\hat{\omega}_\ell] - \omega.$$

- Linear estimator depends on $\mathrm{E}[R_{i\ell}]$ and ω.
- Use DCPC data to estimate ω.
- Use PRS data to estimate $\mathrm{E}[R_{i\ell}]$.
  - Use only PRS data from after August 15^{th} 2012.
  - Adjust for any lag effect with daily recall (not found to be significant).
  - Randomization in PRS helps with various survey-specific effects.
    Ex: Dependence of response errors (weekly value should limit possible daily values).
- Bootstrap respondents to determine distribution of bias estimate:
  - Sample within respondents who took both surveys.
  - Sample within respondents who only took DCPC.
  - Sample within respondents who only took PRS.

For each bootstrapped sample, estimate

- Bias for each ℓ.
- Which recall period minimizes absolute bias.

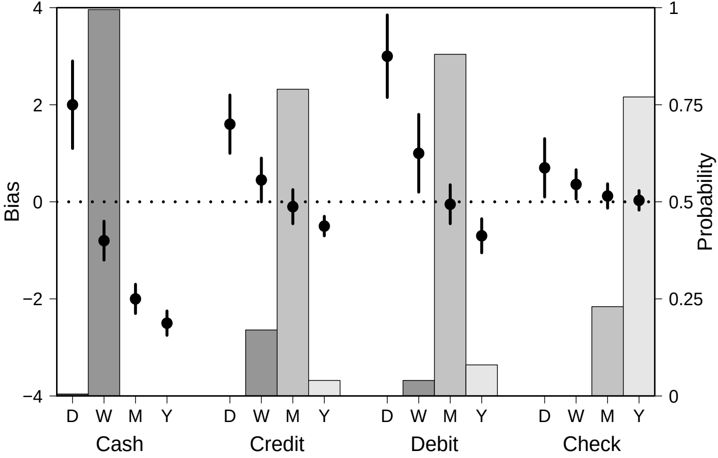


Figure: Bootstrapped bias (lines) and probability of minimizing absolute bias (bars).

Conclusions

- Optimal recall periods differ across payment instruments:
  - Week for cash
  - Month for credit,debit.
  - Year for check.
- Hurd and Rohwedder [9] suggest that optimal recall periods relate to the frequency of behavior.
- Survey of Consumer Payment Choice (SCPC):
  - Taken by those who took DCPC; also in October 2012.
  - Respondents choose recall period (week, month, year) to report typical # of payments.
  - Correspondence between DCPC data and reported SCPC results matches these results.
    Ex: Respondents who report cash on weekly basis show most consistency between SCPC (recall) data and DCPC (diary) data.

Q2: Can we improve estimates by assigning different recall periods to different respondents?

- Recall for individual i is based on recall period ℓ_i.
- ω_i = weekly mean for individual i.
- If $\mathrm{E}[\omega_i] = \omega$ with respect to sampling scheme,

$$\mathrm{E}[\hat{\omega} - \omega] \quad \leq \quad \sum_{i=1}^{N} \mathrm{E}\left| w_{i\ell_i} R_{i\ell_i} - \omega_i \right| .$$

- Minimizing discrepancy between recall-based estimate of ω_i and true ω_i likely improves population estimate.
- Can optimal recall periods be predicted for individuals based on demographic information known ahead of survey?

For any individual i

- $R_{is\ell} = \#$ payments in last ℓ days reported on day s (i.e. phase s).
- $B_{is} =$ recall period that produces closest approximation to ω_i:

$$B_{is} = \text{argmin}_{\ell} \left| w_{is\ell} R_{is\ell} - \omega_i \right| .$$

- If we know ω_i, we can determine B_{is} from PRS data. **Ex:** If $\omega_i = 5$:

Recall Period	Response	Scaled Estimate of ω	Difference
Week	7	$7 \times \frac{7}{7} = 7$	2
Month	20	$20 \times \frac{7}{30} = 4.67$	-0.33
Year	200	$200 \times \frac{7}{365} = 3.83$	-1.17

- Sampling ω_i allows us to sample B_{is}; want to sample from

$$\mathrm{P}(\omega_i \mid \text{DCPC,PRS}) \propto \mathrm{P}(\omega_i \mid \text{DCPC}) \mathrm{P}(\text{PRS} \mid \omega_i)$$

A simple model:

- Distribution of $\omega_i \mid$ DCPC provided from random-effect models; related to μ_i.
- Distribution of $R_{is\ell} \mid \omega_i$ takes form:

$$R_{is\ell} \mid \omega_i, \lambda_{is\ell} \sim \mathrm{Poisson}\left(\lambda_{is\ell} \times \frac{\ell}{7}\omega_i\right)$$

- $\lambda_{is\ell}$ represents degree of reporting bias:
  - $\lambda_{is\ell} = 1 \implies$ unbiased recall
  - $\lambda_{is\ell} > 1 \implies$ overestimation
  - $\lambda_{is\ell} < 1 \implies$ underestimation
- Special case of model based on idea that recall is done via enumeration or rate-based estimation [2, 3].
- $\omega_i = 0 \implies \mathrm{P}(R_{is\ell} = 0) = 1$.

We run a MCMC procedure:

- Restrict data to individuals who took DCPC *and* participated in PRS after July 2012.
- Use cash only:
  - only instrument with very high adoption rates.
  - issue of non-adoption ($\omega_i = 0$) presents modeling computations.
- Compare weekly, monthly, and yearly recall:
  - currently adding daily recall.
- Assume $\lambda_{is\ell} \sim \mathrm{Gamma}(k_\ell, \tau_\ell)$, independent across i, s and ℓ:
  - currently loosening independence assumptions (especially across s).
- Use non-informative hyper-priors: $\mathrm{P}(k_\ell, \tau_\ell) \propto 1$.
- Generate draws of $\omega_i \mid$ DCPC,PRS.

Example 1: Prior vs. posterior estimates of ω_i.

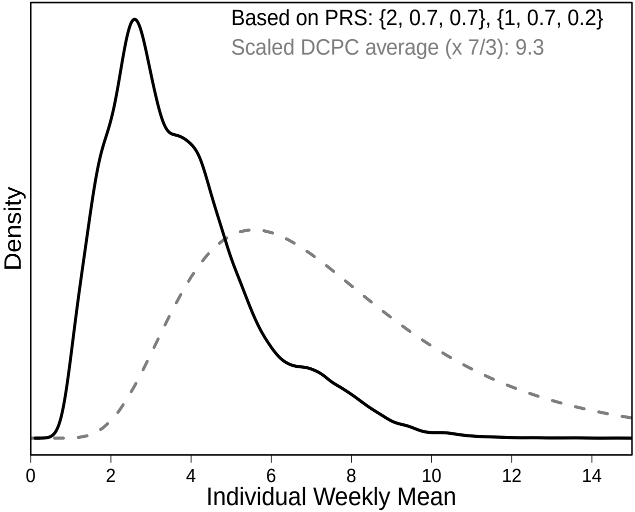


Figure: Prior (dashed) and posterior(solid) distributions of ω_i. PRS estimates are ordered according to $\{W,M,Y\}$ recall.

Example 2: Prior vs. posterior estimates of ω_i.

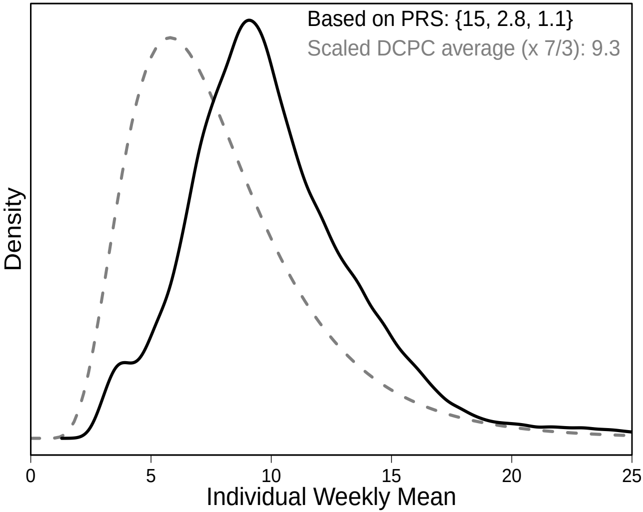


Figure: Prior (dashed) and posterior(solid) distributions of ω_i. PRS estimates are ordered according to {W,M,Y} recall.

Example 3: Prior vs. posterior estimates of ω_i.

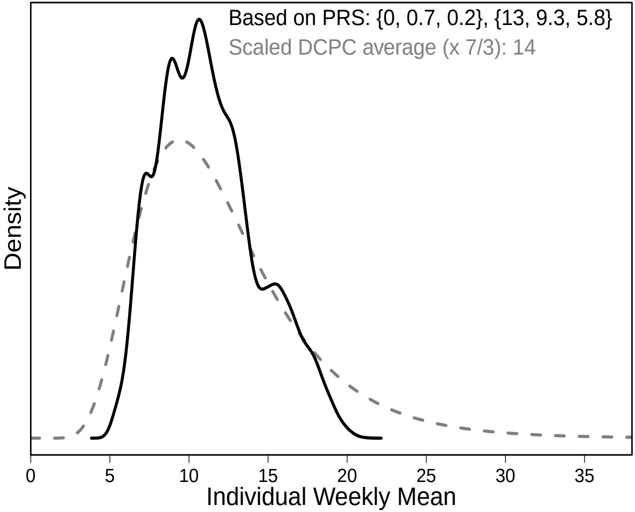


Figure: Prior (dashed) and posterior(solid) distributions of ω_i. PRS estimates are ordered according to {W,M,Y} recall.

In each posterior draw from MCMC algorithm:

- Given ω_i, determine B_{is}:

Individual (i)	1	1	1	2	...
Phase (s)	2	4	5	3	...
ω_i	5.4	5.4	5.4	1.2	...
B_{is}	week	month	week	month	...

- For the generated set $\{B_{is}\}$ fit models:
  - $\mathrm{P}(B_{is} = \ell) \propto \exp(\alpha_\ell)$
  - $\mathrm{P}(B_{is} = \ell) \propto \exp(\mathrm{demo}_i^T \beta_\ell)$.
- Second model suggests that optimal recall period for individual relates to demographic information (demo_i).
- Demographic information includes age, gender, education, and income.

- For each draw calculate deviance between models.

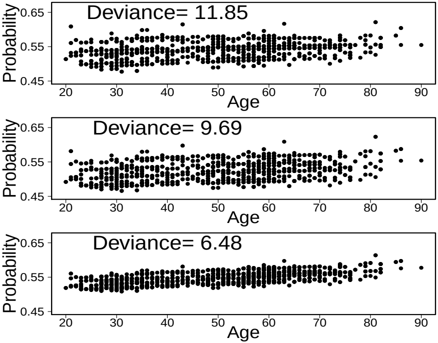


Figure: Fitted probabilities, $\hat{P}(B_{is} = 7)$ for three draws from MCMC.

- Averaging over draws, find little evidence that demographics predict the optimal recall period (p-value$=$ 0.16).
- For all demographic combinations, the weekly recall period is always most likely to be best.

Conclusions

- Important to think carefully about what parameters we are trying to estimate, and whether sampling design is suited to optimize results.
- Evidence that optimal recall periods depend on what is being measured; linked to frequency of behavior?
- Not (yet?) enough evidence of heterogeneity in optimal recall lengths to justify assigning different recall periods to different respondents.

Limiting Factors/Future Work

- Diary data is not necessarily the truth [11].
  - Get more accurate records (if possible).
- Modeling assumptions may not be correct.
  - Expand analysis and flexibility of models.
- Results may not hold for broader populations; the ALP is not representative of US.

[1] Babor, Thomas F., Joseph Brown, and Frances K. Del Boca. 1990. "Validity of Self-Reports in Applied Research on Addictive Behaviors: Fact or Fiction?" *Behavioral Assessment* 12:5–31.

[2] Blair, Edward and Scott Burton. 1986. "Processes Used in the Formulation of Behavioral Frequency Reports in Surveys." *American Statistical Association Proceedings of the Section on Survey Methods* :481–487.

[3] Bradburn, Norman M., Lance J. Rips, and Steven K. Shevell. 1987. "Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys." *Science* 236:157–161.

[4] Chang, Linchiat and Jon A. Krosnick. 2003. "Measuring the Frequency of Regular Behaviors: Comparing the Typical Week to the Past Week." *Sociological Methodology* 33:55–80.

[5] Couper, Mick P., Michael W. Traugott, and Mark J. Lamias. 2001. "Web Survey Design and Administration." *Public Opinion Quarterly* 65(2):230–253.

[6] Deaton, Angus and Valerie Kozel. 2005. "Data and Dogma: The Great Indian Poverty Debate." *World Bank Research Observer* 20(2):177–199.

[7] Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2008. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: Wiley and Sons.

[8] Eisenhower, Donna, Nancy A. Mathiowetz, and David Morganstein. 1991. "Recall Error: Sources and Bias Reduction Techniques." In *Measurement Errors in Surveys*, edited by Paul P. Biermer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman. Wiley.

[9] Hurd, Michael and Susann Rohwedder. 2009. "Methodological Innovations in Collecting Spending Data: The HRS Consumption and Activities Mail Survey." *Fiscal Studies* 30:435–459.

[10] Mahalanobis, P. C. and S.B. Sen. 1954. "On Some Aspects of the Indian National Sample Survey." *Bulletin of the International Statistical Institute* 34.

[11] Silberstein, Adriana R. and Stuart Scott. 1991. "Expenditure Diary Surveys and Their Associated Errors." In *Measurement Errors in Surveys*, edited by Paul P. Biermer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman. Wiley and Sons.

[12] Sudman, Seymour and Norman M. Bradburn. 1973. "Effects of Time and Memory Factors on Response in Surveys." *Journal of the American Statistical Association* 68(344):805–815.

[13] Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco, CA: Jossey-Bass.

[14] Tourangeau, Roger, Mick P. Couper, and Federick Conrad. 2004. "Spacing, Position, and Order: Interpretive Hueristics for Visual Features of Survey Questions." *Public Opinion Quarterly* 68(3):368–398.

[15] Tversky, Amos and Daniel Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185:1124–1131.

Models for $f(t)$

$$\text{Let } \texttt{dow}(t) = \begin{cases} 1 & t \text{ is a Sunday} \\ \vdots & \vdots \\ 7 & t \text{ is a Saturday} \end{cases}$$

index the day of the week and

$$\texttt{pom}(t) \quad = \quad \frac{\sum_{t'} 1\left[t' \leq \ t, \text{ and } t', t \text{ in same month}\right]}{\sum_{t'} 1\left[t', t \text{ in same month}\right]}$$

define location within a month. **Ex:** $\texttt{pom}\left(\text{October } 15^{th}\right) = \frac{15}{31}$.

We consider three models for $f(t)$:

A $f(t) = \sum_{j=1}^{7} \beta_j 1\left[\texttt{dow}(t) = j\right] + \alpha_1 \texttt{pom}(t) + \alpha_2 \texttt{pom}^2(t)$

B $f(t) = \sum_{j=1}^{7} \beta_j 1\left[\texttt{dow}(t) = j\right]$

C $f(t) = \nu$.

$P(\omega_i \mid \text{DCPC})$

The first term in posterior, $\mathrm{P}(\omega_i \mid \text{DCPC})$:

- Represents posterior estimate of $\boldsymbol{\omega}$ given DCPC data.
- Defined via estimates of $f(t)$ and predictions of μ_i in model fits:

$$\begin{aligned} \omega_i \mid \text{DCPC} &= \sum_{j=1}^{7} \exp(\mu_i + \beta_j) \\ &= \exp(\mu_i) \sum_{j=1}^{7} \exp(\beta_j) \end{aligned}$$

  with $\mu_i \mid \text{DCPC} \sim \mathrm{Normal}(\hat{m}_i, \hat{v}_i)$.
- Can be approximated with $\omega_i \mid \text{DCPC} \sim \mathrm{Gamma}(k_i, \tau_i)$ with parameters (k_i, τ_i) determined by matching first two moments of distribution implied by $(\hat{m}_i, \hat{v}_i)$.

Model for Recall Data

The model

$$R_{is\ell} \mid \omega_i, \lambda_{is\ell} \sim \mathrm{Poisson}\left(\lambda_{is\ell} \times \frac{\ell}{7}\omega_i\right),$$

is a special case of more general class of models:

$$R_{is\ell} \mid \omega_i, \lambda_{is\ell} = \begin{cases} \lambda_{is\ell} A_{is\ell} & \text{w.p. } p(\ell) \\ \mathrm{Poisson}(\frac{\ell}{7} \times \gamma_i \omega_i) & \text{w.p. } 1 - p(\ell) \end{cases}$$

- $p(\ell)$ defines probability of using enumeration (presumably decreases as ℓ increases).
- $\lambda_{is\ell}$ defines the bias in the enumeration estimation.
- γ_i defines the bias in the rate-based estimation.