Modeling Efforts

Gavin Corral

USDA NASS

May 15th 2019

Disclaimer

The findings and conclusions in this presentation are those of the author(s) and should not be construed to represent any official USDA or U.S. Government determination or policy.

Goals

1. Identify the fundamental elements of a hog inventory model
2. Analyze the strengths and weaknesses of the two NASS hog models

Outline

- Introduction
  - Fundamentals for modeling hog inventory
  - Model evaluation
- Kalman Filter Model
  - Performance
- Sequential Generalized Linear Model
  - Performance
- Model Comparison
- Shock Diagnostics

Fundamentals of Modeling Hog Inventory

- Purpose:
  - Produce estimates
  - Respect interrelationships (constraints)
  - Efficient
- Simply compiling survey results fails
  - Downward bias
  - Does not reflect hog growth cycle

Model Evaluation

- Overall evaluation
  - Compare ME to official estimate issued by hog board (1 day slaughter ~ 470k)
  - Compare ME to final revised estimate (1 year)

- Differences in official estimate/final estimate
  - Equilibrium vs. shock

Model Evaluation

- Satisfy Constraints
  - Biological considerations (Lifecycle)
    - Births -> growth-> death… and everything in-between
    - ME should be intrinsically connected time/weight gain
  - Slaughter
- Shock
    - Disease: PRRS & PEDv
    - Reaction: before/after

Criteria for Model Evaluation

- Model Evaluation:
  - Captures equilibrium picture
  - Detects and adjusts for shocks when they appear
  - Hog lifecycle (biological considerations)
  - Satisfies external accounting constraints (balance sheet)

Model Performance

- All models were run for board (second of the two runs)
- Each estimate is from its own run
  - December 2015 estimate was produced as if we were running the model for board that quarter
- Epidemic years 2013-2015

KFM

- (KFM) Time series approach with
  - Kalman Filter
    - Update the state of the system after input of observation
  - State space representation. Expressed through
    - Transition equation – how hog inventories change over time
    - Observation equation – relate the state of the system with a set of measurements from that state to another point in time
    - Determine new state of system at future point in time given the state of the current system

KFM

- Constraints
  - Death loss ratio
    - Quantity of pig crop that dies after weaning and cannot be counted in the market weight groups
    - Annual increase of born and weaned > annual increase in market weight groups
  - Weight group transition
    - Assumption about the growth of pigs within weight classes
    - Maps births and their weight groups-> heavier weight groups

KFM

- Constraints (cont.)
  - Pig crop- slaughter ratio
    - Annual increase in slaughter is equal to annual increase in births for the two preceding quarters
  - Market hogs-slaughter ratio
    - The total number of market hogs in a quarter should equal the combined total slaughter numbers for the next two quarters
    - 6 month time period WG1 to slaughter

KFM

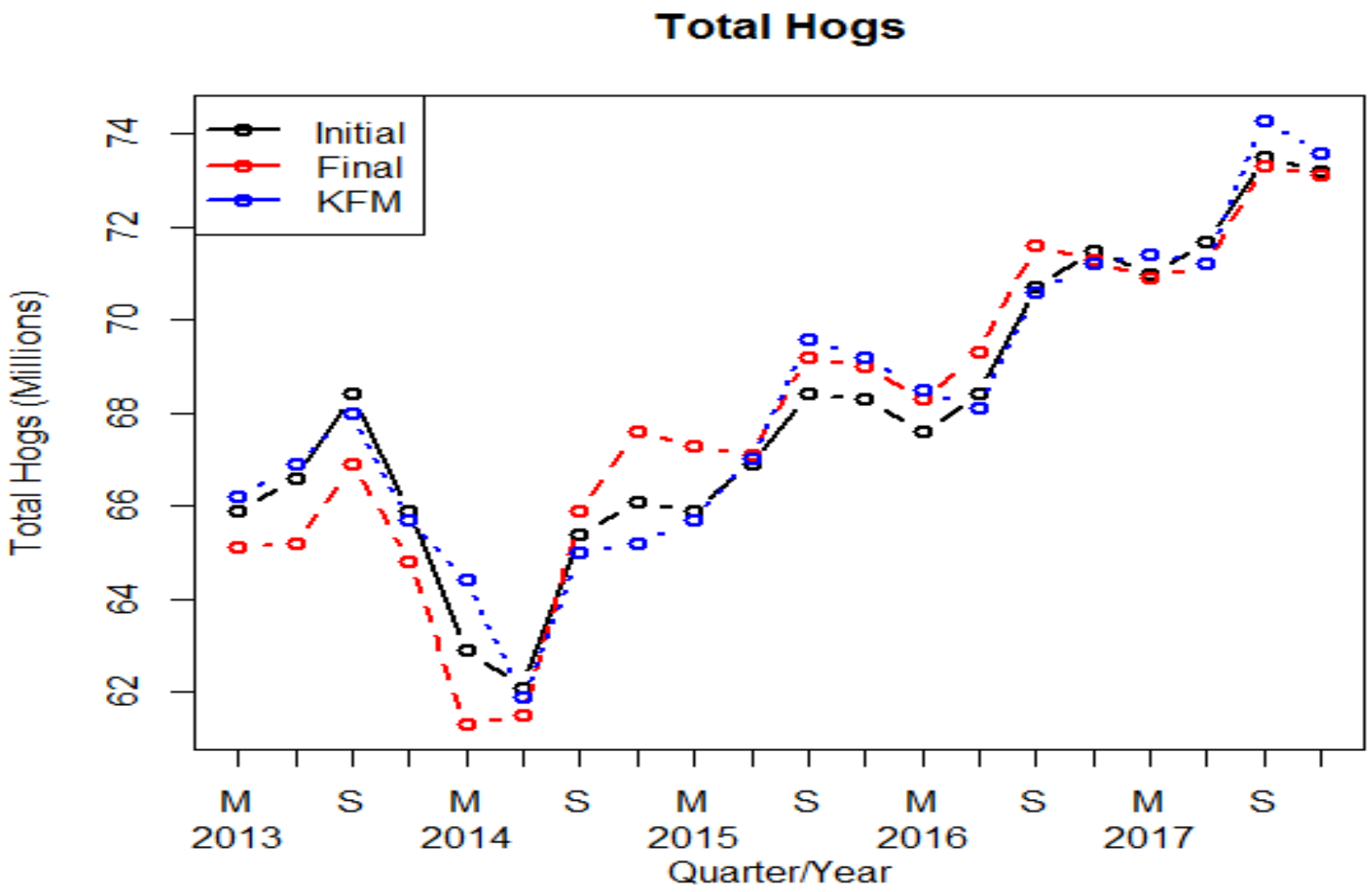
- Constraints (cont.)
  - Market hogs over 180lbs-slaughter ratio
    - Relates mh >180 to slaughter occurring during estimation quarter, following the reference quarter
    - Although quarter is in progress, daily slaughter information is available
    - At time of board there are 2 full weeks of daily slaughter available for the 180+ wg
  - Sows farrowed and breeding herd
    - Sows farrowed make up half of previous quarter breeding herd

KFM

- Additionally:
  - Constant survival rate across all estimated weight classes
  - Previous 5 quarters and current quarter data used to capture
    - Cycle dynamics
    - Annual trend

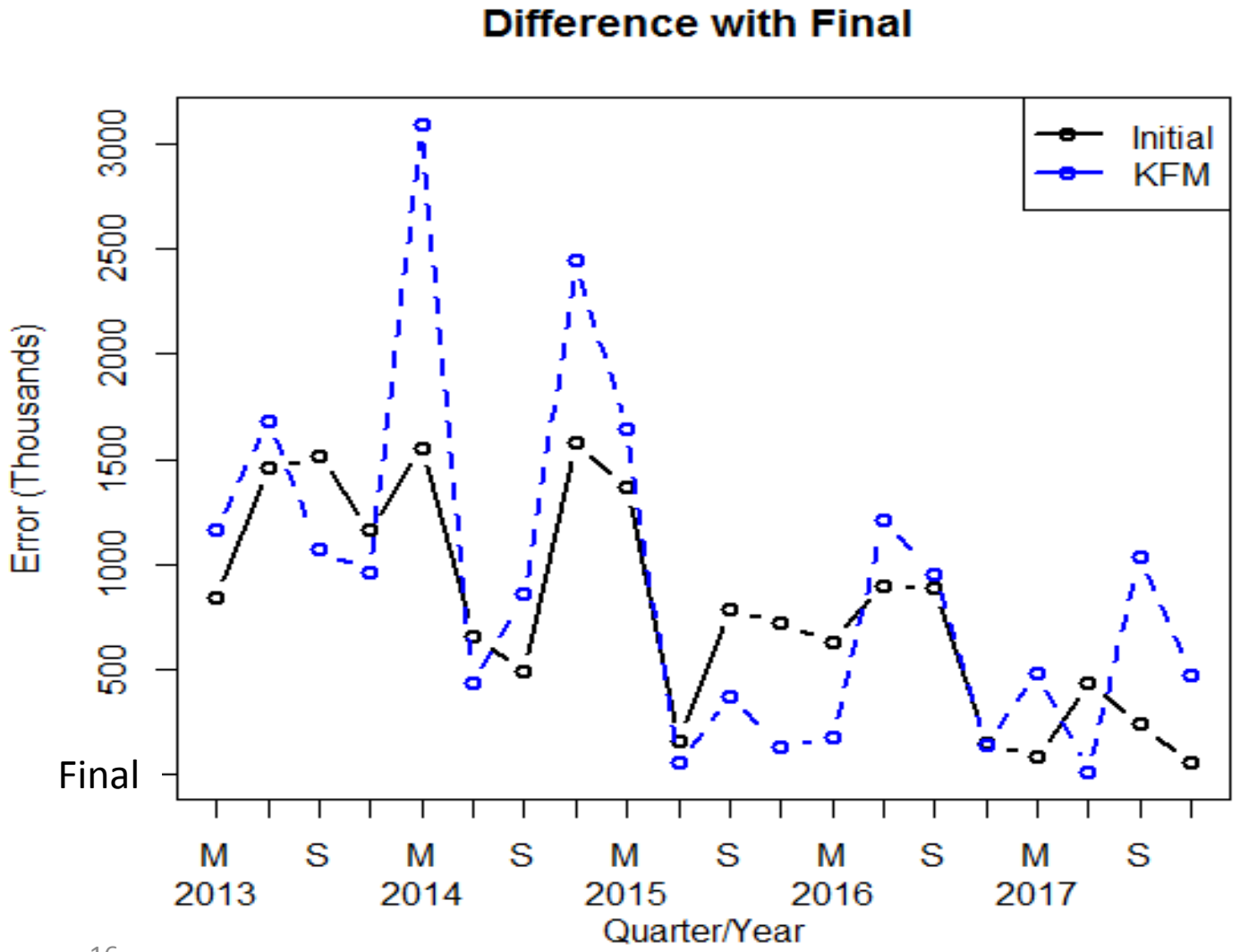
KFM Performance

- Shocks
- Board adjustments
- Model during/after



KFM Performance

- Downward slope of initial and hill pattern of KFM
- Reaction to shock



KFM Performance

- Shocks
- Board adjustments
- Model during/after
- KFM Criteria
  - Captures equilibrium picture
  - Detects/ Adjust shock
  - Accounts hog lifecycle / Allows disruption
  - Satisfies external accounting relationship

SGLM

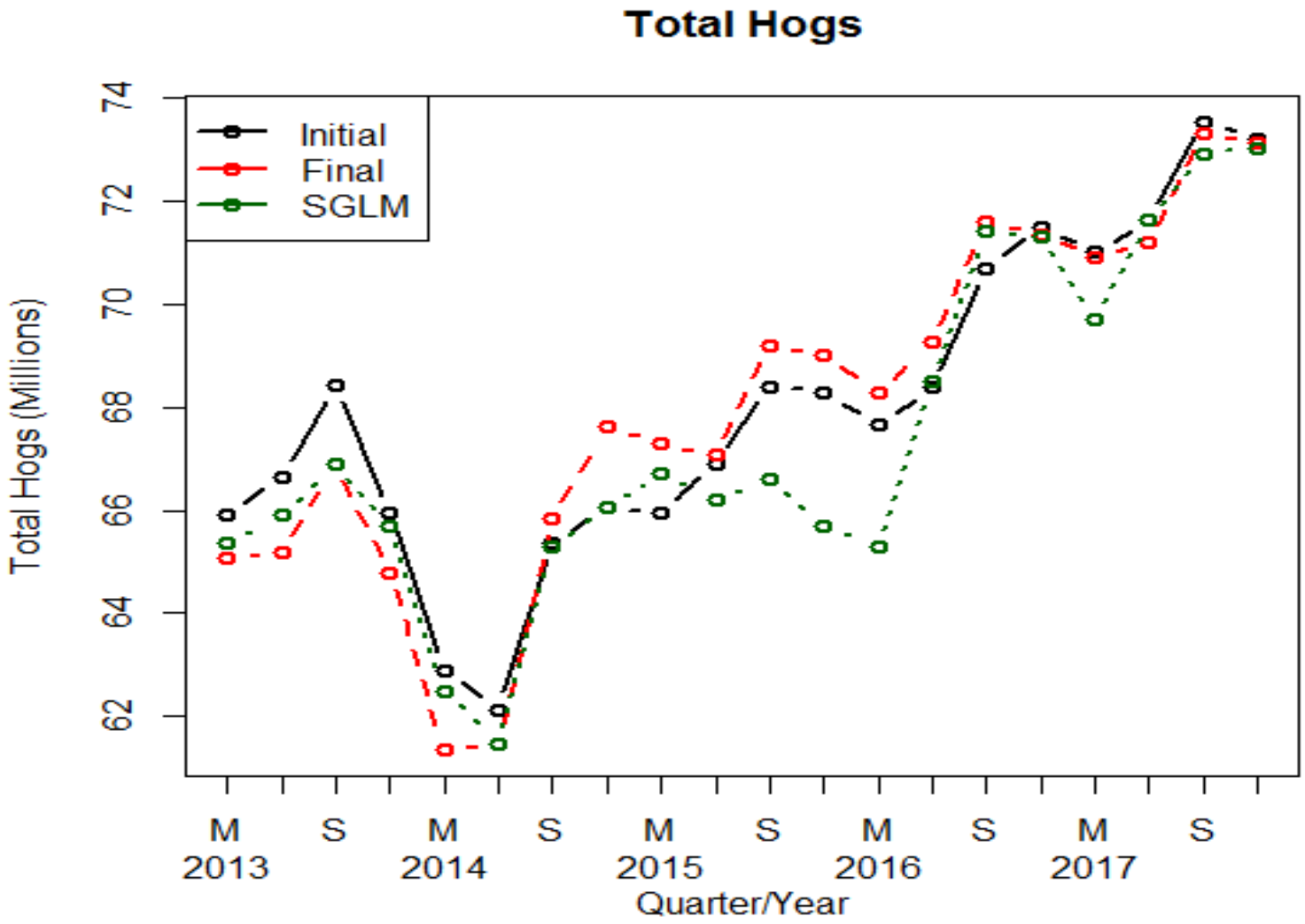
- Developed by Kedem & Pan 2015, attempt to address shocks
- Choice of SGLM based on:
  - More weight to current immediate data/better capture changing dynamics
  - Enable a dynamic selection across a wide range of potential covariates
    - Economic / survey results/board estimates
  - Flexible

SGLM

- SGLM works by
  - Testing large numbers of potential covariates using spectral analysis and selecting among them for the final model
  - 4 year window
  - 4-8 covariates
  - Independent prediction of each inventory item
- Easy/fast implementation via *Shiny*

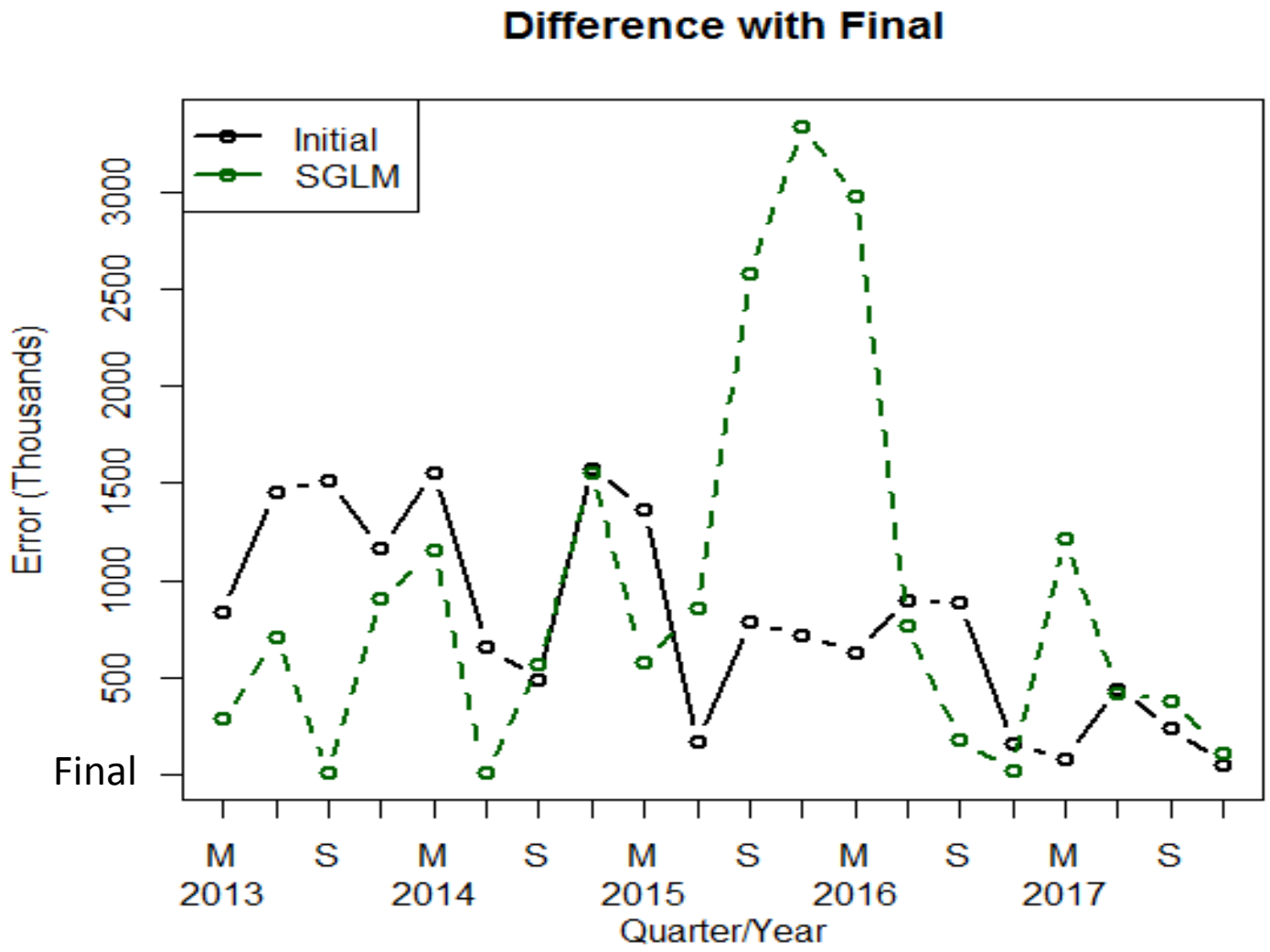
SGLM Performance

- SGLM model estimates could not consistently match measure of accuracy
- Difficulties post-shock period



SGLM Performance

- General decreasing trend vs. large hill
- Post shock spike



SGLM Performance

- Inability of the model to adhere to the biological aspects of the hog life cycle
- Total hogs from the SGLM does not equal the sum of its parts
- Only a snapshot
- No WGT constraint

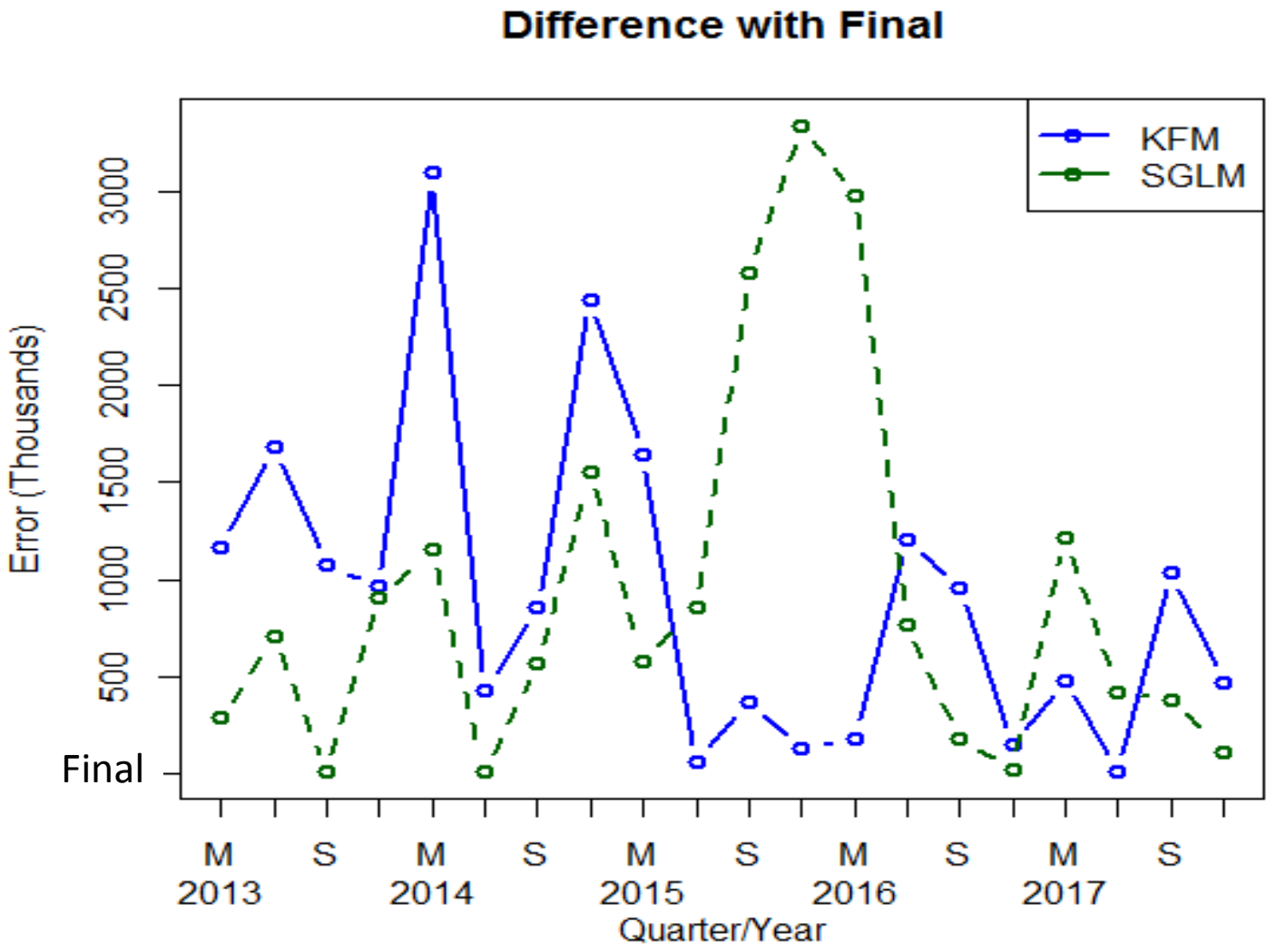
		Model	
		SGLM	KFM
Estimate	G12	38943.00	39657.45
	G3	13324.15	13705.34
	G4	11258.00	11291.74
	BH	6011.72	5976.95
	Total Hogs	70656.00	70631.48
	Sum(G12,G3,G4,BH)	69536.87	70631.48

SGLM

- Shocks
- Model during/after
- SGLM Criteria
  - Captures equilibrium picture 
  - Detects  / Adjust shock 
  - Accounts hog lifecycle  /Allows Disruption 
  - Satisfies external accounting relationship 

Model Comparison

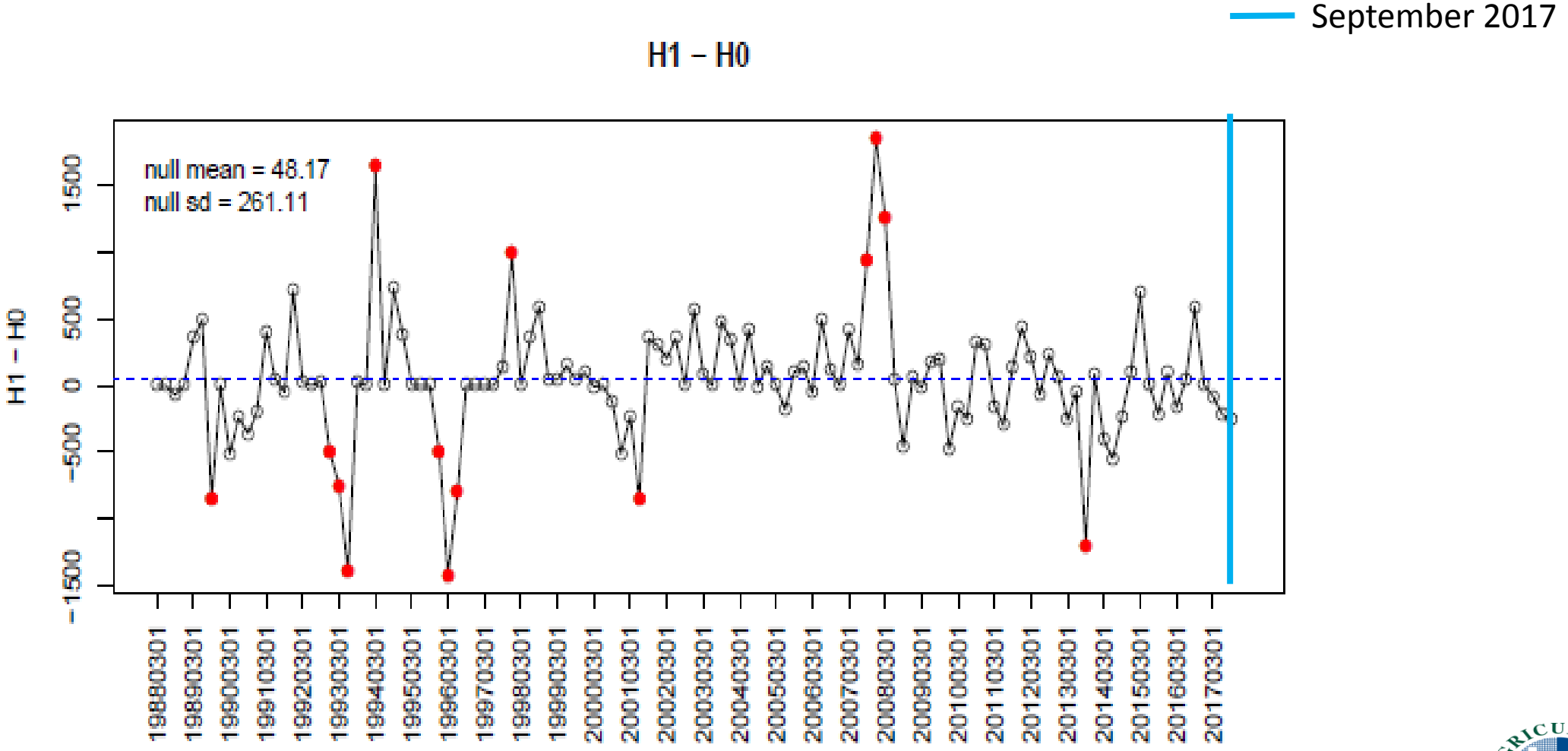
- KFM
  - Consideration biological
  - Satisfies the accounting constraints
  - Unable to adapt quickly to systematic shocks resulting in biased and unrealistic results
- SGLM
  - Captures the economic patterns and departures from an equilibrium state
  - Does not satisfy reasonable biological dynamics of the hog population.



Diagnostics

- Wang et al. (2016) developed a Bayesian model to detect shocks for NASS
- Use a variety of variable
  - Sows farrowed, Pig crop ratios, differences in revisions
- Multiple-hypothesis testing
  - Large scale (temporal) dependent data
  - The dependence structure among hypotheses governed by a hidden Markov model (HMM)
- Uses a Dirichlet mixture model with an unknown number of distributions for the non-null hypothesis.
- Algorithm allows for an optimal false negative rate, while controlling the false discovery rate (Wang et al. 2015)
- No real time detection

Diagnostics



Diagnostics

Date	SR	H0	H1	H2	H3	H4	HP
20130301	74700	75078	73253	56819	60515	58474	58474
20130601	67007	75312	64733	60235	56057	55279	55279
20130901	59122	70914	62156	65304	67074	62057	62057
20131201	61457	75392	59617	66394	62236	73306	73306
20140301	57121	71844	56731	60734	57205	70064	70064
20140601	66148	58857	54278	70330	66381	59681	59681
20140901	74773	64294	58588	75033	60110	57915	57915
20141201	58600	68589	67388	77078	66163	57858	57858
20150301	59097	63485	66291	70916	67259	73175	73175
20150601	72927	69124	68924	67228	75526	66929	66929
20150901	68547	60425	65737	67792	76866	59486	59486
20151201	63799	61448	67705	75617	68210	62416	62416
20160301	71893	75469	61455	63622	68925	60868	60868
20160601	69956	64764	70349	75421	60690	79026	79026
20160901	64959	69083	73250	63205	78297	77828	77828
20161201	65923	78765	77037	81097	71480	80459	80459
20170301	71912	77384	75584	69829	78106	NA	78106
20170601	70650	72341	64368	63316	NA	NA	63316
20170901	62393	68511	69934	NA	NA	NA	69934
20171201	68644	78187	NA	NA	NA	NA	78187

Conclusion

- KFM model is the most useful tool for NASS
  - Although the KFM model has some shortcomings
    - namely the inability to provide reliable estimates during shock periods
    - reliable in periods emerging from shocks, when the SGLM model usually fails.
- Diagnostics are useful and provide needed information but has a time lag

Questions?