Online Appendix

Appendix A: Baseline Regression Estimator

Table EC.1 describes the baseline estimator that was generated using reference literature values from Gidrewicz and Fenton (2014).

	Term		Pre-term	
Average milk age	Fat	Protein	Fat	Protein
[0,3]	1.8	2.0	2.2	2.7
(3,7]	2.6	1.6	3.0	1.7
(7,14]	3.0	1.3	3.5	1.5
(14,28]	3.4	1.1	3.5	1.4
(28,42]	3.6	1.0	3.2	1.1
(42,63]	3.4	0.9	3.3	1.1
> 63	3.4	1.0	3.7	1.0

Table EC.1 Baseline estimator for fat and protein levels of a donation. The baseline value depends on the Term or Pre-term status of the baby and "average milk age", which we define to be the difference between the midpoint between (i) the latest and earliest donation bag and (ii) the baby's date of birth.

The baseline protein and fat estimator achieved MAE values of 0.34 ± 0.010, and 0.82 ± 0.024, respectively. In contrast, our prediction models achieved 0.10 ± 0.008 and 0.79 ± 0.025 MAE, corresponding to reductions of 70.5% and 3.6%.

Appendix B: Feature Importance of Machine Learning Models

Figure EC.1 shows the feature importance for each model using Gini Importance. Note that the age of the baby when the deposit was first pumped, as denoted by the label "earliest day" in Figure EC.1, was by far the most important feature for predicting protein levels. All of the "day" variables are coded in terms of the age of the baby.

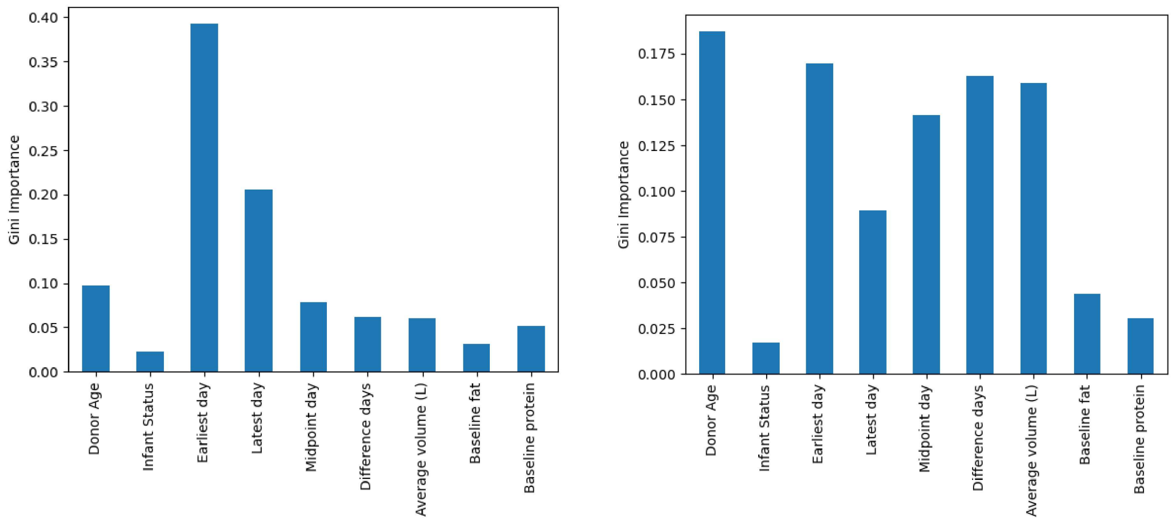
Appendix C: Parameters of the Optimization Problem

A complete list of parameters used in our optimization problem can be found in Table EC.2.

Appendix D: Sensitivity of Model Hyperparameters

The hyperparameters β, $\boldsymbol{\delta}$, and λ of the **PO-FIFO** model where tuned via greedy coordinate descent. We first fix $\lambda = 0$ and $\boldsymbol{\delta} = (0,0)$ (i.e., specializing the objective to ℓ_{T}) and tune $\beta \in [0,1]$. Fixing $\boldsymbol{\delta} = (0,0)$ and the best value found for β, we then tune $\lambda \in [0,1]$. Finally, fixing the β and λ values, we tune $\boldsymbol{\delta} \in [0,0.2] \times [0,0.02]$ for fat and protein, respectively.

Figure EC.2 plots the primary metrics for different β values, fixing $\boldsymbol{\delta} = (0,0)$ and $\lambda = 0$. We observe that pass rates are similar for all values of β. Setting $\beta = 0$, which implies that the objective only penalizes failing to meet macronutrient targets, results in the highest value of 'Protein' MAD. On the other hand, $\beta = 1$ leads to the low 'Fat' and 'Protein' MAD. Figure EC.3 plots the primary metrics for different λ values, fixing



(a) Protein (b) Fat

Figure EC.1 Gini Importance was used to measure feature importance for the final random forest models for (a) protein and (b) fat.

	Parameter	Description	Value
Operational	m^*_{fat}	Target fat content in each pool	3.6 g/dL
	m^*_{pro}	Target protein content in each pool	0.9 g/dL
	V_{tot}	Target volume of milk in a recipe	36 L
	$\overline{V}^r$ ($\underline{V}^r$)	Maximum (minimum) volume of milk in a recipe	36.5 L (36 L)
	$\overline{V}^t$ ($\underline{V}^t$)	Maximum (minimum) volume of milk that can be taken from a deposit for a recipe if the deposit is used	18 L (5 L)
	$\underline{V}^l$	Minimum volume of milk to leave in a deposit if the deposit is used	5 L
	$\underline{W}$	Minimum number of donors that must be used in a recipe	3 donors
	$\overline{X}$	Maximum number of deposits that can be used in a recipe	5 deposits
	$\overline{B}$	Maximum permissible bacteria count in a pool	50 col./plate/L
	$\overline{H}$	Maximum volume of previously refused milk in a pool	$(1/6)V_{\text{tot}}$ L
Deposit ($\mathbb{R}^I$)	$\hat{\mathbf{m}}_{\text{fat}}$	Estimated fat content	
	$\hat{\mathbf{m}}_{\text{pro}}$	Estimated protein content	
	$\hat{\mathbf{H}}$	Indicator of previously refused milk	
	$\hat{\mathbf{V}}$	Total volume	
	$\hat{\mathbf{B}}$	Estimated bacteria content	
	$\mathbf{D}$	Number of days until expiry	

Table EC.2 Parameters of the optimization model. The first set of parameters relate to milk bank operations. The second set of parameters correspond to deposit features and are given as vectors in $\mathbb{R}^I$.

$\beta = 1$ and $\boldsymbol{\delta} = (0,0)$. Setting $\lambda > 0$ usually improves MAD and pass rates, especially for 'Fat', suggesting regularization can partially mitigate prediction error. Specifically, $\lambda = 0.1$ yields the best performance in the protein metrics and is competitive in the fat metrics. Finally, Figure EC.4 plots different $\boldsymbol{\delta}$ values, fixing $\beta = 1$ and $\lambda = 0.1$. Pass rates and MAD values are similar for different $\boldsymbol{\delta}$ values, with (0, 0.01) offering a slight advantage in 'Protein' MAD. Recall from Table 1 that the average fat and protein content of deposits in this data set are 3.7 g/dL and 0.87 g/dL, respectively, meaning deposits already contain more fat than the required 3.6 g/dL on average. Thus, setting $\delta_{\text{fat}} = 0$ is reasonable, and we choose $\boldsymbol{\delta} = (0, 0.01)$.

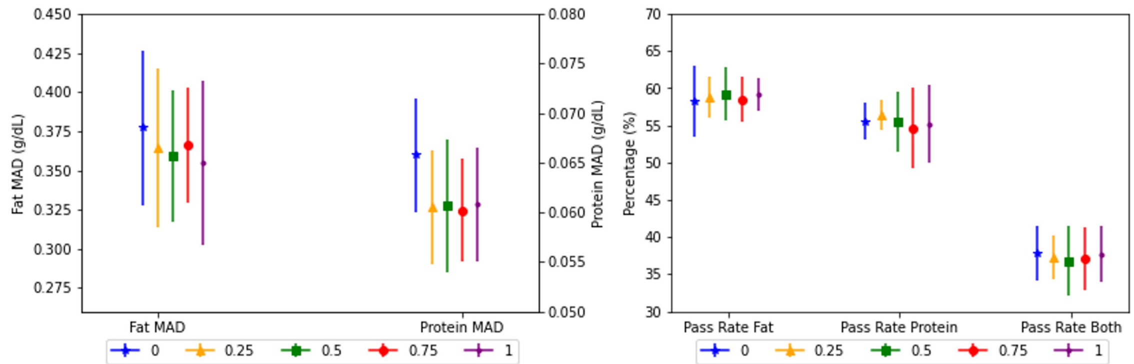


Figure EC.2 **Performance of PO-FIFO on MAD and pass rate with varying β while holding $\boldsymbol{\delta} = (0,0)$ and $\lambda = 0$.**

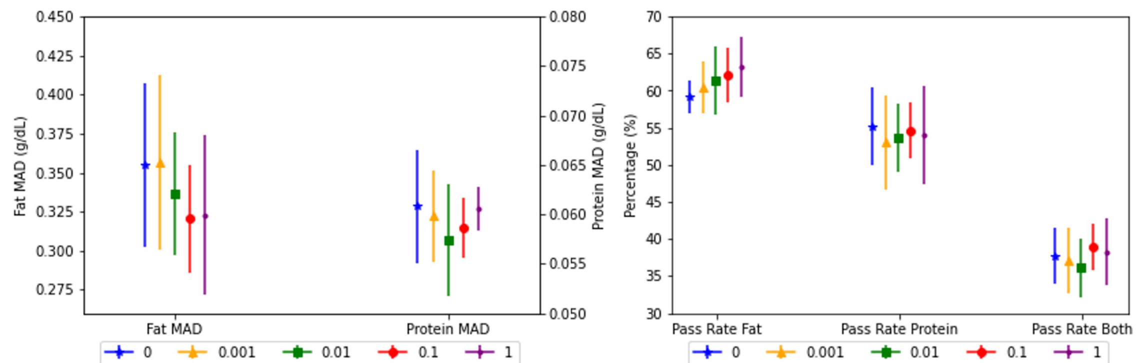


Figure EC.3 **Performance of PO-FIFO on MAD and pass rate with varying λ while holding $\beta = 1$ and $\boldsymbol{\delta} = (0,0)$.**

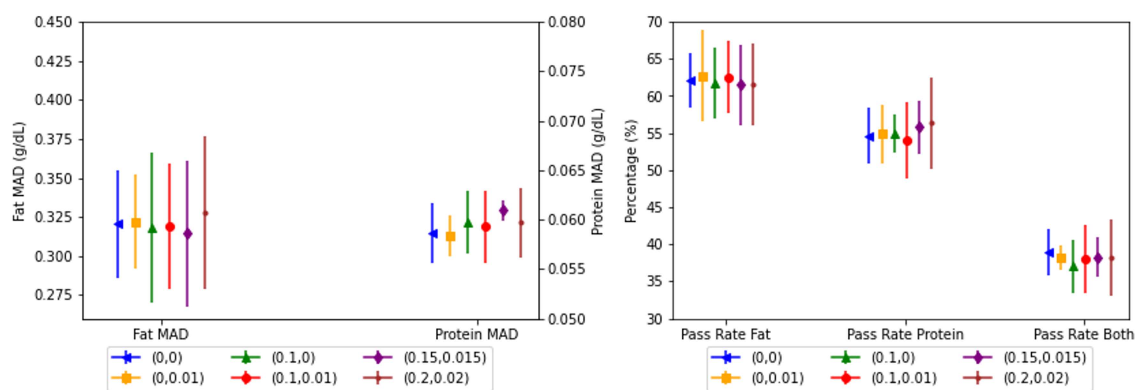


Figure EC.4 **Performance of PO-FIFO on MAD and pass rate with varying $\boldsymbol{\delta}$ while holding $\beta = 1$ and $\lambda = 0.1$.**

Given the above results, our final settings for **PO-FIFO** were $\beta = 1, \lambda = 0.1$, and $\boldsymbol{\delta} = (0, 0.01)$. We tested sensitivity around this regime but did not find any improvement. When facing ties between metrics, we put more emphasis on MAD, given its generalizability to other milk banks (whereas macronutrient targets tend to be milk bank-specific).

Appendix E: Proofs

To prove Theorem 1, we require the original result of Bartlett and Mendelson (2002), which we state below.

THEOREM EC.1 (**Bartlett and Mendelson (2002)**). *Let $\{(\mathbf{w}_j, m_j)\}_{j=1}^n$ be a data set of i.i.d. pairs from $\mathcal{W} \times \mathcal{M}$, let $f : \mathcal{W} \to \mathcal{M}$ be a model from a hypothesis class $f \in \mathcal{F}$. Consider a 1-Lipschitz loss function*

$\ell:\mathcal{M}\times\mathcal{M}\to[0,c]$ for some constant c. Then for any $\gamma>0$, with probability at least $1-\gamma$

$$\mathbb{E}\left[\ell(m,f(\mathbf{w}))\right]\le\frac{1}{n}\sum_{j=1}^{n}\ell(m_j,f(\mathbf{w}_j))+2\mathfrak{R}_n(\mathcal{F})+c\sqrt{\frac{\log(1/\gamma)}{2n}},\qquad\forall f\in\mathcal{F}.$$

We now prove our main results.

Proof of Theorem 1. We first decompose the probability into the likelihoods of the estimates of each deposit:

$$\Pr\left\{\tilde{\mathbf{M}}-\hat{\mathbf{M}}\in\mathcal{U}(\lambda)\right\}=\Pr\left\{\left\|\left(\tilde{\mathbf{M}}-\hat{\mathbf{M}}\right)^{\top}\right\|_{q,1}\le\lambda\right\}\tag{EC.1}$$

$$=1-\Pr\left\{\left\|\left(\tilde{\mathbf{M}}-\hat{\mathbf{M}}\right)^{\top}\right\|_{q,1}>\lambda\right\}\tag{EC.2}$$

$$=1-\Pr\left\{\sum_{i=1}^{I}\|\tilde{\mathbf{m}}_i-\hat{\mathbf{m}}_i\|_q>\lambda\right\}\tag{EC.3}$$

$$\ge 1-\Pr\left\{\sum_{i=1}^{I}\|\tilde{\mathbf{m}}_i-\hat{\mathbf{m}}_i\|_1>\lambda\right\}\tag{EC.4}$$

$$=1-\Pr\left\{\sum_{i=1}^{I}\|\tilde{\mathbf{m}}_i-\hat{\mathbf{m}}_i\|_1-\mathbb{E}\left[\sum_{i=1}^{I}\|\tilde{\mathbf{m}}_i-\hat{\mathbf{m}}_i\|_1\right]>\lambda-\mathbb{E}\left[\sum_{i=1}^{I}\|\tilde{\mathbf{m}}_i-\hat{\mathbf{m}}_i\|_1\right]\right\}\tag{EC.5}$$

$$\ge 1-\exp\left(-\frac{2}{Ic^2}\left(\lambda-\mathbb{E}\left[\sum_{i=1}^{I}\|\tilde{\mathbf{m}}_i-\hat{\mathbf{m}}_i\|_1\right]\right)^2\right)\tag{EC.6}$$

$$\ge 1-\exp\left(-\frac{2}{Ic^2}\left(\lambda-I\mathbb{E}\left[|\tilde{m}_{\text{fat}}-f_{\text{fat}}(\mathbf{w})|\right]-I\mathbb{E}\left[|\tilde{m}_{\text{pro}}-f_{\text{pro}}(\mathbf{w})|\right]\right)^2\right).\tag{EC.7}$$

In the above, (EC.2) first replaces the event with its complement, (EC.3) follows from the definition of the matrix norm, and (EC.4) follows from the dominance of p-norms (i.e., $\|\cdot\|_1\ge\|\cdot\|_q$ for all $q\ge1$). Next, (EC.5) follows from subtracting the expectation on both sides of the inequality and (EC.6) follows from Hoeffding's Inequality. Finally, (EC.7) expands the expectation via our i.i.d. assumption and introduces the regression models.

We now apply Theorem EC.1 to show that with probability at least $(1-\gamma)^2$,

$$I\mathbb{E}\left[|\tilde{m}_{i,\text{fat}}-\hat{m}_{i,\text{fat}}|\right]+I\mathbb{E}\left[|\tilde{m}_{i,\text{pro}}-\hat{m}_{i,\text{pro}}|\right]\le\frac{I}{n}\sum_{j=1}^{n}\|\tilde{\mathbf{m}}_j-\mathbf{f}(\mathbf{w}_j)\|_1+4I\mathfrak{R}_n(\mathcal{F})+2Ic\sqrt{\frac{\log 1/\gamma}{2n}}\le\lambda$$

where the second inequality follows from requirement (5). Then, with probability at least $(1-\gamma)^2$, we have

$$\text{RHS (EC.7)}\ge 1-\exp\left(-\frac{2}{Ic^2}\left(\lambda-\frac{I}{n}\sum_{j=1}^{n}\|\tilde{\mathbf{m}}_j-\mathbf{f}(\mathbf{w}_j)\|_1-4I\mathfrak{R}_n(\mathcal{F})-2Ic\sqrt{\frac{\log 1/\gamma}{2n}}\right)^2\right).\tag{EC.8}$$

Now let $\zeta=1-(1-\gamma)^2$. Rewriting (EC.8) in terms of ζ completes the proof.

Proof of Theorem 2. The proof shares the same initial steps as the proof to Theorem 1. First from (EC.4), we show:

$$\Pr\left\{\tilde{\mathbf{M}}-\hat{\mathbf{M}}\in\mathcal{U}(\lambda)\right\}\ge 1-\Pr\left\{\sum_{i=1}^{I}\|\tilde{\mathbf{m}}_i-\hat{\mathbf{m}}_i\|_1>\lambda\right\}$$

$$=1-\Pr\left\{\sum_{i=1}^{I}\left(|\tilde{m}_{i,\text{fat}}-\hat{m}_{i,\text{fat}}|+|\tilde{m}_{i,\text{pro}}-\hat{m}_{i,\text{pro}}|\right)>\lambda\right\}\tag{EC.9}$$

$$\geq 1 - \Pr\left\{\left(\sum_{i=1}^{I} |\tilde{m}_{i,\text{fat}} - \hat{m}_{i,\text{fat}}| > \frac{\lambda}{2}\right) \bigcup \left(\sum_{i=1}^{I} |\tilde{m}_{i,\text{pro}} - \hat{m}_{i,\text{pro}}| > \frac{\lambda}{2}\right)\right\} \tag{EC.10}$$

$$\geq 1 - \Pr\left\{\sum_{i=1}^{I} |\tilde{m}_{i,\text{fat}} - \hat{m}_{i,\text{fat}}| > \frac{\lambda}{2}\right\} - \Pr\left\{\sum_{i=1}^{I} |\tilde{m}_{i,\text{pro}} - \hat{m}_{i,\text{pro}}| > \frac{\lambda}{2}\right\} \tag{EC.11}$$

$$\geq 1 - \frac{2}{\lambda}\sum_{i=1}^{I}\left(\mathbb{E}\left[|\tilde{m}_{i,\text{fat}} - \hat{m}_{i,\text{fat}}|\right] + \mathbb{E}\left[|\tilde{m}_{i,\text{pro}} - \hat{m}_{i,\text{pro}}|\right]\right) \tag{EC.12}$$

In the above, (EC.9) expands the norm and (EC.10) follows from the fact that if the sum of two numbers is greater than a value λ, at least one of them is greater than $\lambda/2$. Then, (EC.11) follows from the union bound and (EC.12) follows from Markov's inequality.

We first consider the mean absolute deviation term for fat macronutrients and later use the exact steps for the protein term. First, observe that

$$\tilde{m}_{i,\text{fat}} - \hat{m}_{i,\text{fat}} = \frac{1}{K}\sum_{k=1}^{K}\hat{m}_{i,k,\text{fat}} - \frac{1}{\hat{K}}\sum_{k=1}^{\hat{K}}\hat{m}_{i,k,\text{fat}} = \frac{1}{K}\sum_{k=\hat{K}+1}^{K}\hat{m}_{i,k,\text{fat}} + \left(\frac{1}{K} - \frac{1}{\hat{K}}\right)\sum_{k=1}^{\hat{K}}\hat{m}_{i,k,\text{fat}}$$

is a sum of Gaussian random variables and thus, also a Gaussian random variable with mean and variance

$$\mathbb{E}\left[\frac{1}{K}\sum_{k=\hat{K}+1}^{K}\hat{m}_{i,k,\text{fat}} + \left(\frac{1}{K} - \frac{1}{\hat{K}}\right)\sum_{k=1}^{\hat{K}}\hat{m}_{i,k,\text{fat}}\right] = (K - \hat{K})\frac{1}{K}\mu_1 + K\left(\frac{1}{K} - \frac{1}{\hat{K}}\right)\mu_1 = 0$$

$$\operatorname{var}\left(\frac{1}{K}\sum_{k=\hat{K}+1}^{K}\hat{m}_{i,k,\text{fat}} + \left(\frac{1}{K} - \frac{1}{\hat{K}}\right)\sum_{k=1}^{\hat{K}}\hat{m}_{i,k,\text{fat}}\right) = \Sigma_{1,1}\left((K - \hat{K})\frac{1}{K^2} + K\left(\frac{1}{K} - \frac{1}{\hat{K}}\right)^2\right).$$

Moreover, the absolute value of this random variable is a Folded Gaussian random variable with mean

$$\mathbb{E}\left[|\tilde{m}_{i,\text{fat}} - \hat{m}_{i,\text{fat}}|\right] = \sqrt{\frac{2\Sigma_{1,1}}{\pi}\left((K - \hat{K})\frac{1}{K^2} + K\left(\frac{1}{K} - \frac{1}{\hat{K}}\right)^2\right)}$$

Using the same steps, we obtain

$$\mathbb{E}\left[|\tilde{m}_{i,\text{pro}} - \hat{m}_{i,\text{pro}}|\right] = \sqrt{\frac{2\Sigma_{2,2}}{\pi}\left((K - \hat{K})\frac{1}{K^2} + K\left(\frac{1}{K} - \frac{1}{\hat{K}}\right)^2\right)}$$

Substituting these two values back into equation (EC.11) and re-arranging the terms completes the proof.

Appendix F: Bacteria levels and expiration of deposits

Figure EC.5 shows the distribution of bacteria levels and days to expiration for deposits in Phase 1 compared to Phase 3. The Phase 1 and Phase 3 distributions are quite similar for each characteristic.

Roughly 12% of the deposits had bacteria counts higher than 100 cfus, which is the threshold for a pool failing the bacteria criterion. About 8% of deposits had days to expiration of less than 14 days, which generally meant they had to be used immediately. Bacteria and expiration are important secondary metrics for the milk bank since in both cases they could result in milk being discarded. Hence, they want to keep bacteria failures and milk wastage as low as possible. Furthermore, recall that a key motivation for our **PO-FIFO** model (see Section 5.4) is that our original model, which didn't include the expiration constraint (C8), had a much higher expiration rate than desired. Hence, accounting for expiration was critical.

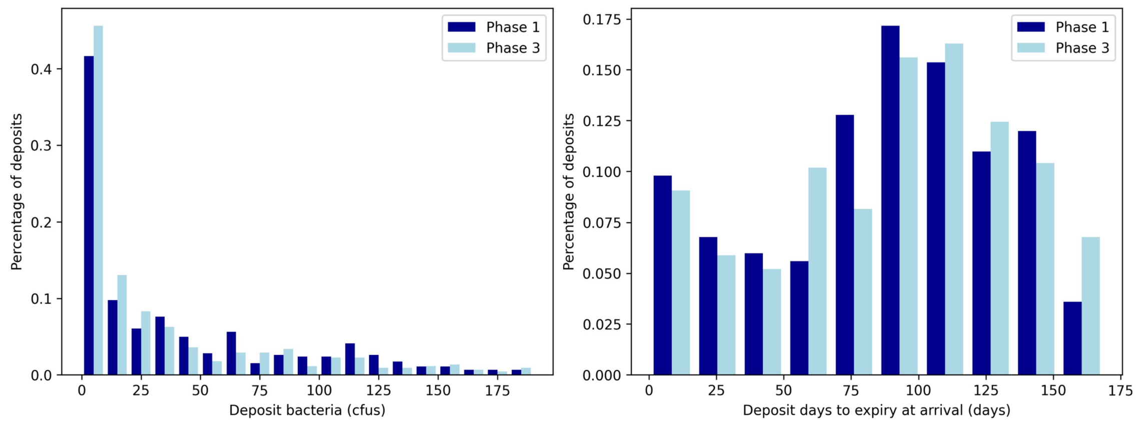


Figure EC.5 **Bacteria counts (left) and days to expiration (right) of deposits in Phase 1 and Phase 3. Days to expiration was measured at the first date at which the deposit could be used in a pool.**

Appendix G: Autocorrelation in the Interrupted Time Series

Figure EC.6 shows partial autocorrelations plotted over a lag value of 25. From these plots, we determined that we should fit our ARIMA models with a lag of two and seven for fat and protein, respectively.

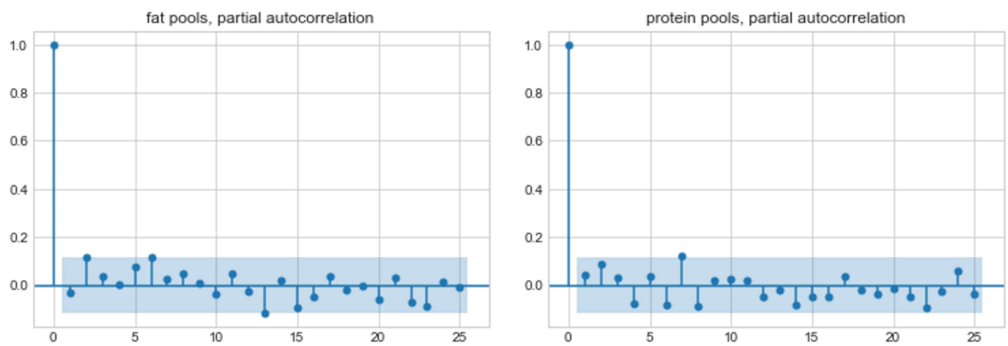


Figure EC.6 **The two figures show partial autocorrelation for fat and protein pool values, respectively. Any lag values that extend beyond the blue upper and lower band, can be considered significant using a 95% confidence interval.**