OPINION DYNAMICS CORPORATION

MEMORANDUM

TO: Roger Baker (ComEd), David Nichols (ComEd)

CC: Jennifer Fagan (Itron), Jeff Erickson (Navigant), Jennifer Hinman (Illinois Commerce Commission)

FROM: Adam Burke, Amanda Dwelley, Bill Norton, Rick Winch (Opinion Dynamics)

DATE: August 10, 2012

RE: Fridge & Freezer Recycle Rewards Program PY4 Metering Study: Savings Results

The PY4 Fridge & Freezer Recycle Rewards (FFRR) program evaluation involved an *in situ* metering study conducted between July 2011 and March 2012. The objectives of this study were to:

- Specify an Illinois-specific regression equation that can be used to estimate gross unit energy consumption (UEC) for the units collected by ComEd's program
- Specify an equation with the same variables as the current lab-based metering regression equation (but using *in situ* metering data) and compare results with the current equation.[1]

This document summarizes the results of gross savings estimation. The methodology and results summary will also be included in the PY4 evaluation report.

1. Metering Study Overview

Data Collection

The evaluation team metered a sample of refrigerators and freezers in participant homes for an average of three weeks prior to their being removed by the program and recycled at JACO's facility. Participants were recruited and screened by telephone. The evaluation team used a monetary incentive and multiple contact attempts to increase the response rate and minimize non-response and selectivity bias.

The metering study collected 5-minute interval demand and average kWh (from power meters), 5-minute interval internal temperature data, and light usage (on/off). We metered the appliance on a

[1] For PY1-PY3, the ComEd program used a regression formula from a meta-analysis of predominantly lab-based metering results that included nearly 1,600 units recycled between 1993 and 2005. See: ADM Associates, Inc., Athens Research, Hiner & Partners, Innovologie LLC (2008). Evaluation Study of the 2004-05 Statewide Residential Appliance Recycling Program.

staggered basis between July 2011 and March 2012. The evaluation team also recorded appliance characteristics that have been associated with energy consumption in previous metering studies, including characteristics that are already recorded by the ComEd FFRR program. Each unit was metered for an average of three weeks before being removed for recycling through the FFRR program.

In total, we metered 121 refrigerators and 34 freezers, resulting in 130 valid sample points for analysis. Table 1 summarizes the units metered and the final metered sample after accounting for unusable data associated with logger malfunctions.

Table 1. Metering Sample Frame and Final Sample

	Refrigerator	Freezer	Total
Total Metered Units	**121**	**34**	**155**
Complete logger failure[a]	13	6	19
Partial logger failure[b]	6	0	6
Units with valid power data	**102**	**28**	**130**

[a] Meter did not record any power data

[b] Meter either (a) recorded less than 1 day of data, or (b) recorded dates & times that could not be aligned with installation times)

Existing Savings Approach

To date, gross savings estimates for FFRR have relied on a regression equation for estimating refrigerator and freezer Unit Energy Consumption (UEC) that is based on a large database of over 2,200 units metered in California using Department of Energy (DOE) laboratory-based metering protocols. The DOE lab test methodology uses a prescribed procedure for metering unit energy consumption, which includes metering each unit at a constant ambient temperature of 90 degrees Fahrenheit. The regression equations derived from the lab-metered data estimate usage as a function of unit characteristics (age, size, configuration, and defrost mode). The characteristics of units collected by JACO for ComEd are then input into these models to estimate full-year UECs (representing kWh savings) that are specific to ComEd's program.

Metering Study Savings Analysis Approach

The energy savings equation was estimated following a two-stage modeling process:

1. During the first stage, we estimated the relationship between observed average hourly demand and outdoor temperature for each unit. We conducted sensitivity analyses to identify the Stage 1 estimation method that provided the best fit of hourly data. We then predicted what average hourly demand would be during typical weather and time periods for each unit in the sample (assuming 30-year typical weather conditions). Hourly estimates are annualized to a full-year UEC by multiplying average hourly demand by 8,766 hours per year.

2. During the second stage, we estimate the relationship between annualized consumption (as predicted in first-stage models) and unit characteristics. We tested the lab-based metering specification and alternative specifications to find the best-fitting model (in terms of explanatory power, relative precision and usefulness for estimating program savings). The coefficients from the second-stage model can be used to re-estimate savings for the PY1-PY3 participant

populations using mean values of appliance characteristics.

The next section provides estimated savings results from the preferred equation (developed from Stage 2 models). The proceeding sections provide more detail on the results of each analysis stage.

2. Savings Results

Savings Algorithm

Based on sensitivity analysis of multiple alternative models conducted to date (including re-estimation of the previous program model) and stakeholder feedback, we recommend the model below for estimating gross UEC for refrigerators and freezers recycled through the ComEd Fridge & Freezer Recycling program.

Table 2. ComEd *in situ* metering model (Model C)
(Dependent variable: Annual UEC in kWh)

(n=130, $R^2 = 0.38$)

Variable Description	Coefficient	Robust t-statistic[2]
Intercept	-103.39	-0.45
Freezer dummy (=1 if freezer)	433.40	2.73
Side-by-side dummy (= 1 if side-by-side)	614.91	3.96
Chest dummy (= 1 if chest freezer)	-490.78	-2.55
Single door dummy (= 1 if single door)	-797.90	-1.80
Age	23.93	3.11
Pre-1993 dummy (=1 if manufactured pre-1993)	289.82	2.00
Cubic Feet	13.52	1.28
Manual defrost dummy (= 1 if manual defrost)	-381.23	-3.03

This model results in lower gross savings estimates than the program has used in previous years. This model is based on primary data from 130 PY4 ComEd program units, and applies to typical weather conditions in ComEd territory.

Savings Results

Here we use the coefficients of the preferred *in situ* regression model to re-estimate gross per unit savings for PY1-PY3 units using each year's summary statistics. Gross per unit savings for refrigerators and freezers in each year are reported in Table 3.

[2] Robust t-statistic use a heteroskedasticity-consistent covariance matrix (HCCM) to adjust standard errors for observed heteroskedasticity (related to magnitude of observed & Stage 1 UEC estimates). We used a version of HCCM called HC3 that has better small-sample properties (n<about 250) than the HC0 robust estimator of variance (a.k.a. Huber or White estimator). The HC3 estimator was first proposed by MacKinnon and White (1985) and is available in Stata 11. Source: Long and Erwin (2000).

Table 3. Re-Estimation of PY1-PY3 Gross Savings by Appliance Type Using Preferred *in situ* Metering Model

Metric	PY1	PY2	PY3
Gross Annual kWh per Refrigerator	**818**	**869**	**937**
Gross kWh RP	10.4%	9.0%	8.4%
Gross Annual kWh per Freezer	**1,238**	**1,083**	**1,220**
Gross kWh RP	12.6%	11.9%	11.8%

Gross and adjusted gross per unit savings (incorporating each year's part-use factor) for the preferred *in situ* model are reported below. Gross savings estimated with this model are about half of gross savings reported in previous evaluation years.

Table 4. Re-estimation of PY1-PY3 Gross and Adjusted Gross Savings Using Preferred *in situ* Metering Model

Metric	PY1	PY2	PY3
Refrigerators and freezers (n)	*11,513*	*25,011*	*39,983*
Gross Annual kWh per unit	**930**	**911**	**980**
Gross kWh RP	7.9%	7.4%	7.4%
Part Use Factor	**0.705**	**0.872**	**0.880**
Part Use RP	7.7%	3.8%	3.4%
Adjusted Annual kWh per unit	**656**	**794**	**862**
Adjusted kWh RP	11.0%	8.4%	8.2%

Discussion

In future program years, the evaluation team believes that the preferred algorithm from this *in situ* metering study will provide more accurate estimates of savings in ComEd territory compared with estimates from the previous algorithm. Gross savings estimates from the ComEd *in situ* models are in line with observed consumption from the metering study sample. Additionally, gross savings estimates are in line with gross savings reported from *in situ* evaluations in Michigan, Ontario and California. Specifically:

1. For their respective program populations, recent *in situ* evaluations in Michigan, Ontario and California reported gross refrigerator UEC of 1,074-1,255 kWh, and gross freezer UEC of 1,173-1,270 per unit

2. Using the ComEd PY3 population characteristics as inputs, regression-based equations from *in situ* metering in Michigan and Ontario predict gross refrigerator UEC of just under 1,000 kWh for refrigerators, and gross freezer UEC of 1,025-1,173 per unit (Table 12)

Estimated gross per unit savings from *in situ* metering are lower than what the program estimated in PY1-PY3. Potential reasons for the difference between *in situ* estimates and the previous lab-based metering estimates include:

3. Annualized, observed unit energy consumption of the metered sample was, on average, about

half of average UEC of previous estimates. Therefore, regression modeling is unlikely to yield a UEC estimate close to the previous estimates, even after accounting for slight differences between the metered sample and program populations (rather, we would expect estimates to be closer to observed UEC).

4. Before adjusting for weather conditions or unit characteristics, the evaluation team observed an average annualized UEC of 957 kWh per year for the 130 units in the metering sample (see the blue box in Table 6). This is about half of the weighted average annual kWh estimates from the previous California lab-based metering model (see Table 11).

5. The previous algorithm was based primarily on units metered in California under DOE protocols, with a constant 90F ambient temperature.

6. About 87% of the units used to develop the previous lab-based metering regression were units whose UEC was estimated using DOE protocols

7. Metering studies using DOE protocols often show higher consumption when compared with *in situ* metering results

8. The sample used to develop the previous model was comprised predominantly of older units (recycled in 1993/1994), with a minority of units recycled after 2000. Partial effects of some appliance characteristics (e.g., age) are likely different within these populations.[3]

9. 72% of the units in the previous algorithm sample were from a 1993-1994 lab-based metering study, 9% from 1998 lab-based metering study, 6% from 2003 lab-based metering study, and 13% from a 2004-2005 dual monitoring study.

The ability of these (or any) models to accurately predict savings from future program populations depends on accurately collecting appliance characteristics that are inputs to the regression equation. Since PY1, the program has improved (reduced) the proportion of units whose configuration or defrost mode is unknown, and should continue these efforts. Supplemental analysis of the sensitivity of these models to alternative age estimation (see Appendix B) revealed some sensitivity in gross savings predictions based on the source of age information (from the metering study or program tracking data). Therefore, we recommend continued focus on data collection quality assurance.

[3] Although a dummy variable for each sample was included, the coefficients are still interpreted as the partial effect of each characteristic (or unit change in the characteristic) holding other factors such as sample constant.

3. Detailed Results

First-Stage Models

Because each unit's weather sensitivity (due to location in home) and time-of-day & day-of-week sensitivity (due to usage patterns) may vary, we tested for different relationships in these factors by including variables other than temperature. We tested seven specifications of bivariate and multivariate models for each unit in the sample, to determine whether inclusion of time-of-day or day-of week terms provided a better fit than temperature alone. The basic form of a model with additional terms is:

$$AvekW_t = Temp_t + PeakHour_t + WeekendHoliday_t + \varepsilon_t$$

With parameters defined as:

$AveKW_t$: Average kW at hour t, based on average of 5-minute interval kW reads across the hour.[4]

$Temp_t$: Average hourly temperature time t, at the weather station closest to the participant's home.[5]

$PeakHour_t$: A dummy variable taking a value of 1 when the hour is in peak hours (1:00-5:00PM CST during standard time and 1:00-5:00PM CDT during daylight savings time)[6]

$WeekendHoliday_t$: A dummy variable taking a value of 1 when the hour falls on a weekend or holiday (a value of 0 would help define PJM performance hours)

ε_t: Idiosyncratic error

We tested each set of models using both hourly data and smoothed hourly data (a moving average of hourly kW and temperature, to smooth spikes in usage related to motor and defrost cycling). Among each set of models (non-smoothed and smoothed), we selected the best-fitting model for each unit primarily based on Akaike information criterion (AIC). Table 5 summarizes the number of appliances whose average hourly demand was best predicted by each model specification. For many units, hourly temperature alone was the best predictor of hourly demand.

[4] In smoothed data, moving average window for hourly kW and temperature includes two hours before and after, to account for some of the longer cycling periods observed.

[5] Weather data for Chicago O'Hare Airport and Rockford Airport comes from the National Climatic Data Center NWS Cooperative Network, provided in hourly and daily format by the Midwestern Regional Climate Center. In some models, we allowed for a non-linear relationship between temperature and hourly kW to account for the possibility that temperature within the home (even in unconditioned spaces) may vary within a narrower range than outdoor ambient temperature.

[6] We also tested using a later peak period (e.g., 3-6pm CDT) and hourly dummies for each peak hour. For most units, a single variable representing the PJM peak period provided better model fit and statistical precision.

Table 5. Hourly Demand Models Used for Estimating Average Hourly Consumption

Model Specification (Independent Variables)	Number of Units with Hourly Demand Best Predicted by Each Model Specification	
	Non-Smoothed Data	Smoothed Data
Temp	44	15
Temp Temp^2	13	18
Temp PeakHour	19	10
Temp WeekendHoliday	11	13
Temp PeakHour WeekendHoliday	10	21
Temp Temp^2 *PeakHour*	16	27
Temp Temp^2 *WeekendHoliday*	17	26

In addition to the models described above, we examined model fit and the distribution of Stage 1 UEC estimates using:

- Temperature only (linear temperature for all units)
- Lagged temperature instead of hourly temperature

These models were used to predict annual UEC at the average Typical Meteorological Year temperature for ComEd territory (50.12 degrees Fahrenheit), and average time-of-day (16.7% peak) and weekend-holiday (31.3%), where appropriate.[7]

We compared different Stage 1 model approaches based on three criteria:

- Explanatory power of Stage 1 models: Smoothed data provided a much better fit of weather-based trends in consumption than hourly data (compare values in second data column of Table 7). (By definition, best-fit models provide a better fit than linear temperature extrapolation.)
- The ability of the models to extrapolate beyond the observation season: We looked at average percentage change in UEC (from observed to predicted) for units metered in cooler periods vs. warmer periods (See rows 9, 11, 13, 15 of Table 6. Since many units were located in unconditioned space, we'd expect that, on average, data collected from units metered in cooler months would show slightly lower average usage for the period we metered than what we might estimate as the unit's annual average if we metered for 365 days; therefore we'd expect a slightly higher UEC after weather adjustment (on average). Similarly, we'd expect that, on average, data collected from units metered in warmer months would show a slightly higher average usage over the period we metered than what might observe as the unit's annual average if we metered for 365 days; therefore we'd expect a slightly lower UEC after weather adjustment (on average). Regardless of extrapolation method, the average predicted UEC of units metered in colder periods is still lower than average predicted UEC of units metered in warmer periods.[8] However, the linear temperature models provided a slightly larger percentage change in observed hourly demand among units metered in warmer periods.
- Explanatory power of Stage 2 models: Finally, we also looked at R-squared and precision of savings estimates for the PY3 population using each set of Stage 1 dependent variables (see

[7] Typical Meteorological Year temperature calculated using 30 year (1982-2011) average daily temperature from O'Hare and Rockford stations, weighted by the PY3 proportion of participants closest to each station (89% O'Hare).

[8] We would not expect complete equality given differences in characteristics within each period.

Stage 2 Model columns of Table 7).[9] Though Stage 1 explanatory power is slightly better for when the best-fit Stage 1 model is used for each unit, there are no major differences in Stage 2 explanatory power.

Considering all of these factors, we recommend using smoothed hourly data for each unit, but including only a linear temperature term (Rows 14 & 15 of Table 6).

Table 6 below shows how predicted UEC from Stage 1 models varies by Stage 1 estimation method (rows 8-17). The blue box shows unadjusted UEC, based on extrapolating average hourly kW for each unit to a full year. It also shows a breakdown of observed and predicted annual UEC by the approximate season of metering (columns). Predicted annual UEC values include:

- Row 8: Predictions using the best-fit Stage 1 modeling approach described above
- Row 10: Predictions using smoothed hourly kW, and taking the best-fit model for each unit
- Row 12: Predictions using hourly temperature alone (linear)
- Row 14: Predictions using smoothed hourly kW, and using temperature only for each unit (linear)

Finally, Table 7 compares average annual gross UEC estimates for the PY3 population using unadjusted UEC as well as predictions from each option for Stage 1 modeling. The bold-outlined box shows average annual UEC estimated from Stage 1 modeling options, compared to unadjusted UEC (top row). In Stage 2 models, using predictions from any Stage 1 model as a dependent variable achieves slightly higher explanatory power and better precision than using unadjusted UEC as the dependent variable (top row). As expected, annual kWh predictions from Stage 1 models are all slightly higher than when using unadjusted UEC, because estimates from Stage 1 models assume a higher temperature than the average temperature observed across metering. Based on these results we recommend using smoothed data and a linear temperature extrapolation for each unit (row 14 in Table 6). Linear temperature extrapolation is also supported by other *in situ* studies (see Consumers Energy Annual Evaluation 2010 Report).

[9] Precision estimates incorporate heteroskedasticity-consistent standard errors (i.e., robust standard errors) that are described in more detail below.

Table 6. Comparison of Observed and Predicted Annual kWh by Season of Metering

Row	Metering Season →	Winter Metering *Under 40 WTHI*	Shoulder Metering[10] *40-60 WTHI*	Summer Metering *60-80 WTHI*	All Units
1	n	55	41	34	130
Season Characteristics					
2	Mean WTHI	37.1	52.1	71.7	50.9
3	Mean Temperature (range)	33.0 (26 - 36.2)	50.8 (39.7 - 59.9)	74.4 (62.2 - 81.2)	49.4 (26 - 81.2)
4	Install Dates	Nov 28, 2011 – Feb 14, 2012	Sep 14, 2011 – Feb 14, 2012	Jul 14, 2011 – Sep 7, 2011	Jul 14, 2011 – Feb 14, 2011
Unit Characteristics					
5	Pct freezers	11%	24%	35%	22%
6	Average Age	21.7	23.1	28.7	24.0
Observed and Predicted kWh					
7	**Avg kWh, Observed** (Annualized from hourly data)	**780**	**923**	**1,284**	**957**
8	**Avg kWh, Predicted** (from *best-fit* Stage 1 models)	870	922	1,226	980
9	Average % Change from Observed	*+11.6%*	*-0.1%*	*-4.5%*	*+2.4%*
10	**Avg kWh, Predicted** (from ***smoothed best-fit*** Stage 1 models)	868	922	1,230	980
11	Average % Change from Observed	*+11.3%*	*-0.1%*	*-4.2%*	*+2.4%*
12	**Avg kWh, Predicted** (from ***linear hourly temperature*** models)	867	918	1,206	972
13	Average % Change from Observed	*+11.1%*	*-0.5%*	*-6.0%*	*+1.6%*
14	**Avg kWh, Predicted** (from ***smoothed linear hourly temperature*** models)	867	918	1,201	971
15	Average % Change from Observed	*+11.2%*	*-0.6%*	*-6.4%*	*+1.4%*
16	**Peak kW** (Predicted for PJM demand models)	0.139	0.135	0.156	0.142

[10] The shoulder season units were metered under similar temperature conditions as TMY temperature in ComEd territory.

Table 7. Predicted Average UEC and Regression Fit among Stage 2 Models, based on Stage 1 Models[11]

Dependent Variable	Stage 1 Modeling Method	Predicted Average UEC (from Stage 1)		California Lab-Based Metering Equation (Stage 2)		Model A (Stage 2)		Model B (Stage 2)	
		UEC Estimate and (CV)	*Average Adjusted R^2*	*R^2*	*PY3 Estimate (RP)*	*R^2*	*PY3 Estimate (RP)*	*R^2*	*PY3 Estimate (RP)*
Observed	*Unadjusted UEC (Annualized from hourly kW)*	*957 (0.67)*	*n/a*	*0.41*	*916 (16.3%)*	*0.40*	*1,063 (8.0%)*	*0.39*	*1,090 (10.5%)*
Stage 1 Prediction Options	Hourly data & Best-fit Stage 1 models	980 (0.66)	0.078	0.41	980 (15.3%)	0.42	1,106 (7.4%)	0.40	1,143 (9.9%)
	Smoothed data & Best-fit Stage 1 models	980 (0.66)	0.210	0.42	958 (15.0%)	0.42	1,088 (7.2%)	0.41	1,127 (9.6%)
	Hourly data & Linear temperature only	972 (0.64)	0.064	0.41	1,002 (15.1%)	0.41	1,114 (7.4%)	0.40	1,162 (9.7%)
	Smoothed data & Linear temperature only *(preferred)*	971 (0.64)	0.178	0.42	956 (14.9%)	0.42	1,088 (7.3%)	0.41	1,127 (9.6%)

Note: Description of Stage 2 Models is below. Summary information is provided here for Stage 1 comparison purposes only.

[11] R-squared (R^2) can be used to compare explanatory power within a group of similar models (e.g., column) but should be interpreted with caution across models (across columns) because it generally increases as more independent variables are added to the model. In this case, the California Lab-Based Metering equation has the most independent variables

Second-Stage Models

After predicting annual UEC at the average Typical Meteorological Year temperature for ComEd territory (50.12 degrees Fahrenheit) for each unit using smoothed hourly data and linear temperature models for each unit, we specified a Stage 2 model identical to the lab-based metering regression equation, to examine similarity of coefficients and model fit. Table 8 compares coefficients in both models.

Table 8. Comparison of Coefficients in California Lab-Based Metering Equation

Variable Description	California Lab-Based Metering Equation (R^2 = 0.43)		ComEd *In situ* Metering Equation (R^2 = 0.42)	
	Coefficient	t-stat	Coefficient	Robust t-stat[12]
Intercept	-422.41	-0.77	-55.79	-0.04
Freezer dummy (=1 if freezer)	169.05	1.84	76.54	0.47
Bottom freezer dummy (=1 if unit is bottom freezer)	595.38	2.91	34.92	0.21
Side-by-side dummy (= 1 if unit is side-by-side)	-129.36	-0.34	117.59	0.23
Single door dummy (= 1 if unit is single door)	-417.10	-4.73	-575.86	-2.40
Frost free dummy (= 1 if unit is frost free)	-445.03	-1.00	-1561.40	-1.17
Natural log of unit age	405.21	2.15	113.69	0.24
Cubic Feet of unit (per tracking system data)	43.65	4.59	17.44	1.03
Label Amps	104.10	4.83	11.28	0.78
Freezer dummy x frost free dummy	319.11	1.94	329.34	1.11
Bottom freezer dummy x frost free dummy	-302.05	-1.28	(omitted due to collinearity)	
Side-by-side dummy x frost free dummy	1451.32	3.80	633.63	1.42
Side-by-side dummy x amps	-126.43	-2.88	-9.60	-0.19
Frost free dummy x ln(age)	299.82	2.09	519.59	1.28
Dummy if mfg. year is 1990 or earlier[13]	1197.83	2.61	-289.36	-0.17
Ln(age) x age 15 up dummy	-524.98	-3.08	158.39	0.30

[12] Robust t-statistic use a heteroskedasticity-consistent covariance matrix (HCCM) to adjust standard errors for observed heteroskedasticity (related to magnitude of observed & Stage 1 UEC estimates). We used a version of HCCM called HC3 that has better small-sample properties (n<about 250) than the HC0 robust estimator of variance (a.k.a. Huber or White estimator). The HC3 estimator was first proposed by MacKinnon and White (1985) and is available in Stata 11. Source: Long and Erwin (2000).

[13] This dummy variable was intended to represent units manufactured before 1990, though it is sometimes stated as a dummy variable for age. Therefore, for comparison purposes we used a dummy variable equal to 1 if the unit was manufactured prior to 1990.

All but three coefficients in the *in situ* model share a similar direction as the lab-based model, though the magnitudes vary. The last column of Table 8 shows that when all of the coefficients and interaction effects are included in a single model, few partial effects are significant at a 90% confidence level (two-tailed), including terms involving side-by-side units, age, manufacturing year, and size. Relatively low statistical significance is likely due to collinearity between terms (that may be more pronounced in the smaller ComEd metering sample than the California lab-based metering sample).

If this *in situ* model is used to re-estimate savings for PY1-PY3, the relative precision around the per unit gross savings estimates (at 90% confidence in a one-tailed test) exceeds 10%. Table 11 shows potential results. Because precision falls below evaluation standards, we searched for more appropriate model specifications.

We considered all relevant appliance characteristics available in program tracking data for inclusion in alternative models, such as:

1. Dummy variables for all configurations (top freezer, bottom freezer, side-by-side, single-door, chest and upright freezers).
2. Dummy variables for appliance features (e.g., manual defrost and through-door ice)
3. Alternate specification of continuous variables – age, cubic feet, label amps (e.g., squared term or natural log of age)[14]
4. Appliance vintage - Dummy variable for manufacturing year before 1990 (when first National Appliance Energy Conservation Act standards became effective) or before 1993 (first update of NAECA standards)
5. Location in home - The program tracks location in home, but does not currently track summer or winter space conditioning. Therefore, location in home served as a proxy for potential weather sensitivity – for example, units located in a garage, porch or patio may show more sensitivity to climactic conditions.
6. Number of occupants in the home - The program does not track the number of occupants, though this variable was available through primary data collection.
7. Interaction terms

We examined model fit, precision, and statistical significance of individual terms under different specifications.[15] We also tested separate refrigerator and freezer models.

To select the most appropriate model for future FFRR program savings estimation, we weighed criteria such as:

- Model fit (explanatory power)
- Relative precision (using the PY1-PY3 participant population characteristics)
- Savings estimates for different appliance configurations (relative to observed UEC (annualized from hourly), Stage 1 estimates, evaluation of savings for each configuration in other recent *in*

[14] We also examined sensitivity to modeling age based on manufacturing year recorded in the program tracking database rather than age collected in the metering study. Results are shown in Appendix B.

[15] We began by assessing the variables and interaction terms that were significant in peak demand model, used to estimate peak demand for PJM purposes.

situ models, and evaluation of savings for each configuration in original California lab-based metering model.

In addition to re-estimation of the California lab-based metering model, we identified four models that met at least one of these criteria. The main difference between the models is how age and vintage are specified. All models include a continuous age term, which accounts for degradation over time. It is believed that the marginal impact of degradation decreases over time, and this appears to be confirmed by the negative sign on Age-squared variable in Model A.[16] There may also be a "vintage" effect based on manufacture before 1990 or 1993 NAECA efficiency standards, which may cause a difference in efficiency independent of other characteristics (that are included in the model). Models B-C allow for age as well as a separate "vintage" partial effect related to efficiency standards. Model B.1 allows for a difference in the slope of age (degradation year-over-year) based on vintage cohort.[17]

The regression coefficients and t-statistics of versions of Model A are presented in Table 9. This table also shows coefficients of analogous models run only with refrigerators and freezers. Because results are consistent between the pooled and separate models (Average UEC by appliance type and configuration) but precision is better for the pooled model, we recommend using a pooled model. Results from other models are shown with pooled models only.[18]

The regression coefficients of Model B, C and B.1 are shown in

[16] Language developed through collaboration with the Ameren evaluation team (The Cadmus Group)

[17] Though Lawrence-Berkeley National Laboratory research on average annual energy consumption for new refrigerators and freezers over time (Figure 1) suggests a more pronounced change in slope in 1993, our sample supported a more detectable change in slope in 1990 rather than 1993 (based on significance of age & vintage interaction in model B.1 and an analogous model with a dummy for 1993 and its interaction), likely because a greater majority of units in the sample (and PY3 population) were manufactured before 1993 (providing fewer sample points to determine a different marginal effect).

[18] Only results from separate appliance models are shown for Model A, because separate models for Models B and C yielded lower explanatory power and less precision than separate models for Model A.

Table 10. Here we discuss the trade-offs of each model:

Re-Estimation of the California lab-based metering model: This model controls for multiple interactions between configurations and features. Similar to Models A-C, point estimates for different configurations are in line with observed UEC. However, the lower precision around estimates (related to collinearity) does not make this model ideal for evaluation purposes.

Model A: Model A controls for multiple configurations and represents age non-linearly (with an age and age-squared term). It has slightly higher explanatory power that Models B and C, and better relative precision than B.1.

Models B: Model B controls for multiple configurations and represents age as well as manufacturing year prior to 1990. While explanatory power is not quite as high as Model A, the inclusion of a vintage dummy representing manufacture pre- or post-NAECA standards is consistent with other *in situ* and lab-based metering studies.

Model C: Model C is similar to Model B, but includes a dummy for manufacturing year prior to the 1993 update to NAECA standards rather than 1990. Research on average annual energy consumption of new appliances suggests a more pronounced change in slope before vs. after 1993 compared with before vs. after 1990 (see Appendix Figure 1). Coefficients and results of models B and C are similar. Both models have relatively strong precision.

Model B.1: This variation on Model B allows for a different slope on age in the periods before and after the original NAECA standards, and significant coefficient on the analogous ln(age) x pre-1990 interaction in the California lab-based metering model). Though a lower marginal effect of age among pre-1990 units is supported by the model, precision is not as strong as simpler models.

Table 9. Coefficients and T-Statistics of Model A

	Pooled Model A ($n=130$, $R^2=0.42$)		Refrigerators Only ($n=102$, $R^2=0.42$)		Freezers Only ($n=28$, $R^2=0.51$)	
Variable Description	Coefficient	Robust t-statistic	Coefficient	Robust t-statistic	Coefficient	Robust t-statistic
Intercept	-695.70	-2.42	-1034.59	-2.37	-60.50	-0.09
Freezer dummy (=1 if freezer)	410.79	2.66			-551.57	-2.53
Side-by-side dummy (= 1 if side-by-side)	655.92	4.35	631.99	4.31		
Chest dummy (= 1 if chest freezer)[19]	-399.27	-2.21				
Single door dummy (= 1 if single door)	-567.58	-3.11				
Age	87.55	6.02	96.01	5.41	79.28	1.75
Age-squared	-0.94	-4.20	-1.09	-4.09	-0.74	-0.86
Cubic Feet	7.40	0.72	18.13	1.22	12.23	0.41
Manual defrost dummy (= 1 if manual defrost)[20]	-350.58	-2.90	-202.90	-1.46	-706.23	-2.74
Overall PY3 Estimate (RP)	1,088 (7.3%)					
Refrigerator Estimate (RP)	1,037 (7.9%)		1,037 (8.4%)			
Freezer Estimate (RP)	1,344 (10.5%)				1,385 (13.1%)	

[19] We also tested analogous models without the chest dummy, because before controlling for features like age and defrost, the chest freezers in our sample showed a smaller difference in UEC from upright freezers than model coefficients predict. However, the chest coefficient remained strong across models, and other sources suggest that chest freezers may be more energy efficient than upright, because less cold air flows out when you open chest freezers (whereas upright freezers may lose more cold air as it flows down and out. (Source: Energy Star - http://www.energystar.gov/index.cfm?fuseaction=find_a_product.showProductGroup&pgw_code=FRZ and Natural Resources Canada - http://oee.nrcan.gc.ca/equipment/appliance/3906)

[20] We also tested the interaction between manual defrost and freezers in pooled models, and although this interaction is significant and adds explanatory power to the models, it results in less realistic estimates for different configurations (e.g., larger overstatements and understatements of average UEC per configuration relative to what was observed and what other *in situ* studies have found).

Table 10. Coefficients and T-Statistics of Models B and C

	Model B ($n=130$, $R^2=0.38$)		Model C ($n=130$, $R^2=0.38$)		Model B.1 ($n=130$, $R^2=0.41$)	
Variable Description	Coefficient	Robust t-statistic	Coefficient	Robust t-statistic	Coefficient	Robust t-statistic
Intercept	-3.91	-0.02	-103.39	-0.45	-412.46	-1.57
Freezer dummy (=1 if freezer)	406.94	2.63	433.40	2.73	429.88	2.67
Side-by-side dummy (= 1 if side-by-side)	596.29	3.85	614.91	3.96	628.11	4.18
Chest dummy (= 1 if chest freezer)[21]	-471.28	-2.52	-490.78	-2.55	-429.42	-2.18
Single door dummy (= 1 if single door)	-805.01	-2.17	-797.90	-1.80	-732.71	-3.56
Age	20.19	2.42	23.93	3.11	49.12	4.24
Pre-1990 dummy (=1 if manufactured pre-1990)	344.49	2.02			1,030.80	2.86
Pre-1993 dummy (=1 if manufactured pre-1993)			289.82	2.00		
Cubic Feet	14.20	1.31	13.52	1.28	10.84	1.00
Manual defrost dummy (= 1 if manual defrost)	-362.05	-2.86	-381.23	-3.03	-363.39	-2.94
Age X Pre-1990					-36.27	-2.47
Overall PY3 Estimate (RP)	997 (7.5%)		980 (7.4%)		1,127 (9.6%)	
Refrigerator Estimate (RP)	956 (8.4%)		937 (8.4%)		1,081 (10.2%)	
Freezer Estimate (RP)	1,232 (11.5%)		1,220 (11.8%)		1,340 (12.1%)	

[21] We also tested analogous models without the chest dummy, because before controlling for features like age and defrost, the chest freezers in our sample showed a smaller difference in UEC from upright freezers than model coefficients predict. However, the chest coefficient remained strong across models, and other sources suggest that chest freezers may be more energy efficient than upright, because less cold air flows out when you open chest freezers (whereas upright freezers may lose more cold air as it flows down and out. (Source: Energy Star - http://www.energystar.gov/index.cfm?fuseaction=find_a_product.showProductGroup&pgw_code=FRZ and Natural Resources Canada - http://oee.nrcan.gc.ca/equipment/appliance/3906)