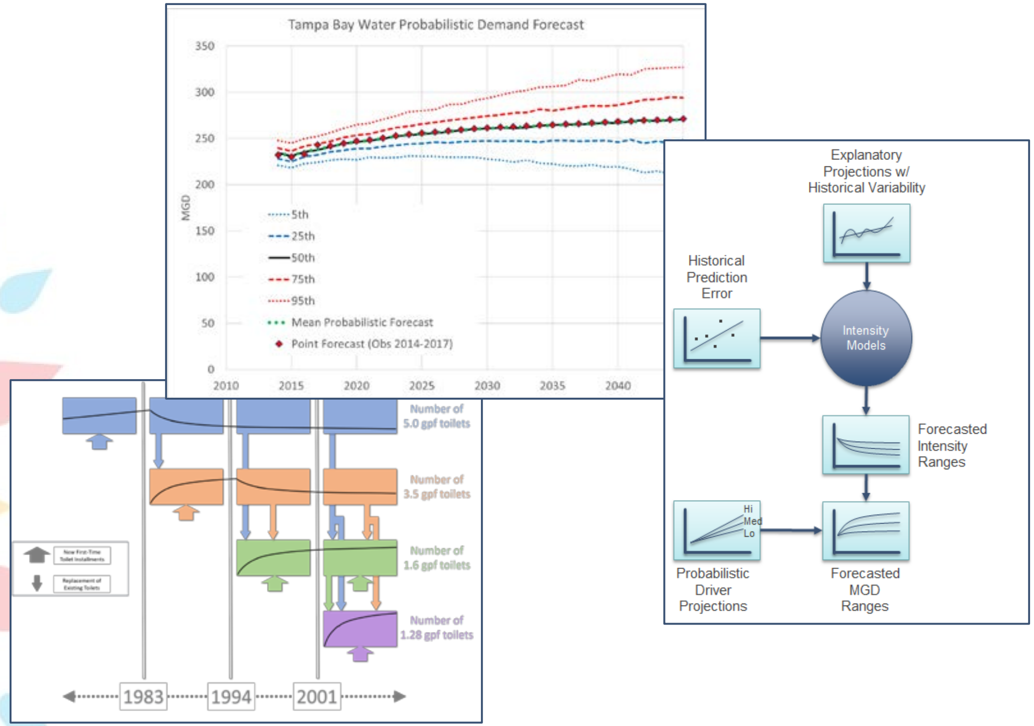
This presentation premiered at WaterSmart Innovations

watersmartinnovations.com

watersmart
INNOVATIONS

Hazen



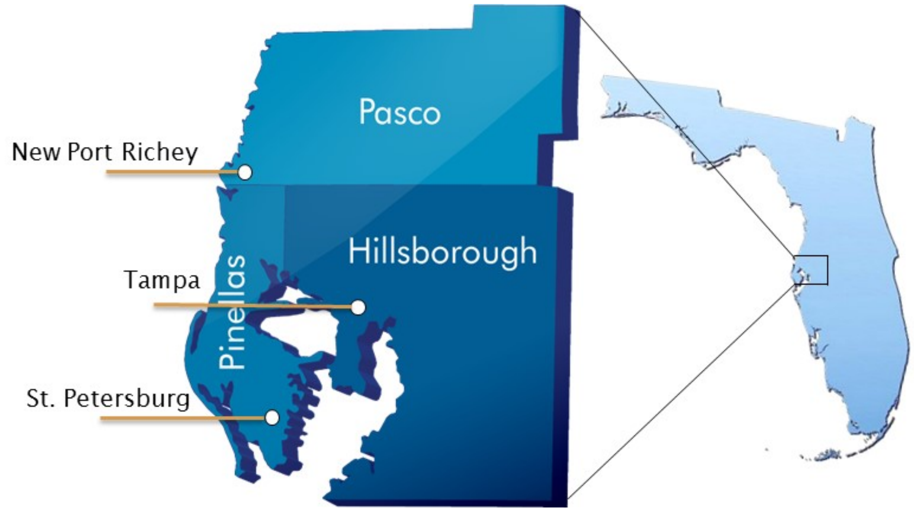
Is Future Water Use Efficiency a Significant Demand Influence Given Forecast Uncertainty?

John Clayton, Jack Kiefer, Lisa Krentz
Hazen and Sawyer

Dave Bracciano
Solomon Erkyihun, Tirusew Asefa
Tampa Bay Water

Example: Tampa Bay Water

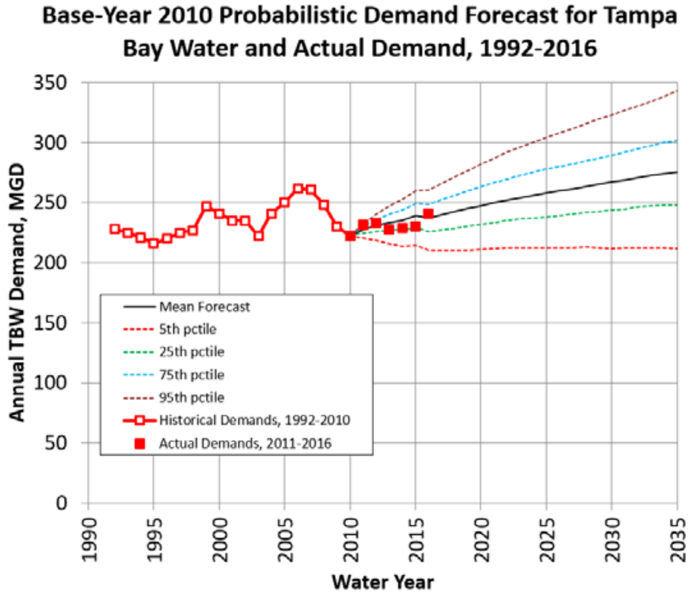
- Wholesaler to 6 utilities in Tampa Bay Region
  - 7 Water Demand Planning Areas (WDPAs)
- Fastest growing water provider in state
  - 2.5 million people
  - Approx. 250 MGD demand
  - Demands susceptible to swings in economy and strong seasonal climate variation



Demand Forecasting at Tampa Bay Water

Since 2002

- Per-unit models: econometric regressions
- Annual updates with new observations and projections
- Probabilistic forecasts every few years
    - Distributions of predictors at future times



Model Development: Sectoral Rate of Use Times Drivers

- Determine past number of water using entities (drivers)
  - SF and MF: number of dwelling units
  - NR: developed square footage in thousands (ksf)
- Determine past demand per driver (demand intensity)
  - SF and MF: gallons per housing unit per day
  - NR: gallons per ksf per day

$$Q_t = n_t f(\mathbf{X}_t)$$

Total demand at t

Number of drivers

Intensity model: function of explanatory variables

Model Development: Sectoral Rate of Use Times Drivers

- Intensity models: historical intensity regressed against explanatory variables
  - Income, Price of water
  - Housing density, PPH
  - Reclaimed service penetration
  - End use efficiency
  - Weather assumptions

$$Q_t = n_t f(\mathbf{X}_t)$$

Total demand at t

Number of drivers

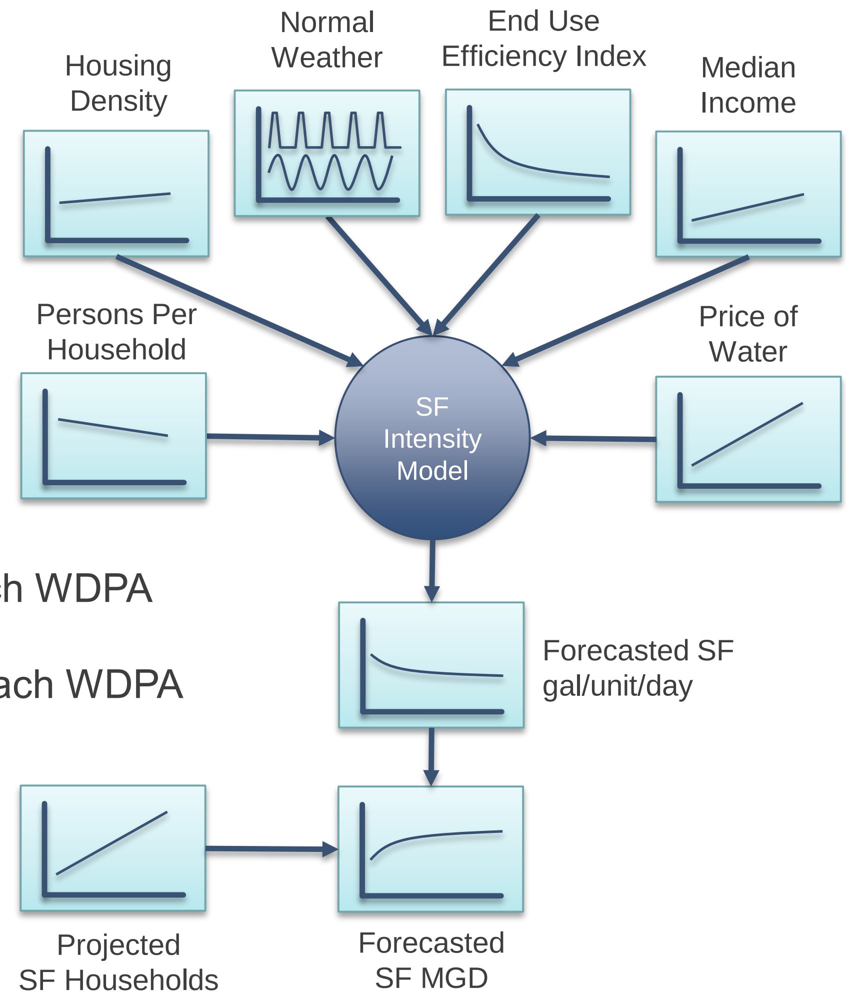
Intensity model: function of explanatory variables

“Point” Forecast

- Input projection time series for each WDPA

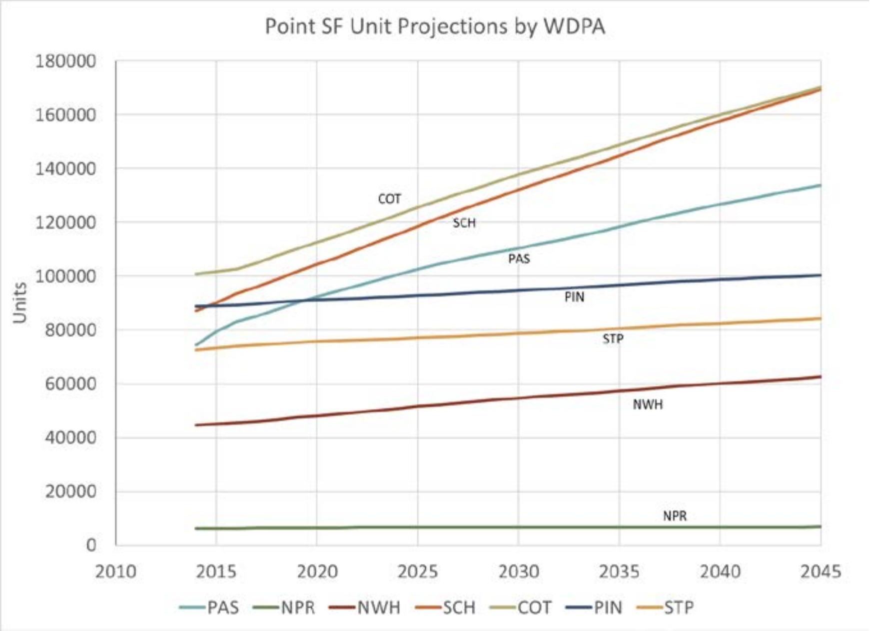
→ Projected explanatory variables
→ Intensity models
→ Forecasted sectoral intensities for each WDPA
→ Multiply by projected driver variables
→ Forecasted sectoral demands for each WDPA

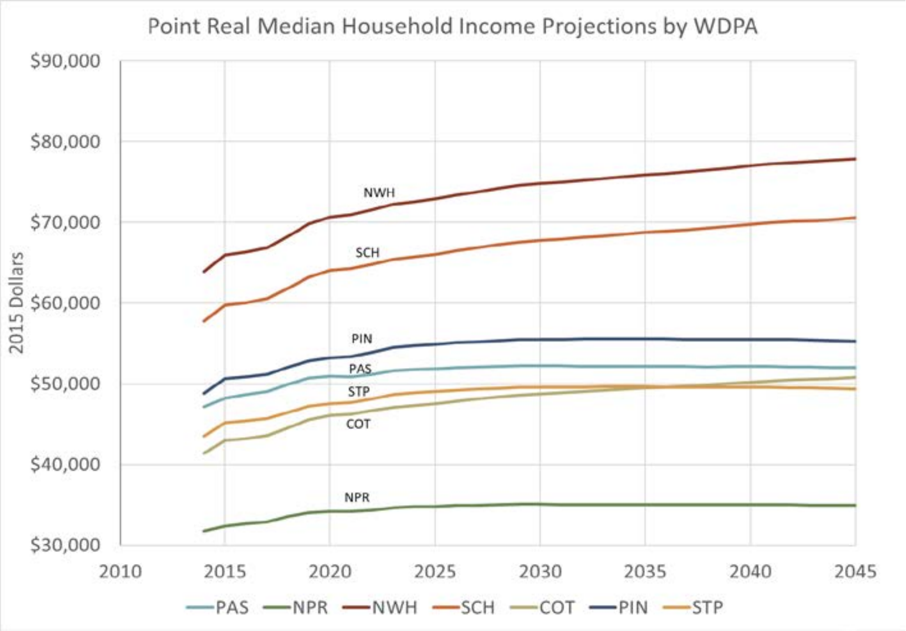
- “Point”: one demand value for each future year, WDPA, sector



Point Projection Examples

Number of SF Households and Median Income for Each WDPA

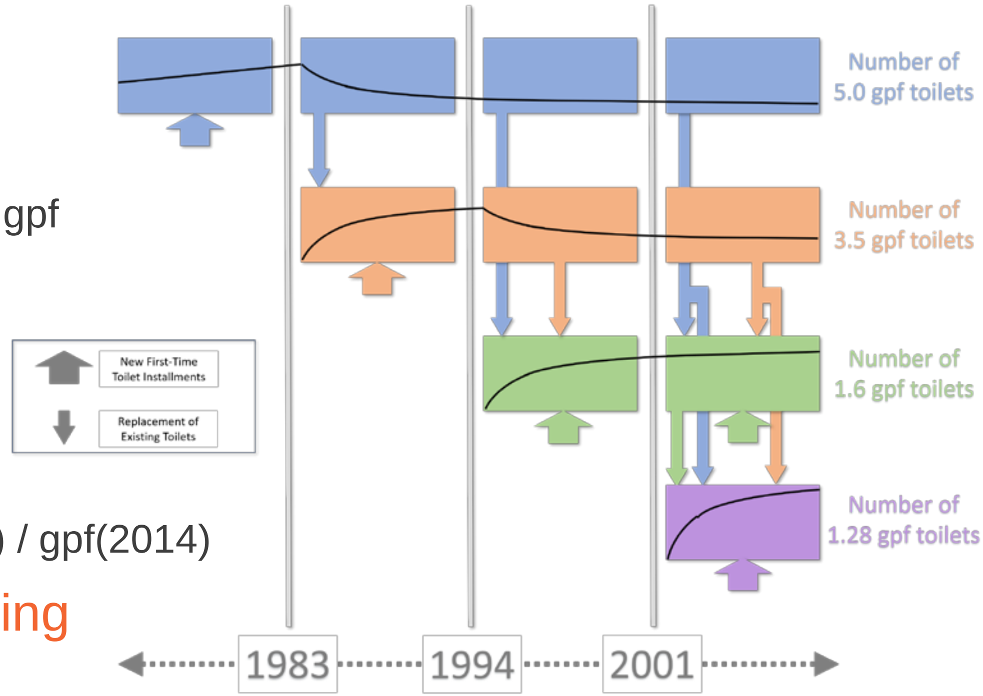




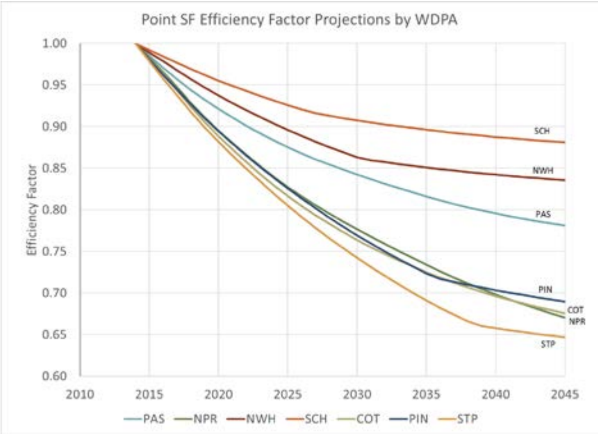
Efficiency Factor

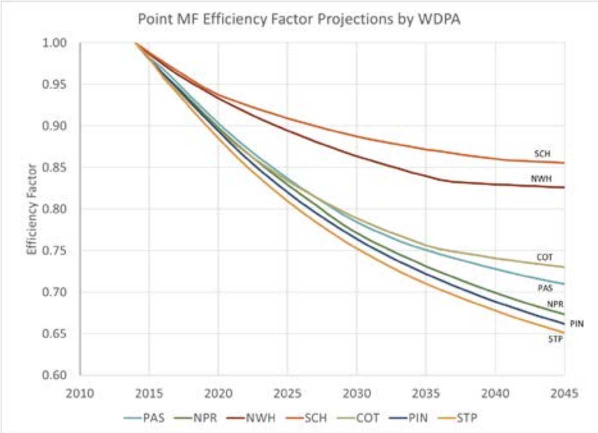
Simplified End-Use Model to Project Passive Efficiency Effects

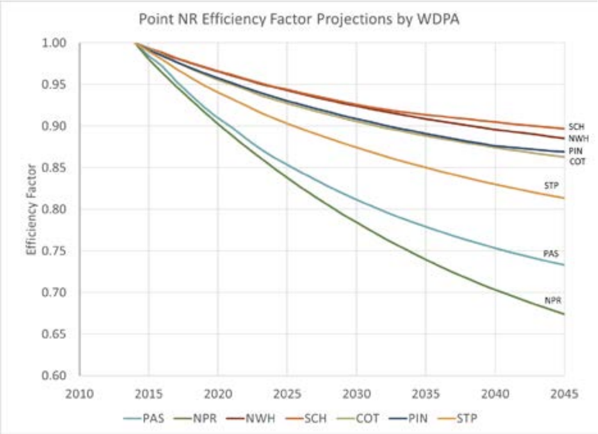
- Stock model for toilets by gpf levels
  - Simulate # of 5.0, 3.5, 1.6, 1.28 gpf toilets from past into future
  - Add/replace certain % of existing toilets each year at standard gpf level for year
  - Determine average gpf and index to a single year e.g. gpf(y) / gpf(2014)
- Use factors as X values during regression
  - Toilet-based factors as proxy for all efficiency



Efficiency Factor Point Projections by WDPA and Sector







Efficiency Factor is a Significant Predictor

- Elasticities
  - SF: 0.35 (p = 0)

Table 3-2: Single-Family Model

Dependent Variable: $\ln(q_{SF,m,y,g})$	Method: Panel Least Squares	Date: 07/11/18 Time: 15:28
Periods included: 141	Cross-sections included: 566	Total panel (unbalanced) observations: 76326
Effects Specification: Month fixed (shown), Cross-section fixed (not shown)	Sample: 2001M10 2013M09 IF SF_GPUD>25 AND UNITS_SFAM>29 AND HURRICANE=0 AND HURRICANE2*PAS=0 AND COT*BAN=0 AND SFK>0	

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Intercept	2.814422	0.619799	4.540861	0
$\ln(INC_{m,y,g})$	0.279786	0.055439	5.046747	0
$\ln(RMP_{m,y,g})$	-0.33135	0.044889	-7.38143	0
$\ln(PPH_{SF,m,y,g})$	0.36065	0.0577	6.2504	0
$\ln(DEN_{SF,m,y,g})$	-0.12691	0.016815	-7.54739	0
$RECL_{SF,m,y,g}$	-0.39061	0.026821	-14.5635	0
$\ln(EFF_{SF,m,y,g})$	0.349646	0.062689	5.57748	0
δ_{Jan}	-0.02373	0.007981	-2.9733	0.0029
δ_{Feb}	-0.01368	0.009876	-1.38551	0.1659
δ_{Mar}	0.038455	0.010807	3.558523	0.0004
δ_{Apr}	0.135433	0.011408	11.87127	0
δ_{May}	0.178004	0.011843	15.03053	0
δ_{Jun}	0.086731	0.011951	7.25707	0
δ_{Jul}	-0.01559	0.011718	-1.33003	0.1835
δ_{Aug}	-0.03328	0.011594	-2.87063	0.0041
δ_{Sep}	-0.01772	0.011179	-1.58551	0.1129
δ_{Oct}	0.025505	0.01004	2.540248	0.0111
δ_{Nov}	0.031506	0.007951	3.962319	0.0001
$dMAXT_{m,y,g}$	0.212444	0.06484	3.276423	0.0011
$dMAXT_{m1,y1,g}$	0.218619	0.065303	3.347789	0.0008
$dMAXT_{m2,y2,g}$	0.13272	0.062627	2.119207	0.0341
$dPRCP_{m,y,g}$	-0.02456	0.002302	-10.6711	0
$dPRCP_{m1,y1,g}$	-0.01811	0.002613	-6.93177	0
$dPRCP_{m2,y2,g}$	-0.00419	0.002303	-1.8192	0.0689
φ	0.563281	0.011514	48.92306	0

Effects Specification: Cross-section fixed (dummy variables)

R-squared	0.907431	Mean dependent var	5.325801
Adjusted R-squared	0.906712	S.D. dependent var	0.356105
S.E. of regression	0.108766	Akaike info criterion	-1.59154
Sum squared resid	895.9545	Schwarz criterion	-1.5201
Log likelihood	61328.03	Hannan-Quinn criter.	-1.56959
F-statistic	1260.485	Durbin-Watson stat	1.598854
Prob(F-statistic)	0	Inverted AR Roots	0.56

Efficiency Factor is a Significant Predictor

- Elasticities
  - SF: 0.35 (p = 0)
  - MF: 0.20 (p = 0)

Table 3-3: Multifamily Model

Dependent Variable: $\ln(q_{MF,m,y,g})$	Method: Panel Least Squares	Date: 07/11/18 Time: 15:31
Periods included: 141	Cross-sections included: 407	Total panel (unbalanced) observations: 48704
Effects Specification: Month fixed (shown), Cross-section fixed (not shown)	Sample: 2001M10 2013M09 IF MF_GPUD>25 AND PID_TOTALMF>2 AND COT*BAN=0 AND HURRICANE=0 AND HURRICANE2*PAS=0	

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Intercept	4.572712	0.45831	9.977338	0
$\ln(INC_{m,y,g})$	0.074048	0.039776	1.861612	0.0627
$\ln(RMP_{m,y,g})$	-0.183221	0.039907	-4.591195	0
$\ln(DEN_{MF,m,y,g})$	-0.106424	0.013996	-7.603786	0
$\ln(EFF_{MF,m,y,g})$	0.203825	0.045988	4.432108	0
δ_{Jan}	0.009893	0.005429	1.82227	0.0684
δ_{Feb}	0.018756	0.006643	2.823654	0.0047
δ_{Mar}	0.02626	0.00731	3.592104	0.0003
δ_{Apr}	0.034297	0.007739	4.431637	0
δ_{May}	0.031516	0.008047	3.916633	0.0001
δ_{Jun}	0.011118	0.007921	1.403617	0.1604
δ_{Jul}	-0.00336	0.007845	-0.428255	0.6685
δ_{Aug}	-0.005582	0.007773	-0.718181	0.4726
δ_{Sep}	-0.018143	0.007554	-2.401862	0.0163
δ_{Oct}	-0.009925	0.006756	-1.469112	0.1418
δ_{Nov}	-0.004585	0.005327	-0.860719	0.3894
$dMAXT_{m,y,g}$	0.031377	0.043703	0.717975	0.4728
$dMAXT_{m1,y1,g}$	0.066076	0.042541	1.553256	0.1204
$dPRCP_{m,y,g}$	-0.006244	0.001543	-4.045667	0.0001
$dPRCP_{m1,y1,g}$	-0.002885	0.001552	-1.859113	0.063
φ	0.574551	0.008928	64.35672	0

R-squared	0.833902	Mean dependent var	4.738196
Adjusted R-squared	0.832436	S.D. dependent var	0.408044
S.E. of regression	0.167031	Akaike info criterion	-0.732547
Sum squared resid	1346.896	Schwarz criterion	-0.655452
Log likelihood	18265.98	Hannan-Quinn criter.	-0.708368
F-statistic	568.9584	Durbin-Watson stat	1.716229
Prob(F-statistic)	0	Inverted AR Roots	0.57

Hazen

Efficiency Factor is a Significant Predictor

- Elasticities
  - SF: 0.35 (p = 0)
  - MF: 0.20 (p = 0)
  - NR: 0.30 (p = 0)

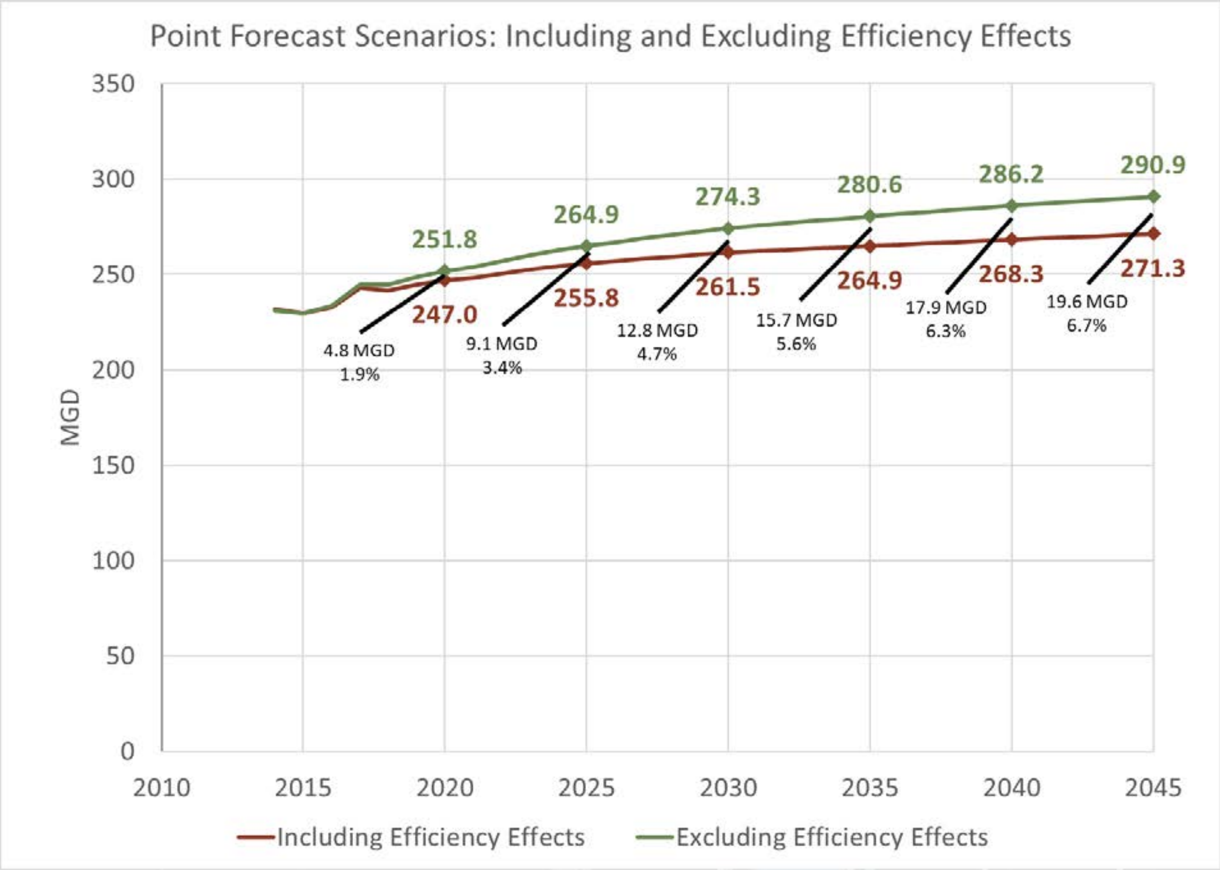
Table 3-4: Nonresidential Model

Dependent Variable: $\ln(q_{NR,m,y,g})$	Method: Panel Least Squares	Date: 07/11/18 Time: 15:41
Periods included: 141	Cross-sections included: 598	Total panel (unbalanced) observations: 80251
Effects Specification: Month fixed (shown), Cross-section fixed (not shown)	Sample: 2001M10 2013M09 IF GPKSFD>0 AND HURRICANE=0 AND HURRICANE2*PAS=0 AND COT*BAN=0 AND LAND_TO_BUILD>1	

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Intercept	1.868465	0.739679	2.526048	0.0115
$\ln(INC_{m,y,g})$	0.305654	0.06617	4.619213	0
$\ln(RMP_{m,y,g})$	-0.30873	0.059961	-5.14881	0
$RECL_{NR,m,y,g}$	-0.45621	0.075847	-6.01489	0
$\ln(EFF_{NR,m,y,g})$	0.300565	0.072236	4.160909	0
$SQFR_{EDUC,m,y,g}$	-0.63858	0.03366	-18.9717	0
$SQFR_{GOVT,m,y,g}$	-0.47841	0.071345	-6.7056	0
$SQFR_{HCAR,m,y,g}$	0.163439	0.060703	2.692441	0.0071
$SQFR_{HMFG,m,y,g}$	0.087706	0.16429	0.533848	0.5934
$SQFR_{HOTL,m,y,g}$	0.711577	0.066937	10.63062	0
$SQFR_{LMFG,m,y,g}$	-0.13417	0.076499	-1.75391	0.0795
$SQFR_{OFFC,m,y,g}$	-0.24981	0.045792	-5.45533	0
$SQFR_{FOOD,m,y,g}$	2.309579	0.102592	22.51224	0
$SQFR_{RETL,m,y,g}$	-0.12994	0.028561	-4.54971	0
$SQFR_{RETR,m,y,g}$	0.378146	0.054798	6.900708	0
δ_{Jan}	0.01534	0.006066	2.528845	0.0114
δ_{Feb}	0.048201	0.007674	6.280919	0
δ_{Mar}	0.060785	0.008699	6.987341	0
δ_{Apr}	0.082565	0.009365	8.816807	0
δ_{May}	0.098664	0.009822	10.04564	0
δ_{Jun}	0.049379	0.009768	5.055241	0
δ_{Jul}	0.014062	0.009664	1.455157	0.1456
δ_{Aug}	0.027549	0.009419	2.924673	0.0034
δ_{Sep}	0.040224	0.008942	4.498074	0
δ_{Oct}	0.051633	0.007701	6.704787	0
δ_{Nov}	0.037068	0.005885	6.298583	0
$dMAXT_{m,y,g}$	0.094306	0.049083	1.921377	0.0547
$dMAXT_{m1,y1,g}$	0.114894	0.047925	2.397362	0.0165
$dPRCP_{m,y,g}$	-0.00954	0.001709	-5.57805	0
$dPRCP_{m1,y1,g}$	-0.00659	0.001725	-3.81917	0.0001
φ	0.722715	0.008785	82.26225	0

R-squared	0.865369	Mean dependent var	4.479423
Adjusted R-squared	0.864309	S.D. dependent var	0.782385
S.E. of regression	0.288201	Akaike info criterion	0.357481
Sum squared resid	6613.493	Schwarz criterion	0.430202
Log likelihood	-13716.09	Hannan-Quinn criter.	0.37977
F-statistic	816.2578	Durbin-Watson stat	2.177301
Prob(F-statistic)	0	Inverted AR Roots	0.72

Point Forecast (2014-2016 Base Period)



Demand Forecast Uncertainty

Nobody knows the future perfectly

$$Q_t = n_t f(\mathbf{X}_t)$$

Unit growth or development rates may differ from expectations

Demand Forecast Uncertainty

Nobody knows the future perfectly

$$Q_t = n_t f(\mathbf{X}_t)$$

Unit growth or development rates may differ from expectations

Intensity variations can arise from deviations in explanatory variables

Demand Forecast Uncertainty

Nobody knows the future perfectly

$$Q_t = n_t f(\mathbf{X}_t) + \epsilon_t$$

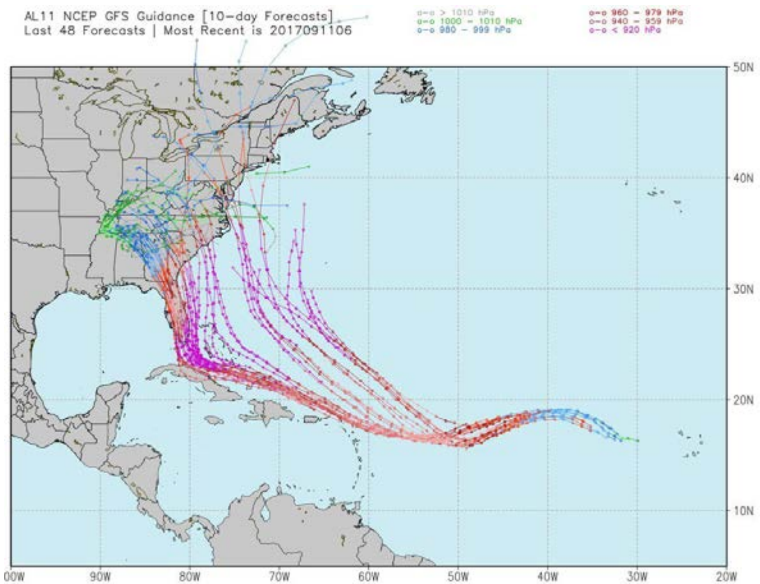
Unit growth or development rates may differ from expectations

Intensity variations can arise from deviations in explanatory variables

Possible statistical error even if drivers and explanatories known perfectly

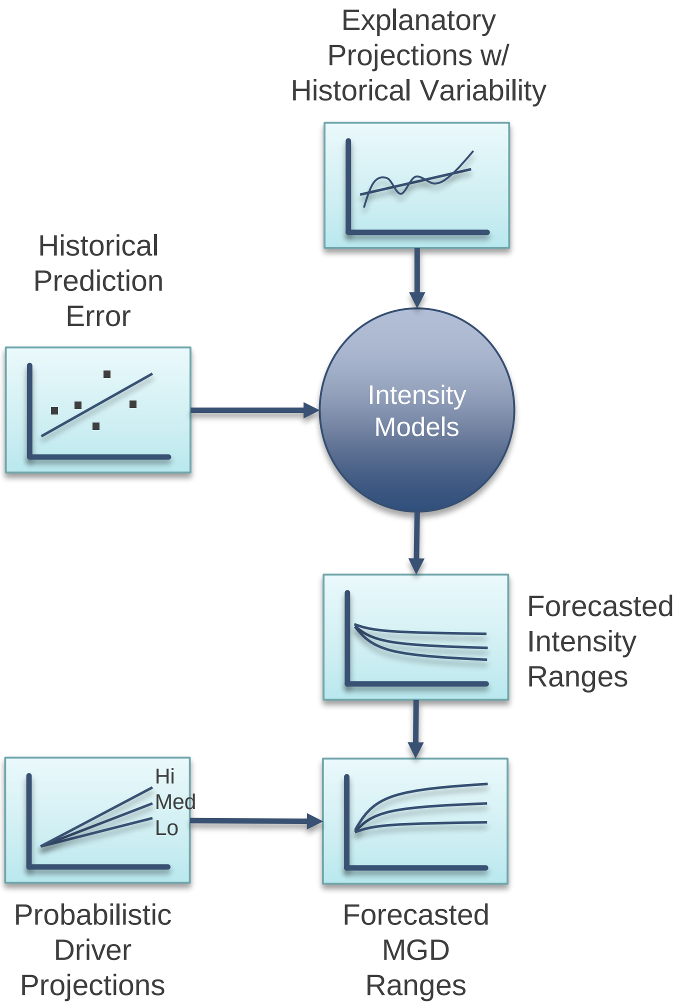
Incorporating Uncertainty in Forecasts

- Determine ranges/scenarios/distributions of projected model inputs
- Produce multiple forecast outcomes and summarize range of possible demands
- Think "hurricane track forecasts", but with demand vs time



Tampa Bay Water Approaches

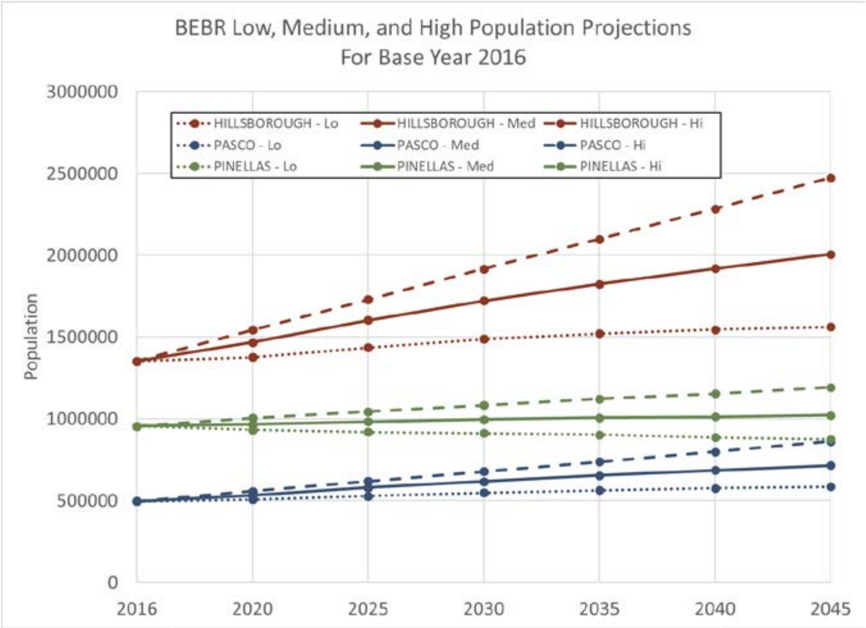
- Uncertainty incorporated in multiple aspects
  - Drivers – population projection distributions provided externally
  - Explanatory variables – uncertainty derived from historical variability
  - Model uncertainty – derived from comparing historical model estimates with observations
- Monte Carlo approach
  - Sample input distributions and calculate demands repeatedly



Uncertainty Assumption Examples

Direct Third-Party Projections

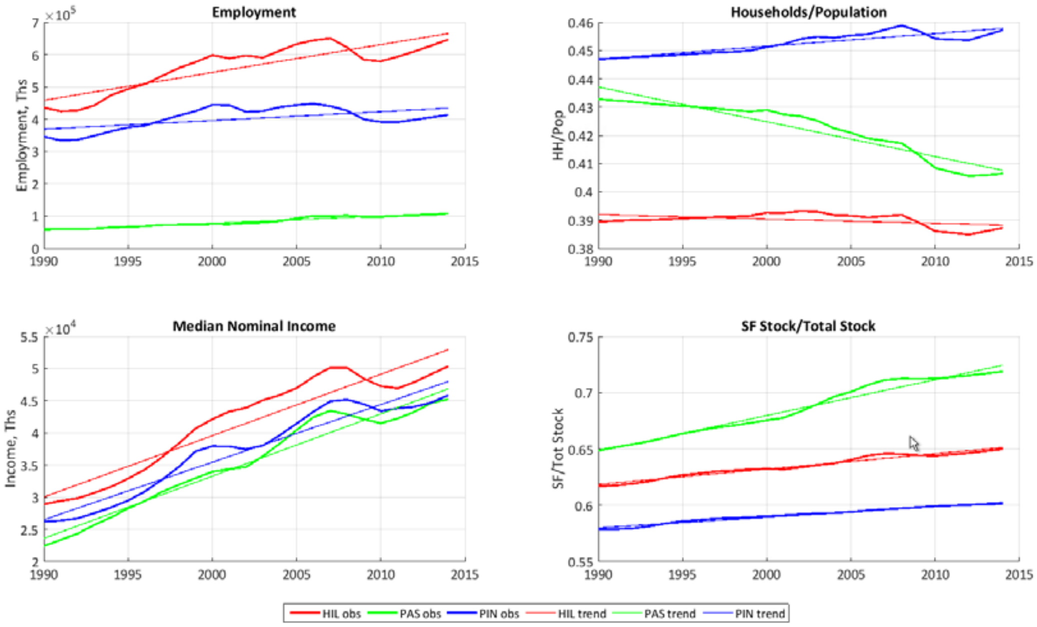
- Population
  - Provided by Bureau of Economic and Business Research (BEBR): U. of Fla.
  - Use medium growth rate to point-project units and ksf
  - High/Medium/Low projections: “75% of outcomes”
  - Use all three to define population distributions, then corresponding units and ksf distributions
  - Sample during Monte Carlo



Uncertainty Assumption Examples

Historical Short-Term Variability Imposed on Point Projections

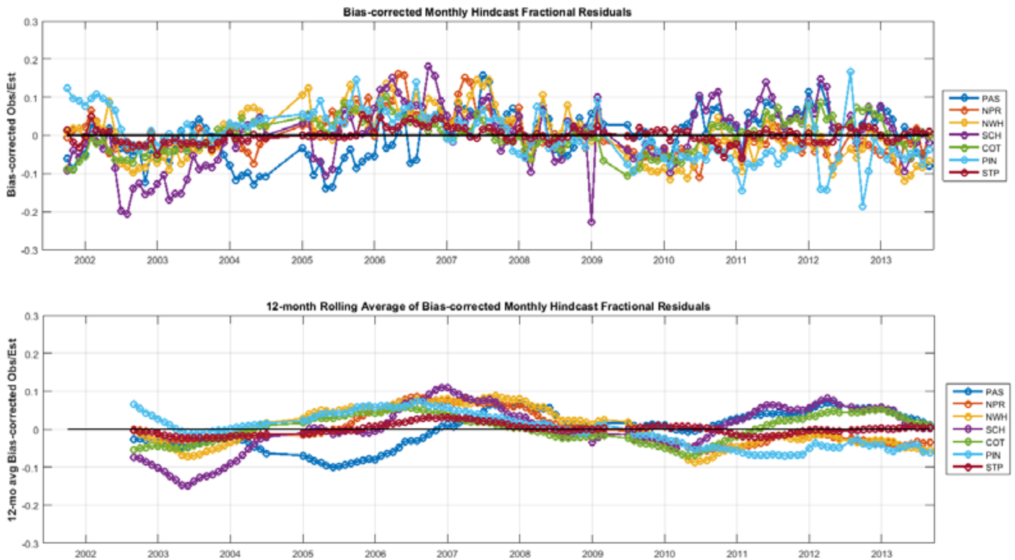
- Explanatory variables
  - Historical data: trend and annual departures
  - Pretend point projection is the future trend
  - Add historical short-term variations to future – sample during Monte Carlo



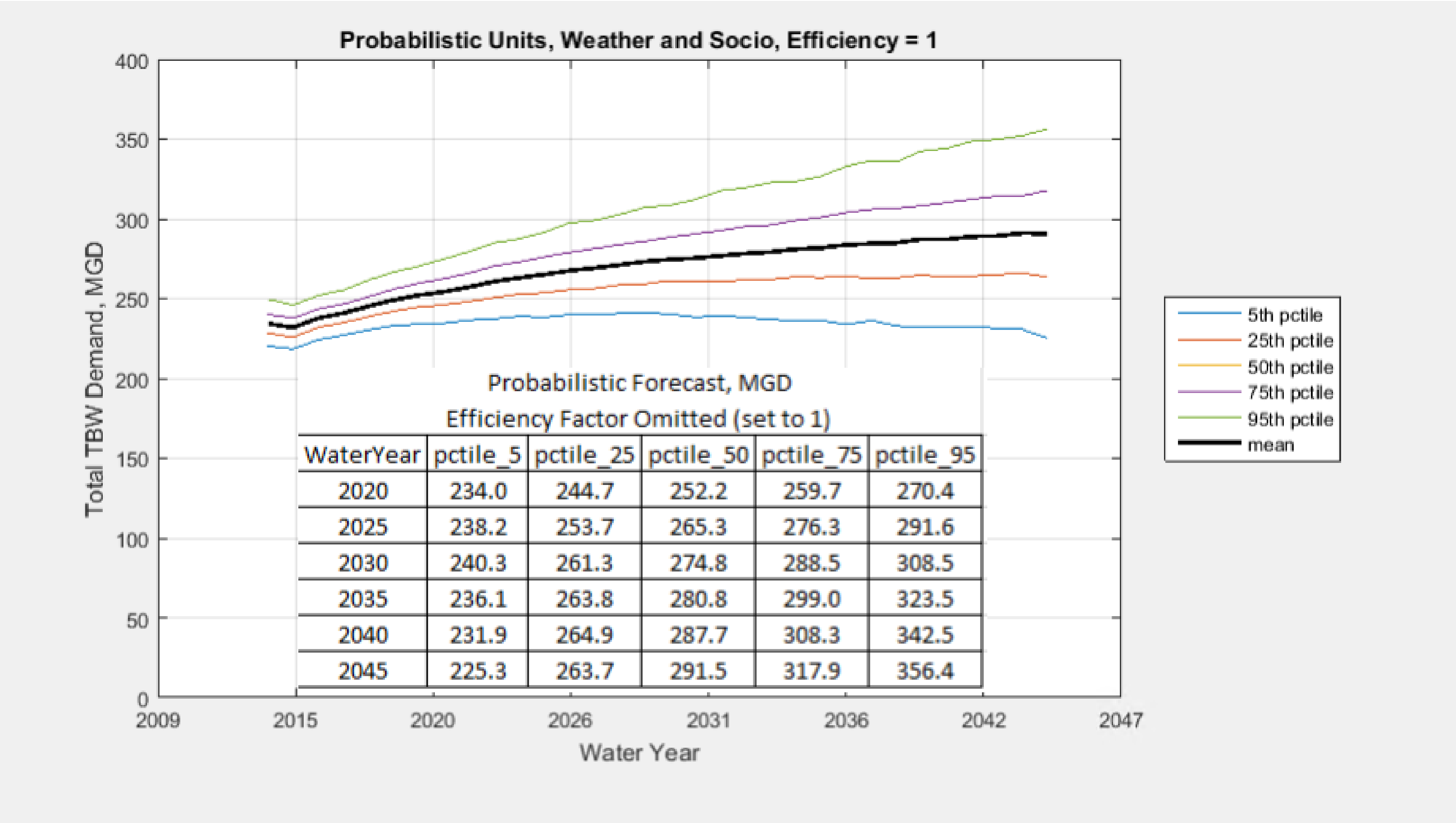
Uncertainty Assumption Examples

Model Noise

- Fitting set observations and estimates: residuals
  - Reflects parameter and specification uncertainty
  - Sample residuals during Monte Carlo



Probabilistic Forecast Ignoring Efficiency Factor (set to 1)

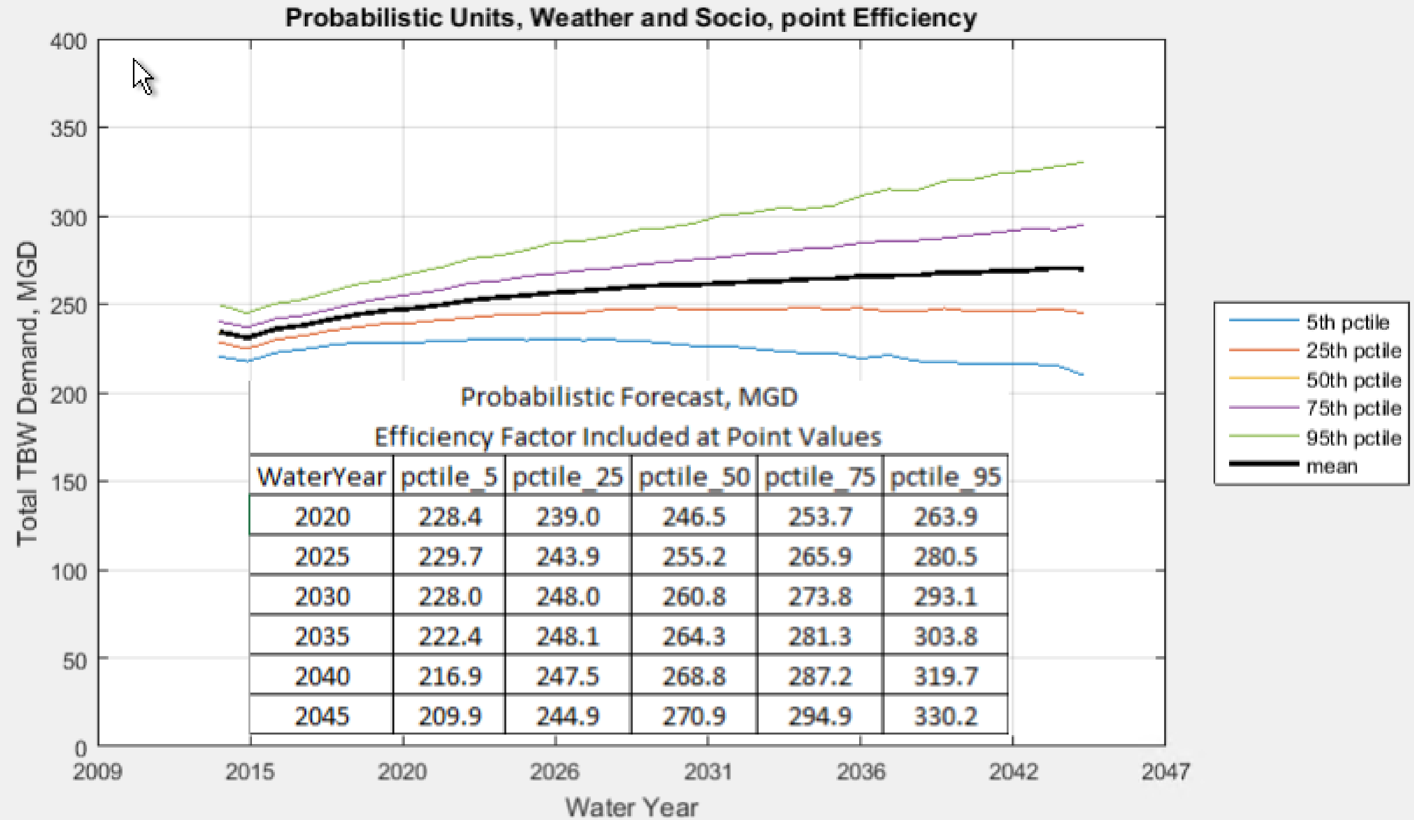


Probabilistic Forecast, MGD

Efficiency Factor Omitted (set to 1)

WaterYear	pctile_5	pctile_25	pctile_50	pctile_75	pctile_95
2020	234.0	244.7	252.2	259.7	270.4
2025	238.2	253.7	265.3	276.3	291.6
2030	240.3	261.3	274.8	288.5	308.5
2035	236.1	263.8	280.8	299.0	323.5
2040	231.9	264.9	287.7	308.3	342.5
2045	225.3	263.7	291.5	317.9	356.4

Probabilistic Forecast With Point-Projected Efficiency Factor



Probabilistic Forecast, MGD

Efficiency Factor Included at Point Values

WaterYear	pctile_5	pctile_25	pctile_50	pctile_75	pctile_95
2020	228.4	239.0	246.5	253.7	263.9
2025	229.7	243.9	255.2	265.9	280.5
2030	228.0	248.0	260.8	273.8	293.1
2035	222.4	248.1	264.3	281.3	303.8
2040	216.9	247.5	268.8	287.2	319.7
2045	209.9	244.9	270.9	294.9	330.2

Impacts of Including Efficiency in Probabilistic Forecast

Entire forecast discernably lowered

- Median
  - 5.7 MGD (2.3%) lower in 2020
  - 20.7 MGD (7.1%) lower in 2045
- Stands out from forecast noise
  - Inclusion of efficiency moves the prior median demands to the 75th percentile

Probabilistic Forecast, MGD
Efficiency Factor Omitted (set to 1)

WaterYear	pctile_5	pctile_25	pctile_50	pctile_75	pctile_95
2020	234.0	244.7	252.2	259.7	270.4
2025	238.2	253.7	265.3	276.3	291.6
2030	240.3	261.3	274.8	288.5	308.5
2035	236.1	263.8	280.8	299.0	323.5
2040	231.9	264.9	[illegible]	308.3	342.5
2045	225.3	263.7	291.5	317.9	356.4

Probabilistic Forecast, MGD
Efficiency Factor Included at Point Values

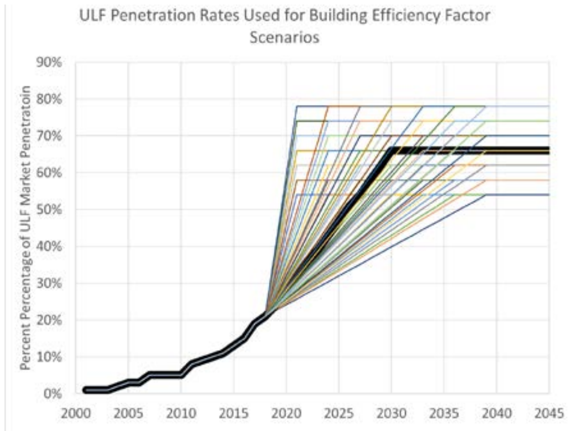
WaterYear	pctile_5	pctile_25	pctile_50	pctile_75	pctile_95
2020	228.4	239.0	246.5	253.7	263.9
2025	229.7	243.9	255.2	265.9	280.5
2030	228.0	248.0	260.8	273.8	293.1
2035	222.4	248.1	264.3	281.3	303.8
2040	216.9	247.5	268.8	287.2	319.7
2045	209.9	244.9	270.9	294.9	330.2

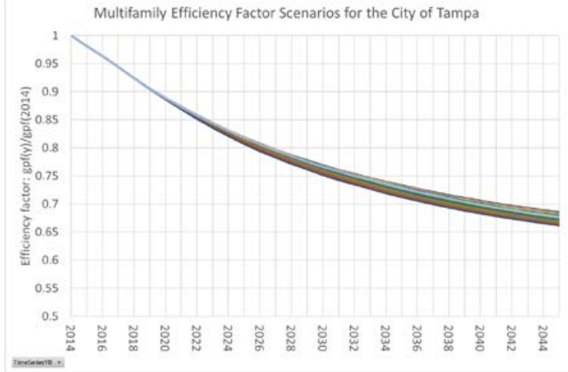
Conclusions

- Efficiency factor approach
  - Less data intensive than a full end use model
  - Focus on fixtures where data and assumptions are easiest to come by
  - Impacts of future efficiency on forecast are reasonable
- Even when considering the uncertainty inherent in a forecast, end use efficiency increases can produce significant demand impacts

Future Work

- Uncertainty in future end-use efficiency
  - Already began this as part of probabilistic forecast
  - Assumed wide range of future 1.6/1.28 efficiency standards
  - Produced ranges of efficiency factors
  - Not much additional uncertainty
- What about uncertainty in average lifetime?
  - Could be large contributor
  - Need statistical definitions (confidence intervals)



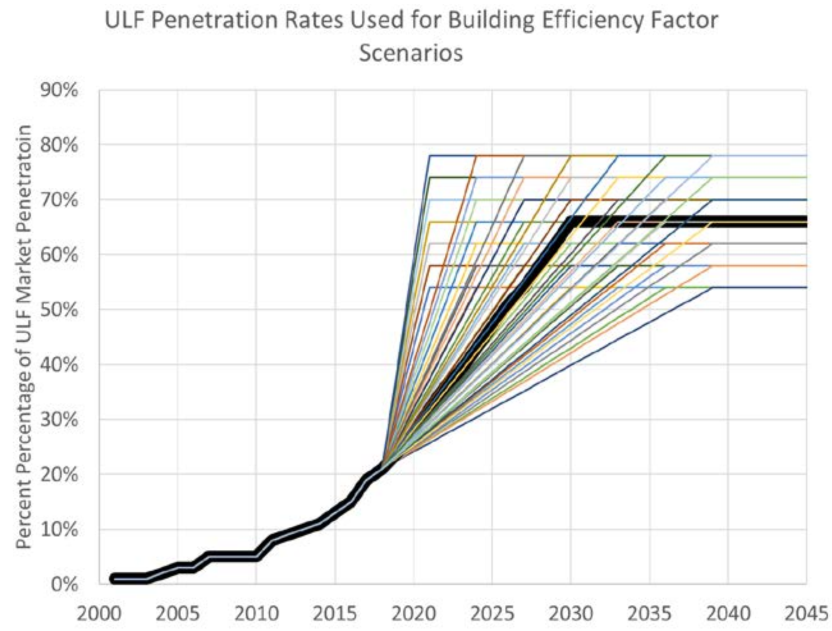


Questions

Efficiency Factor Uncertainty

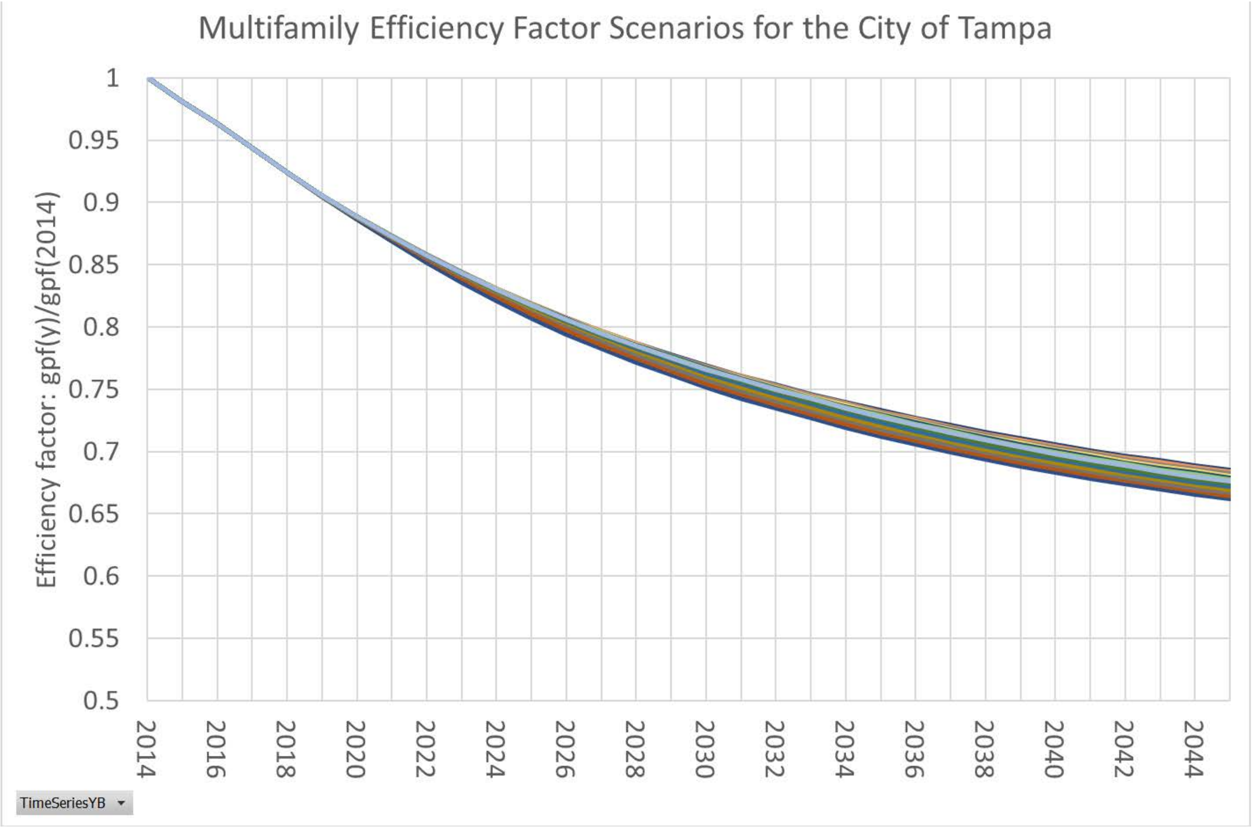
Possibly the First Ever Probabilistic Treatment of Efficiency in Forecasting

- Parameters of stock model
  - Average toilet lifetime
  - Future 1.28 market presentation
- Uncertainty: 49 scenarios of penetration rates
  - Fastest: 78% 1.28 gpf by 2021
  - Slowest: 52% 1.28 gpf by 2039
- Did not vary average lifetime
  - Projection sensitivity
  - Need for well-supported bounds

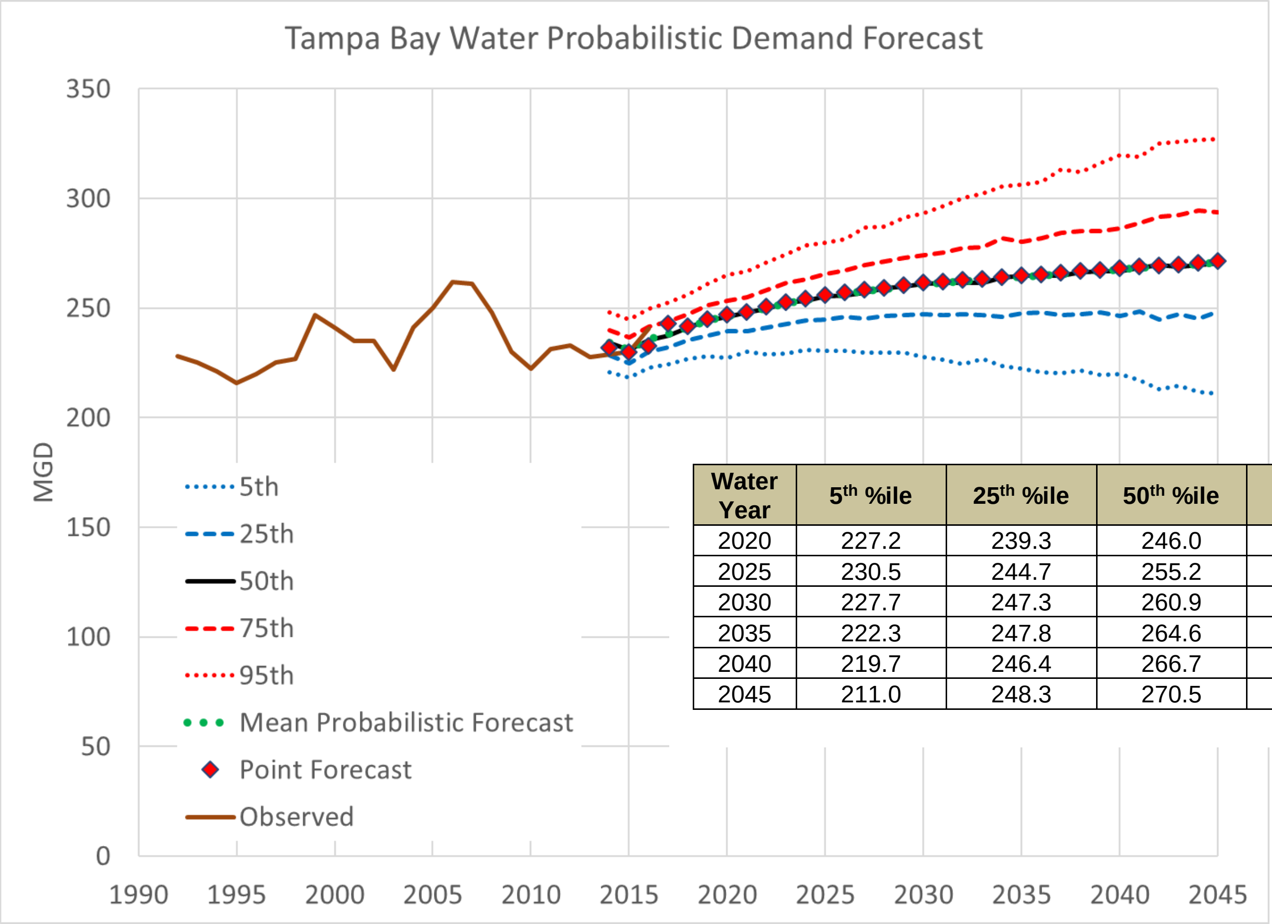


- Apply penetration scenarios to stock model: corresponding factors

Efficiency Factor Uncertainty



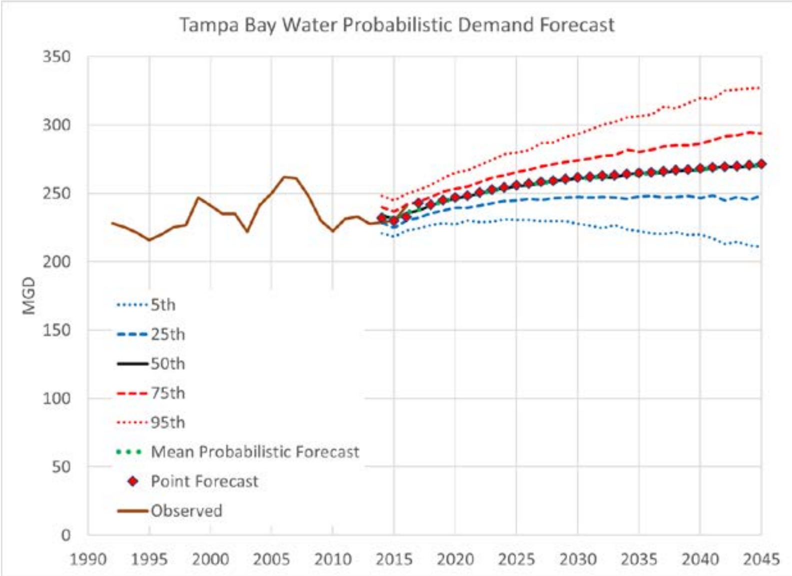
Probabilistic Forecast (2014-2016 Base Period)



Water Year	5th %ile	25th %ile	50th %ile	75th %ile	95th %ile	Mean	Point Forecast
2020	227.2	239.3	246.0	253.4	265.0	246.4	247.0
2025	230.5	244.7	255.2	265.7	280.0	255.2	255.8
2030	227.7	247.3	260.9	274.2	293.4	260.9	261.5
2035	222.3	247.8	264.6	280.2	306.3	264.3	264.9
2040	219.7	246.4	266.7	286.2	319.8	267.6	268.3
2045	211.0	248.3	270.5	293.8	327.0	270.7	271.3

Largest Contributors to Uncertainty

- Population
  - 1:1 demand variation
- Balance between price and income
  - both strong controls on intensity
  - opposite directions of influence
- Intensity model uncertainty
- Details of how variables are correlated

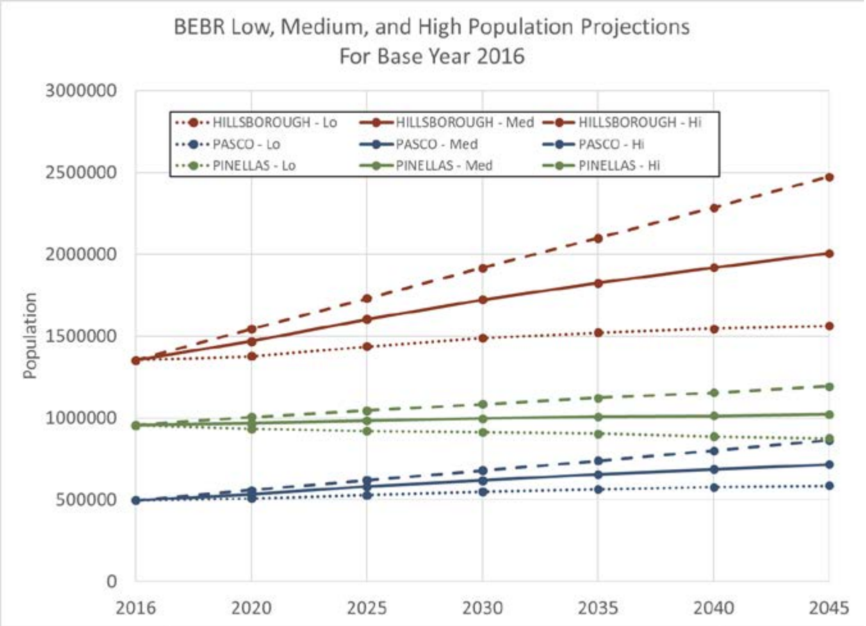


Summary

- New Tampa Bay Water probabilistic approach
  - Incorporate uncertainty in drivers, explanatory variables, intensity model
  - Uncertainty leverages a variety of information
  - Novel uncertainty application for future efficiency
- Different uncertainty sources have different degrees of influence
  - Uncertainty treatment can be selective – just where forecasts are most sensitive
- Additional work
  - Uncertainty treatment for more variables
  - Efficiency: expression of uncertainty for average fixture lifetime

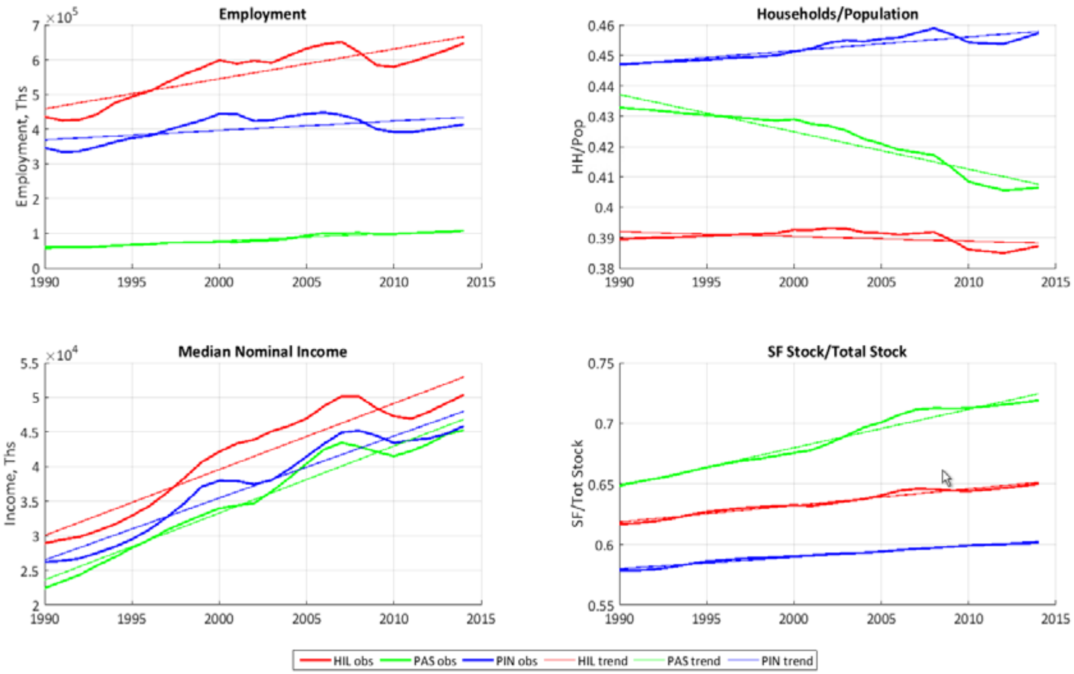
Driver Uncertainty

- Population Projections from Bureau of Economic and Business Research (BEBR) at University of Florida
  - County-level
  - High, Medium, and Low Scenarios
  - Probabilistic: High-to-low represents 75% prediction interval
- Translate population distributions onto forecast drivers



Explanatory Uncertainty

- Historical, projected socioeconomic data from Moody's Analytics
    - Historical data show year-to-year variations around trends
    - Statistically describe variations, center them onto point projections



Model Uncertainty

- Plug historical driver and explanatory data into model
- Compare with historical demands
- Statistically describe residuals