Back to the Future

Do recent developments tell us anything about the future of statistics?

Jon A. Wellner

University of Washington

- Talk at Joint Statistical Meetings
- Washington, D.C.
- August 3, 2009
- *Email: jaw@stat.washington.edu*
  http: //www.stat.washington.edu/jaw/jaw.research.html

Outline

- Background

Outline

- Background
- What recent developments?

Outline

- Background
- What recent developments?
- Possible future directions?

Outline

- Background
- What recent developments?
- Possible future directions?
- A few specific problems.

1. Background

Prediction is very difficult, especially when it involves the future.
(Niels Bohr).

Some Boundaries and Restrictions:

- Focus here on:
  - Statistical inference (theory and methods)
  - Research problems

Some Boundaries and Restrictions:

- Focus here on:
  - Statistical inference (theory and methods)
  - Research problems
- Other recent surveys, reviews, and reports:
  - *Report on the Future of Statistics* (Stat. Sci. 2004); Lindsay, Kettenring, and Siegmund.
  - Longer on-line version of above: *Statistics: Challenges and Opportunities for the Twenty-First Century*.
  - *Statistics in the 21st Century*; Raftery, Tanner, and Wells. 22 vignettes (organized by George Casella) on “Theory and Methods”

Some Boundaries and Restrictions:

- Focus here on:
  - Statistical inference (theory and methods)
  - Research problems
- Other recent surveys, reviews, and reports:
  - *Report on the Future of Statistics* (Stat. Sci. 2004); Lindsay, Kettenring, and Siegmund.
  - Longer on-line version of above: *Statistics: Challenges and Opportunities for the Twenty-First Century*.
  - *Statistics in the 21st Century*; Raftery, Tanner, and Wells. 22 vignettes (organized by George Casella) on "Theory and Methods"
- Personal, subjective, and biased view, with lots ignored mostly due to personal ignorance.

Problems for the future (of Statistical Inference) mentioned by Kiefer, Savage, and Le Cam, 1967;
Conference at University of Wisconsin on "The Future of Statistics"

- Problems mentioned by Kiefer:
  - Theory of nonparametric inference: testing and estimation.
  - Theory of nonparametric Bayes procedures. (Theory developing over last 10+ years: e.g. Le Cam lecture by van der Vaart at this meeting.)
  - Rates of convergence. (Götze, Bickel, van Zwet, Peter Hall, ...)
  - Nonparametric regression / curve estimation. (Hints of model selection, penalized estimation.)

- Problems, Le Cam:
  - Alternatives to Bayesian statistics?
  - Specification of stochastic structure?
  - Stability of experiments? (Distance between experiments? Discrete versus continuous parameterizations?)
  - Stability and relations to invariance?

- Problems, Le Cam:
  - Alternatives to Bayesian statistics?
  - Specification of stochastic structure?
  - Stability of experiments? (Distance between experiments? Discrete versus continuous parameterizations?)
  - Stability and relations to invariance?
- Problems mentioned by Savage:
  - Descriptive statistics (look at the data).
  - Multi-parameter problems and nonparametric problems
  - Weather forecasting?
  - Medical diagnosis?

- Problems mentioned by Audience:

- Problems mentioned by Audience:
  - Colin Mallows:

- Problems mentioned by Audience:
  - Colin Mallows:
    - Effect of computing facilities on statistical inference?

- Problems mentioned by Audience:
  - Colin Mallows:
    - Effect of computing facilities on statistical inference?
    - Subset regression?

- Problems mentioned by Audience:
  - Colin Mallows:
    - Effect of computing facilities on statistical inference?
    - Subset regression?
    - What should we be teaching in inference?

- Problems mentioned by Audience:
  - Colin Mallows:
    - Effect of computing facilities on statistical inference?
    - Subset regression?
    - What should we be teaching in inference?
  - George H. Ball (Stanford Research Institute).

- Problems mentioned by Audience:
  - Colin Mallows:
    - Effect of computing facilities on statistical inference?
    - Subset regression?
    - What should we be teaching in inference?
  - George H. Ball (Stanford Research Institute).
    - “... select some subset of the variables if we have more than five or six ... ”

- Problems mentioned by Audience:
  - Colin Mallows:
    - Effect of computing facilities on statistical inference?
    - Subset regression?
    - What should we be teaching in inference?
  - George H. Ball (Stanford Research Institute).
    - "... select some subset of the variables if we have more than five or six ... "
    - Change of perspective with computing power?

2. What recent developments?

The future ain't what it used to be. (Yogi Berra)

Data: Numbers of papers per year for several topics.

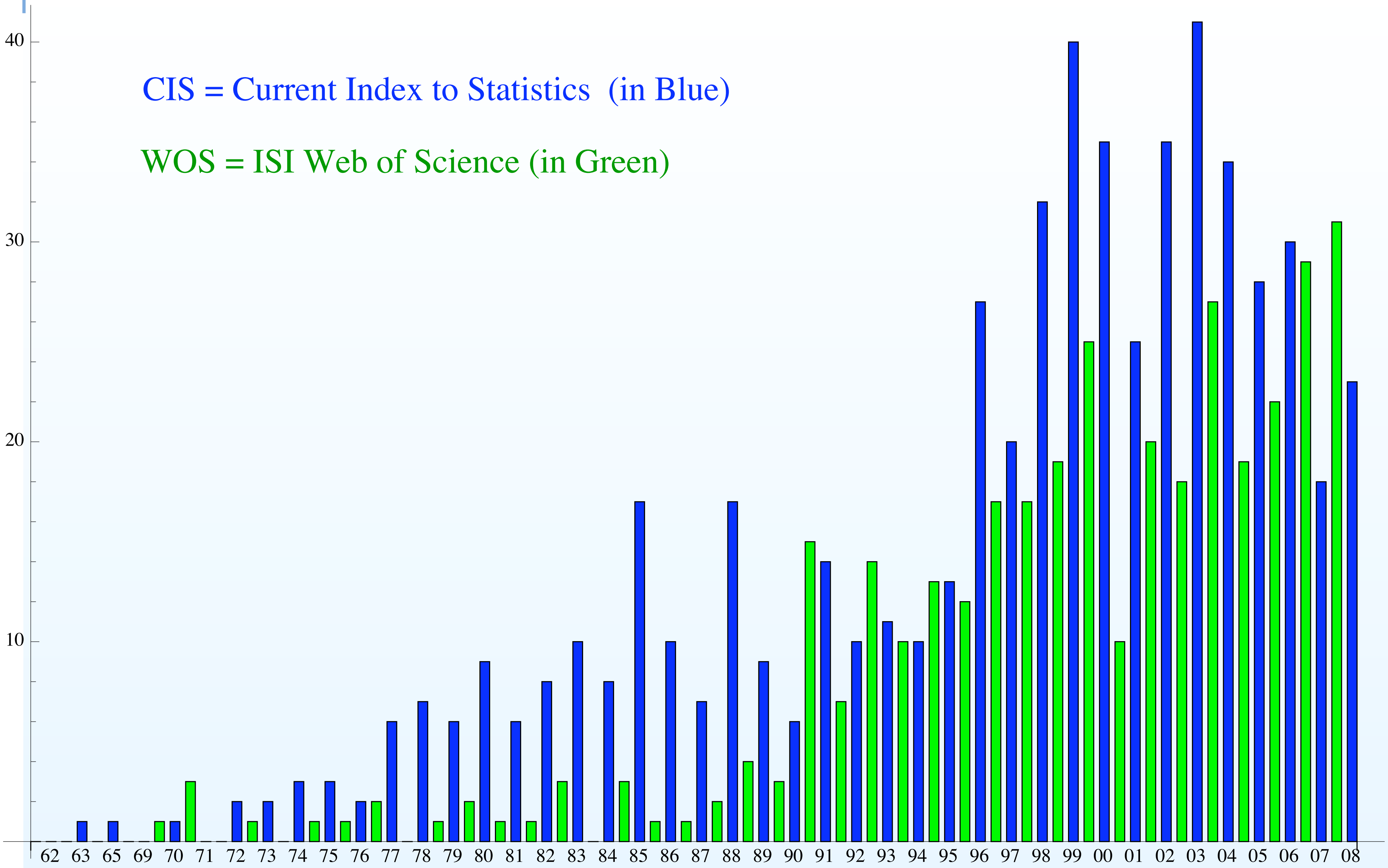
- Topics explicitly mentioned at the 1967 Madison meeting
  - Nonparametric Bayes methods (Kiefer)
  - Robustness (Kiefer)
  - Subset regression (Mallows, Ball) $\longrightarrow$ model selection

Data: Numbers of papers per year for several topics.

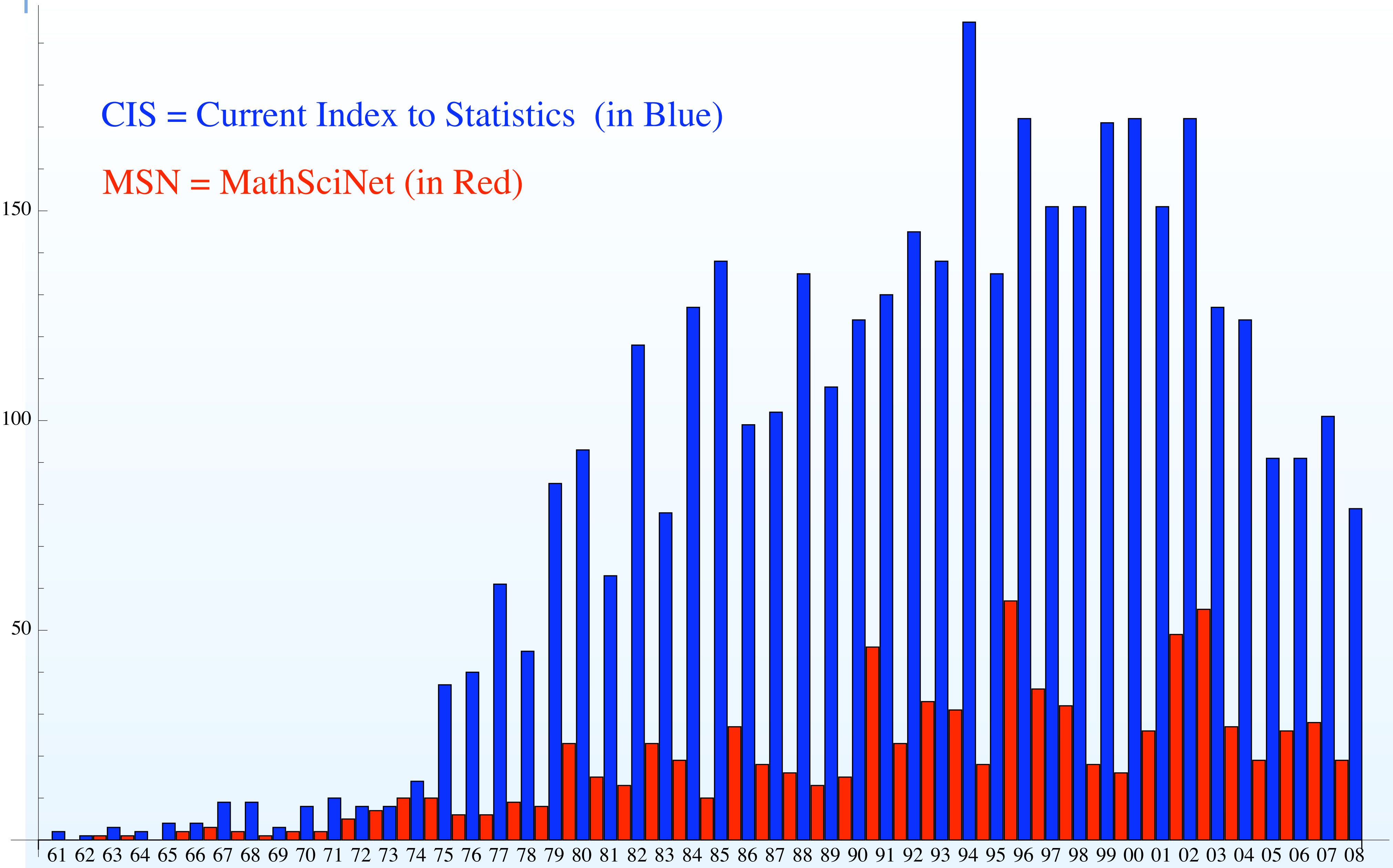
- Topics explicitly mentioned at the 1967 Madison meeting
  - Nonparametric Bayes methods (Kiefer)
  - Robustness (Kiefer)
  - Subset regression (Mallows, Ball) $\longrightarrow$ model selection
- Topics implicit or “hinted at” in discussion at the 1967 Madision meeting
  - Alternatives to Bayes methods (Le Cam) $\longrightarrow$ empirical Bayes estimation
  - Structure of stochastic models (Le Cam) $\longrightarrow$ graphical models
  - Multiple comparisons / multiple testing (Savage) $\longrightarrow$ false discovery rate
  - Multiparameter - nonparametric models (Savage) $\longrightarrow$ semiparametric models

- Topics not mentioned at the 1967 UW meeting
  - Bootstrap methods
  - Markov chain monte carlo
  - Empirical process theory / methods
  - Lasso + regression

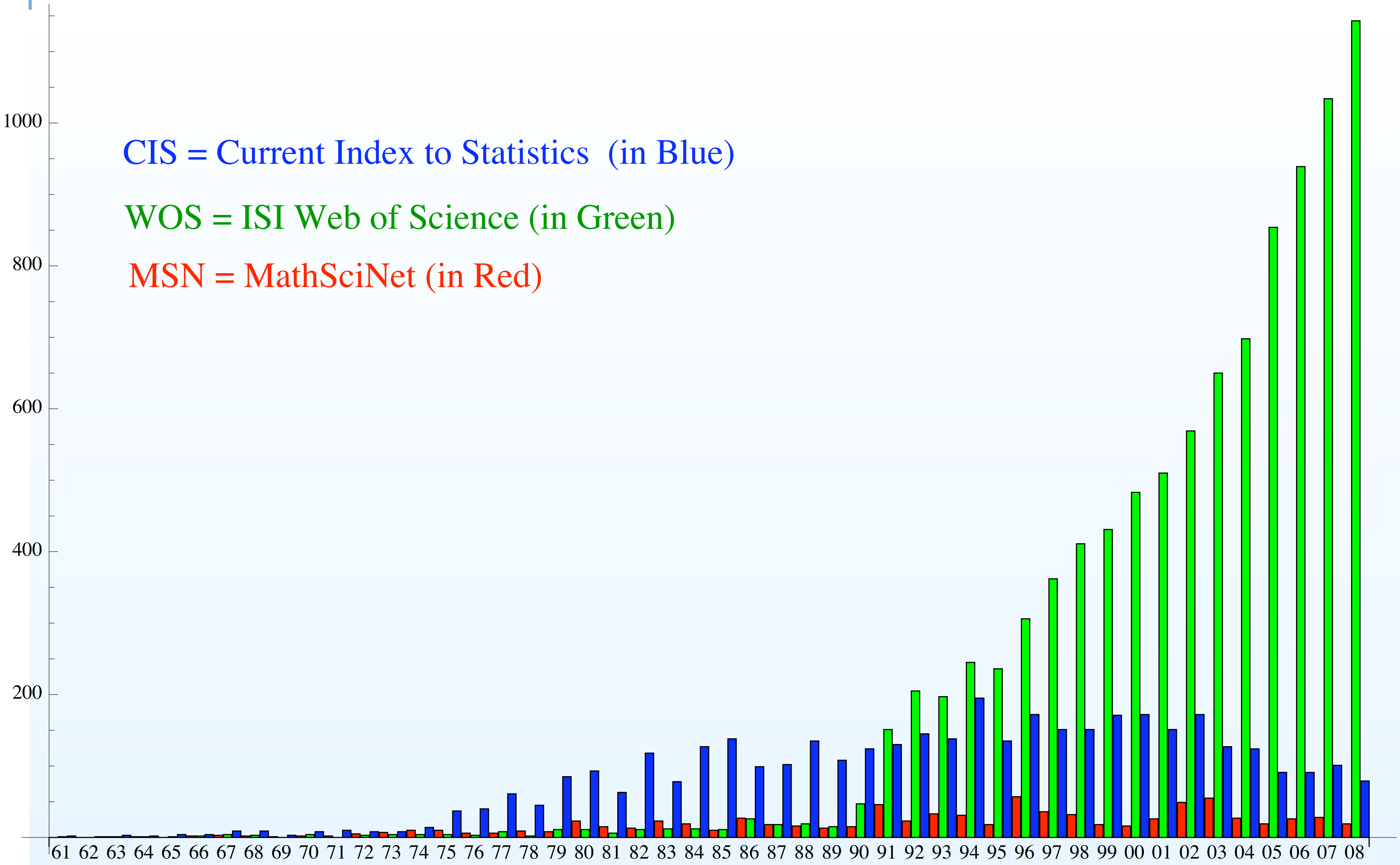
- Topics not mentioned at the 1967 UW meeting
    - Bootstrap methods
    - Markov chain monte carlo
    - Empirical process theory / methods
    - Lasso + regression
- Many other possible topics (not pursued here)
    - Hierarchical models
    - Metaanalysis
    - Nonparametric function estimation
    - Causal inference
    - Missing data



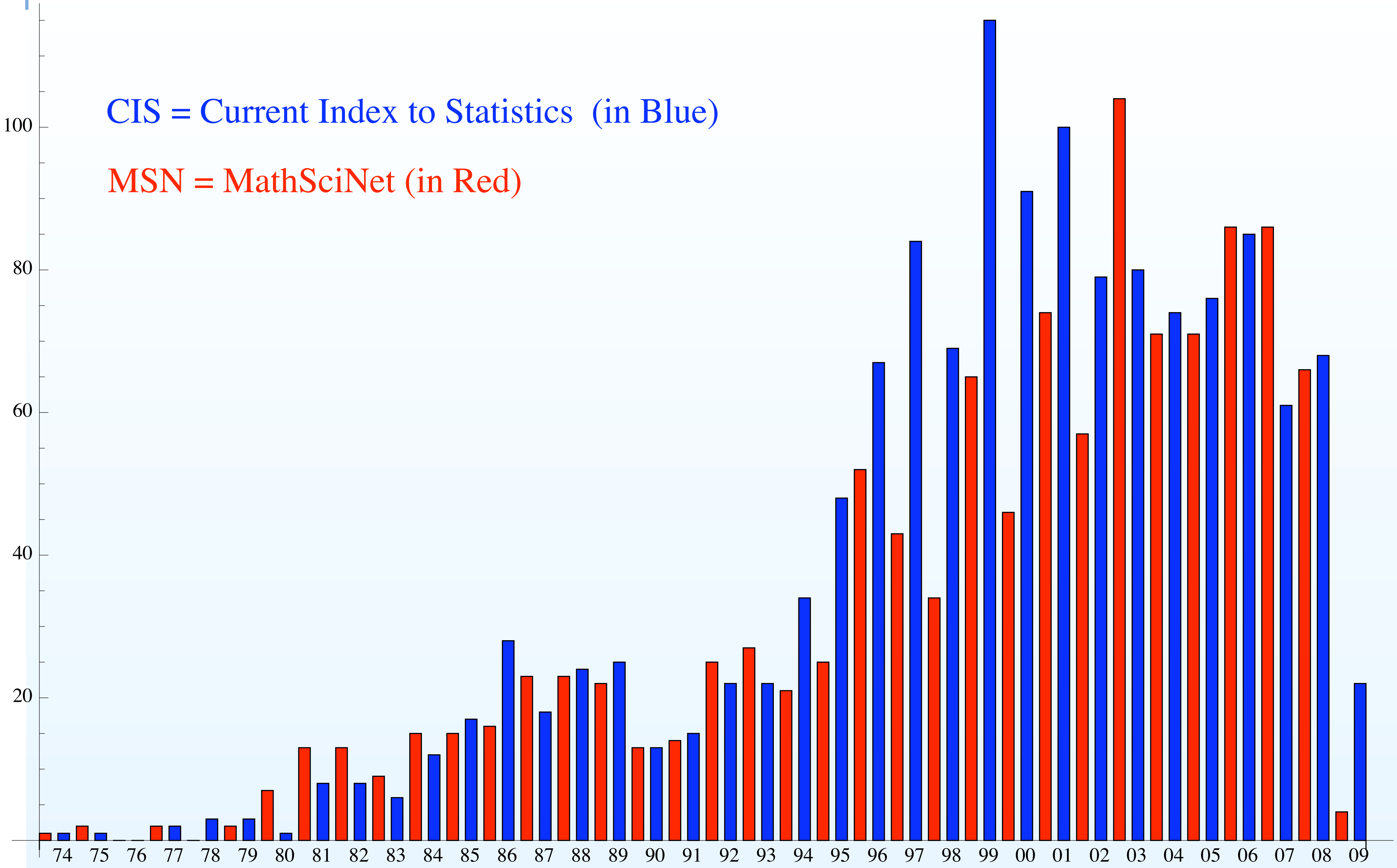
Numbers of papers by year, Nonparametric Bayes



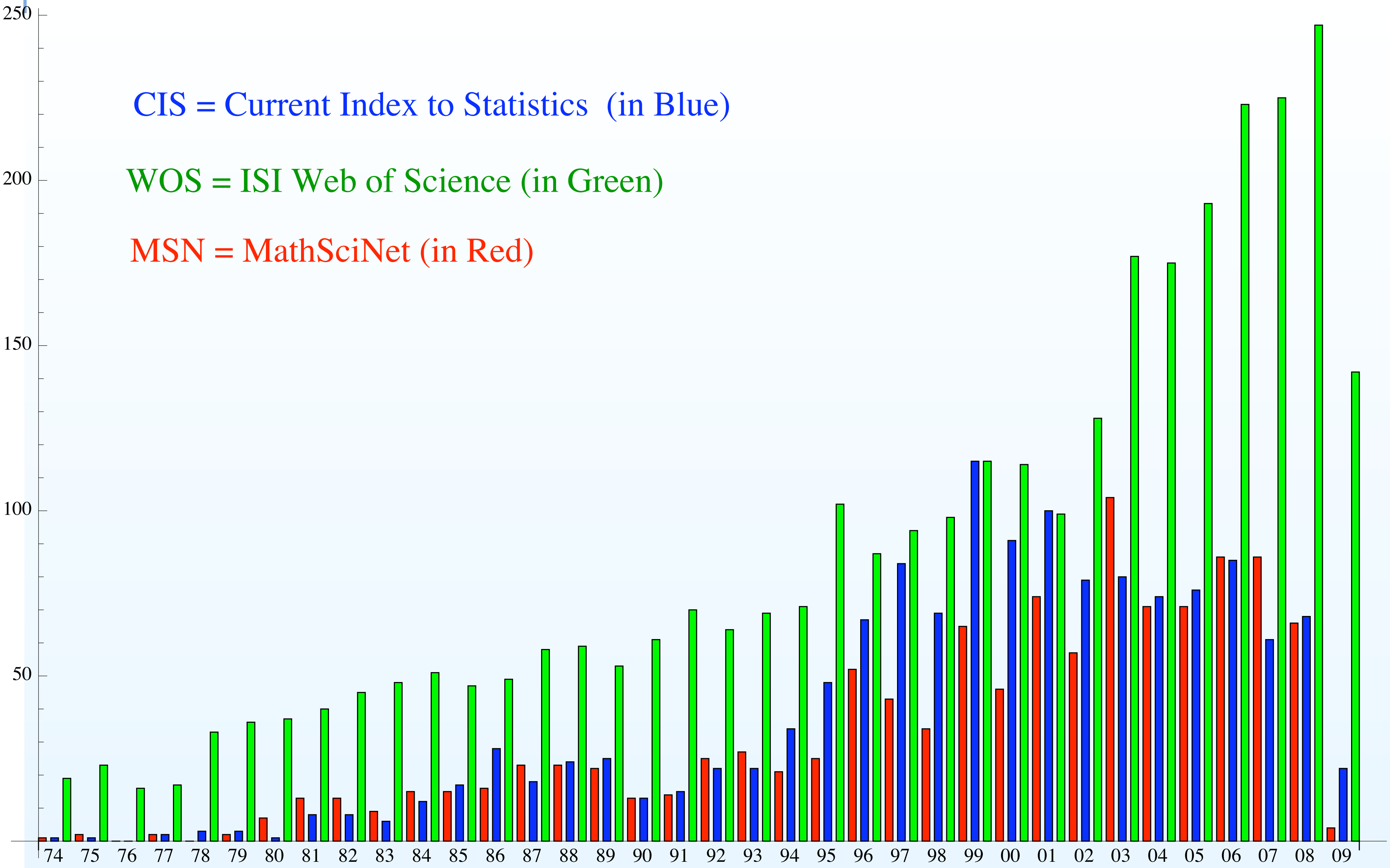
Numbers of papers by year, Robustness



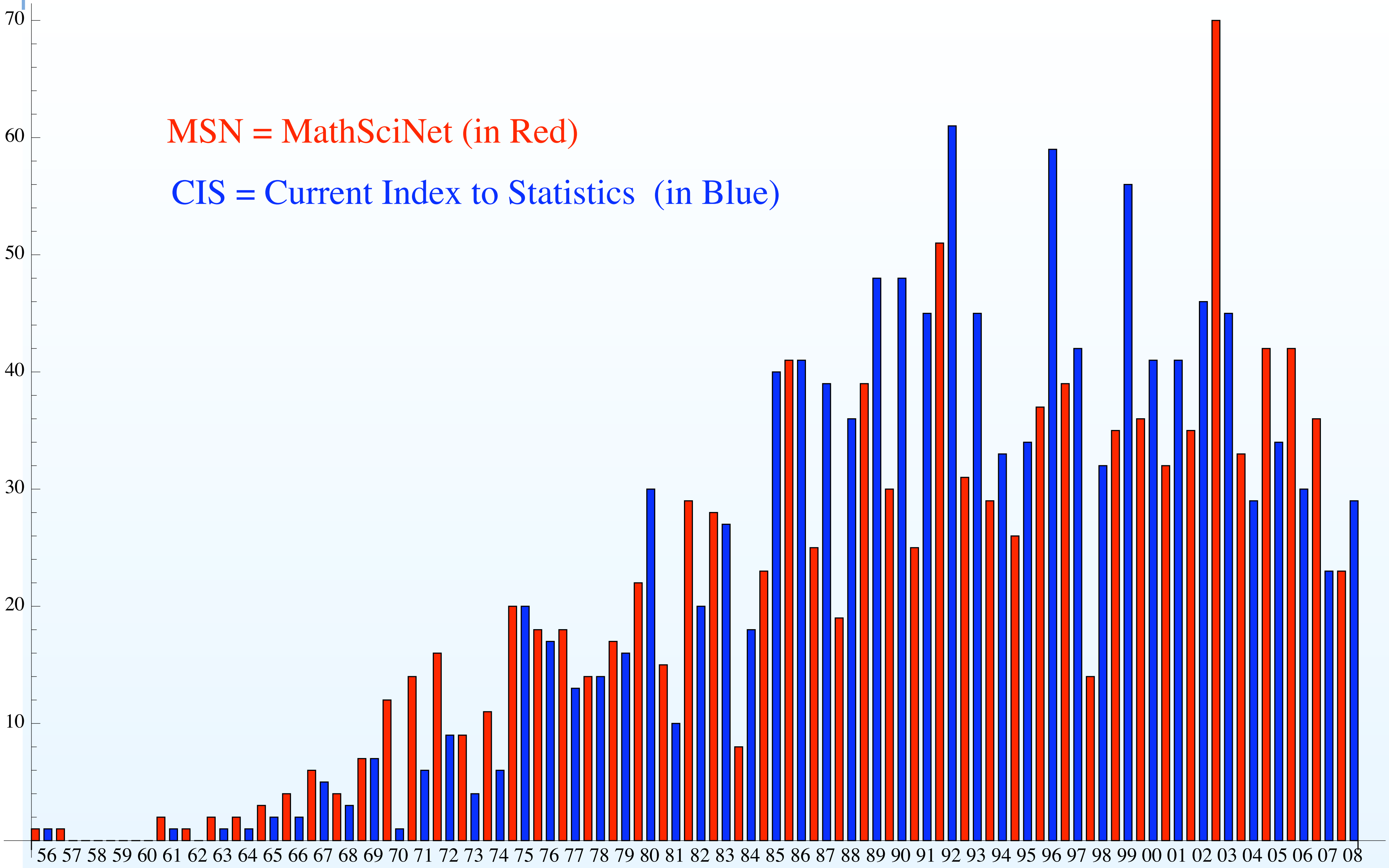
Numbers of papers by year, Robustness, all of MSN, CIS, WOS



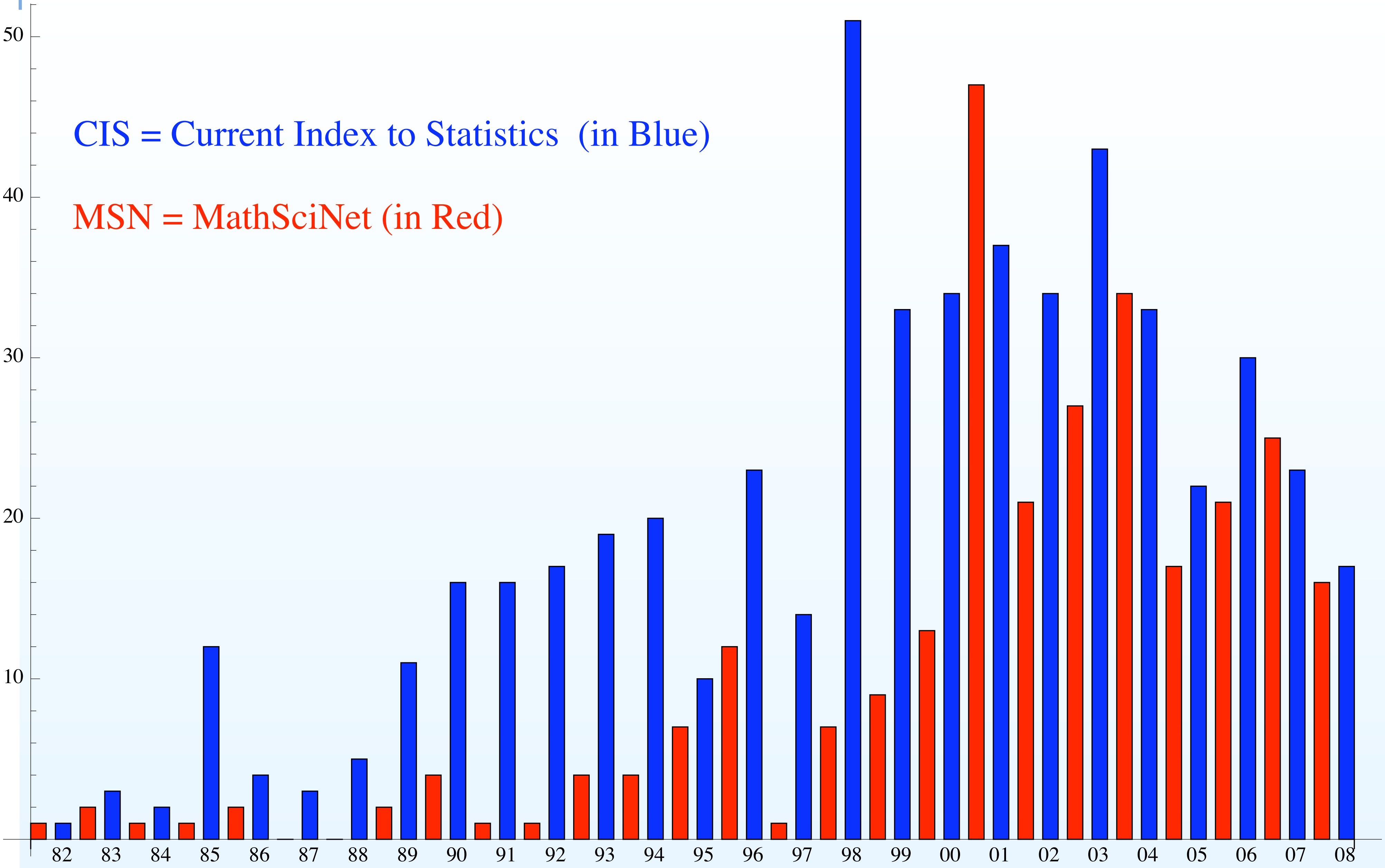
Numbers of papers by year, Model Selection



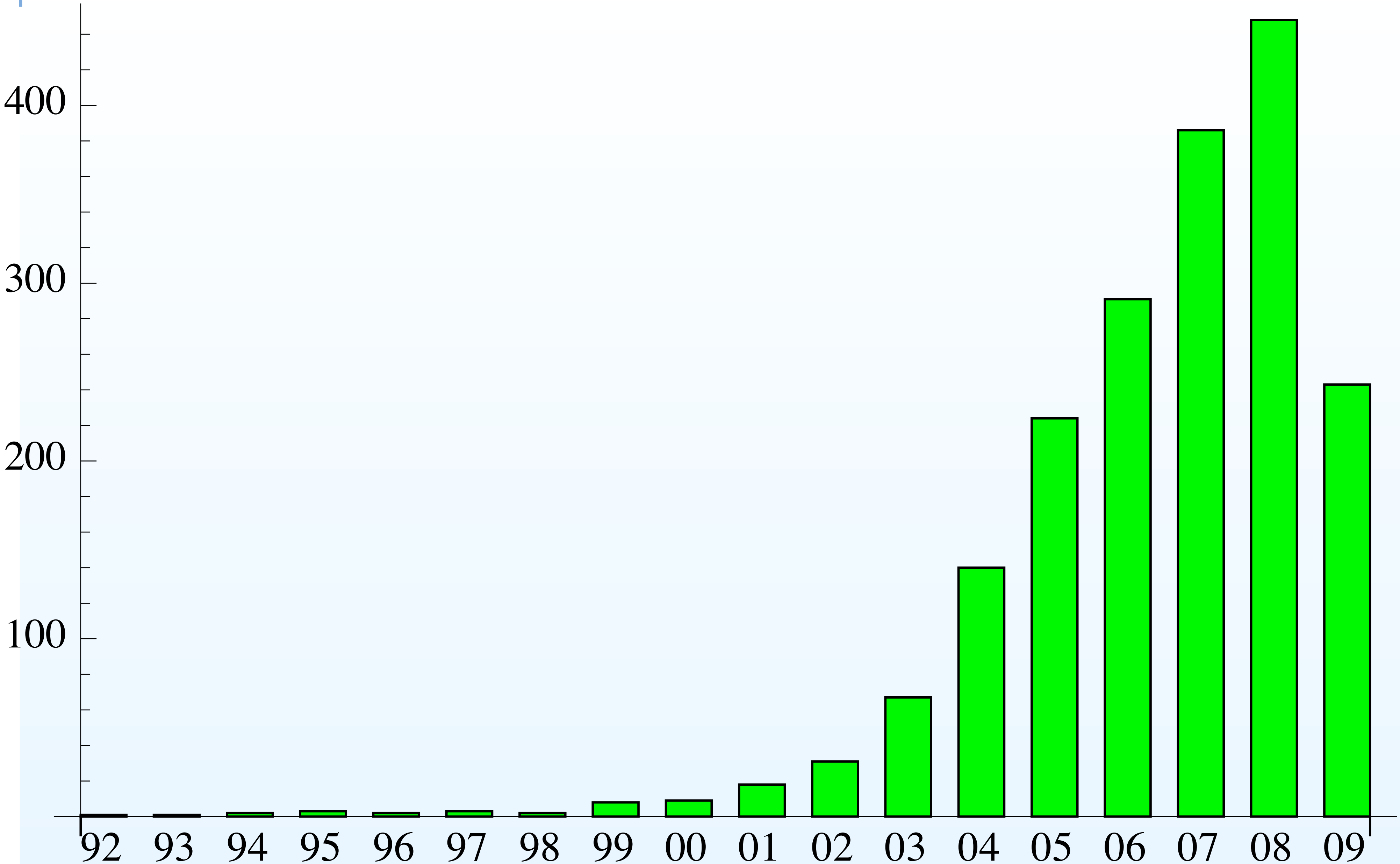
Numbers of papers by year, Model Selection



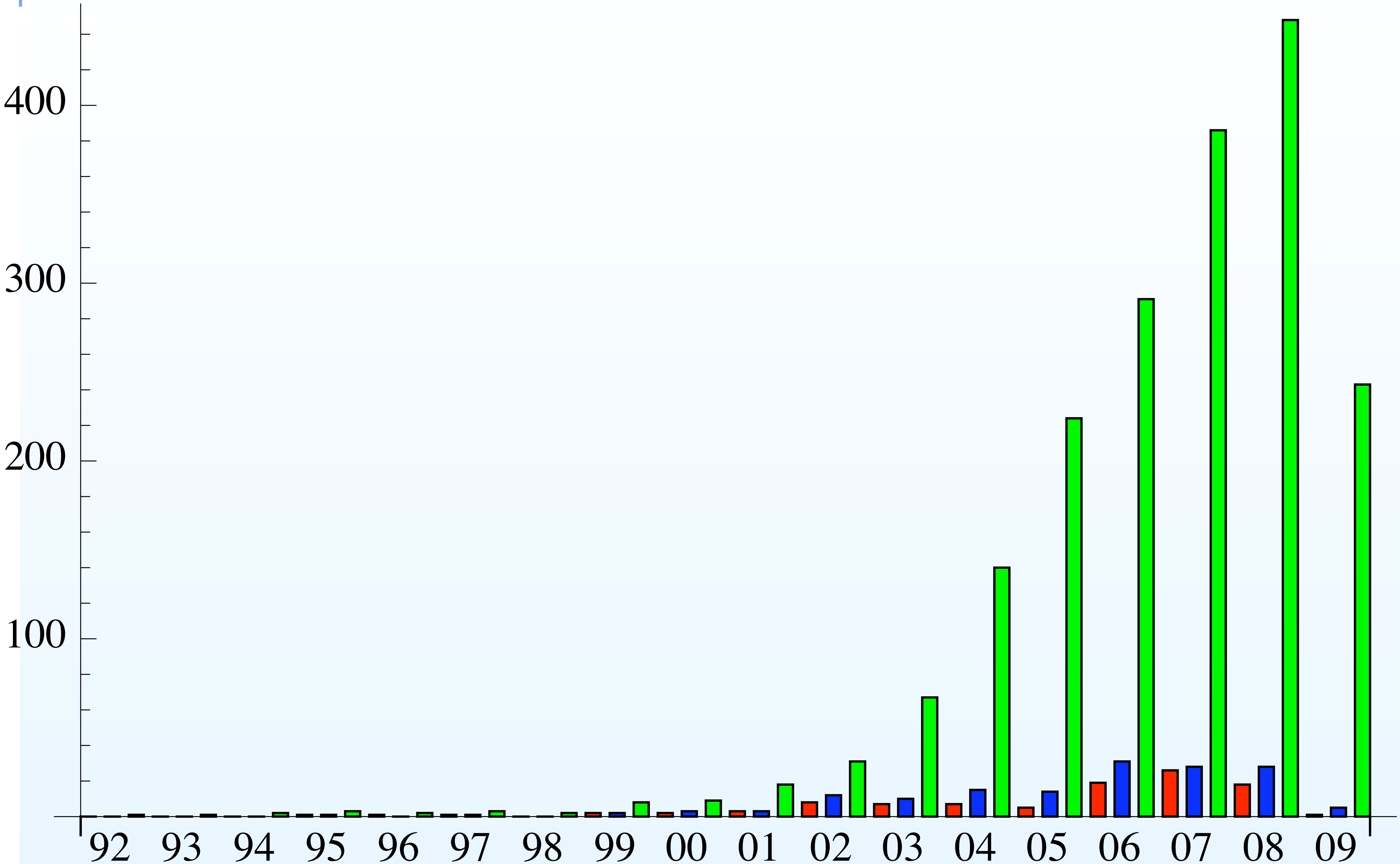
Numbers of papers by year, Empirical Bayes estimation



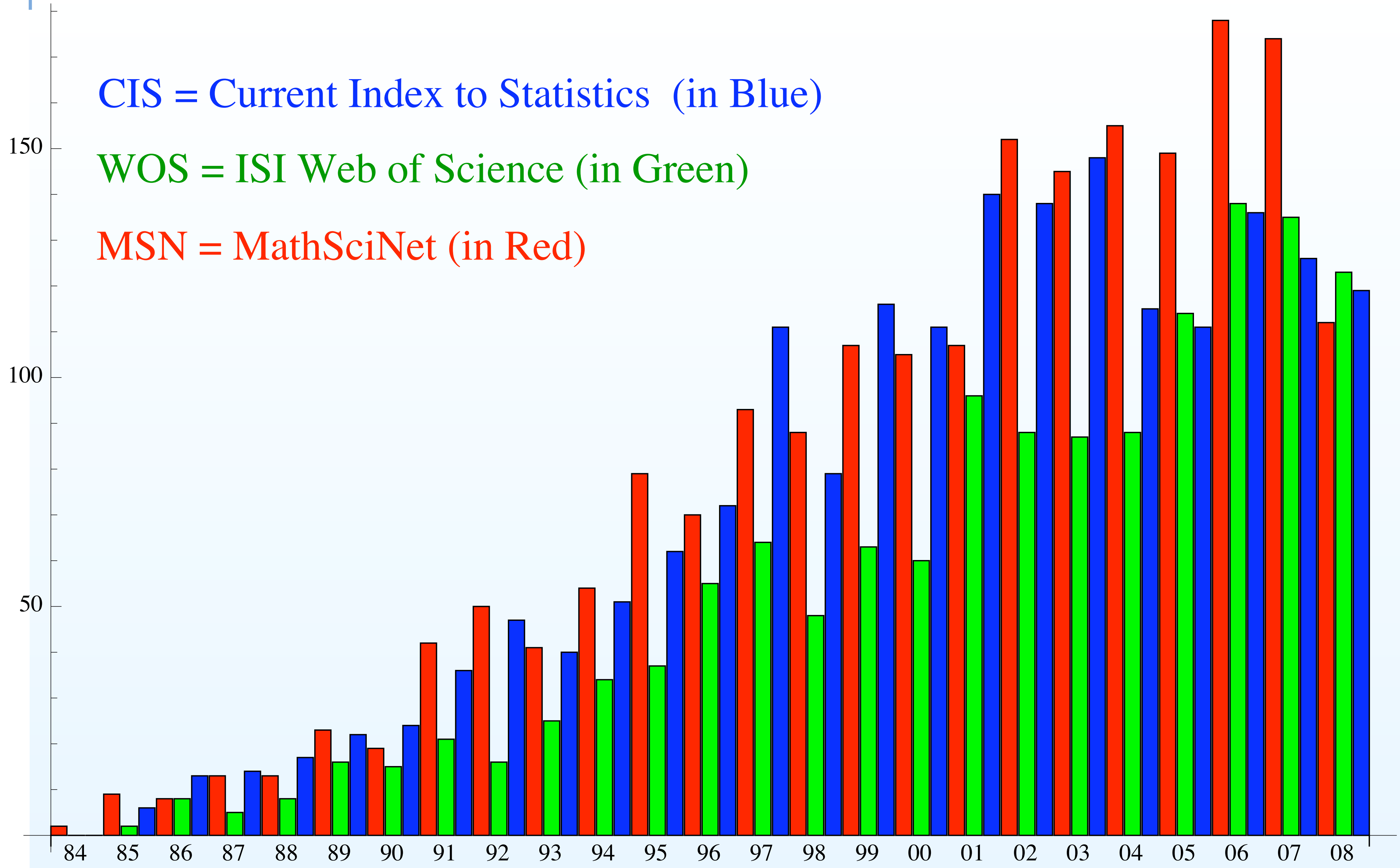
Numbers of papers by year, Graphical Model



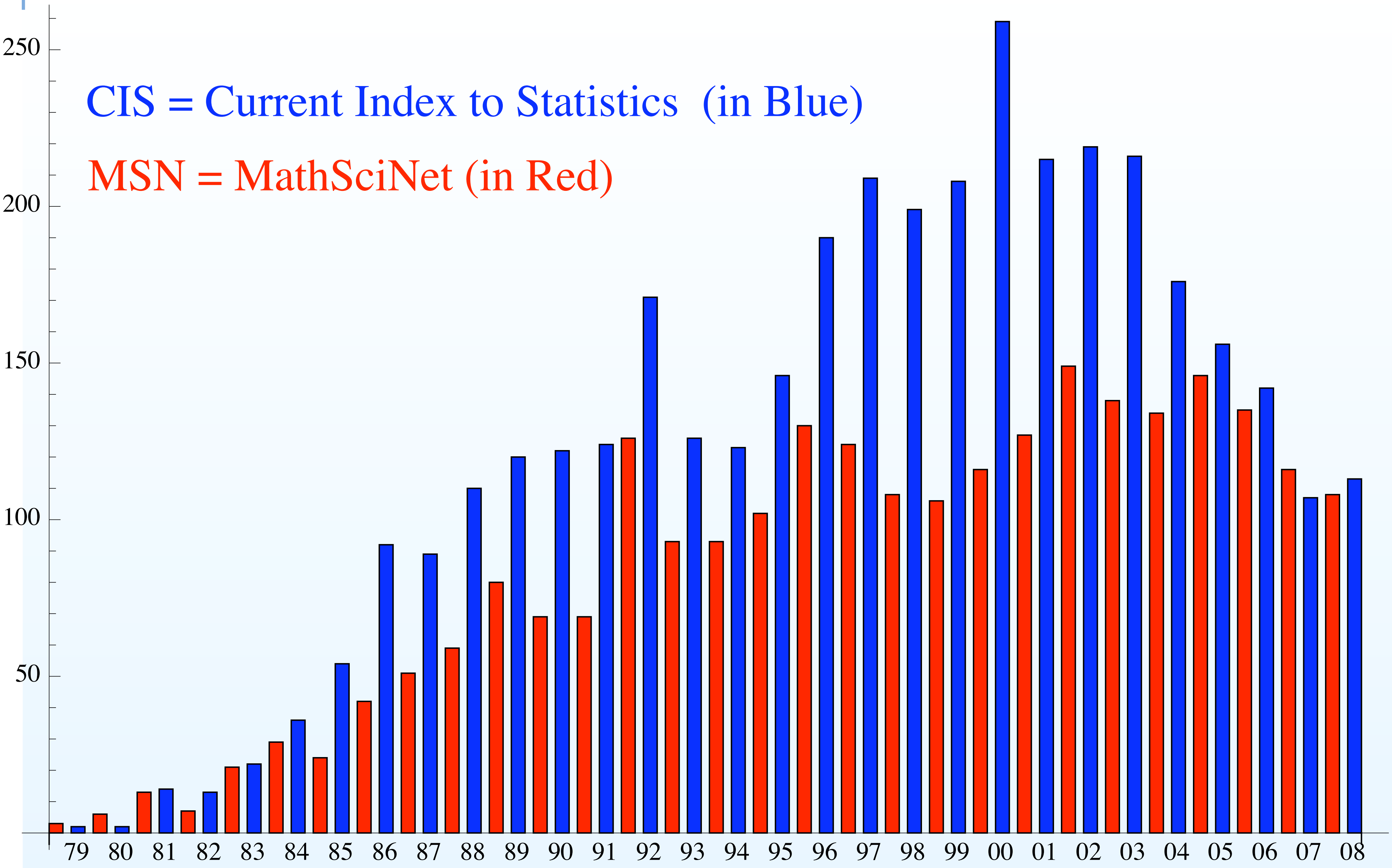
Numbers of papers by year, False Discovery Rate: green=WOS



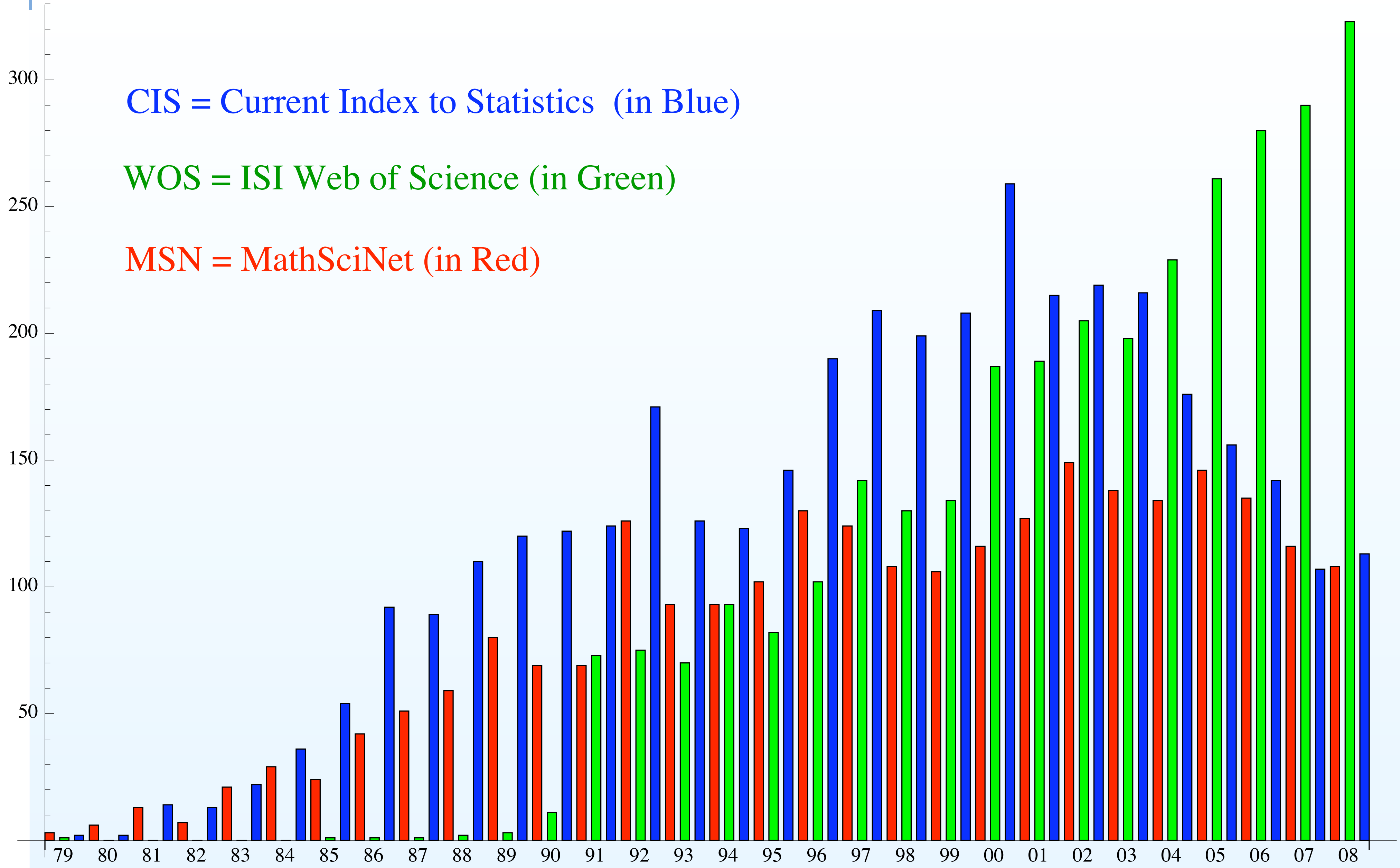
Numbers of papers by year, False Discovery Rate, combined MSN, CIS, WOS



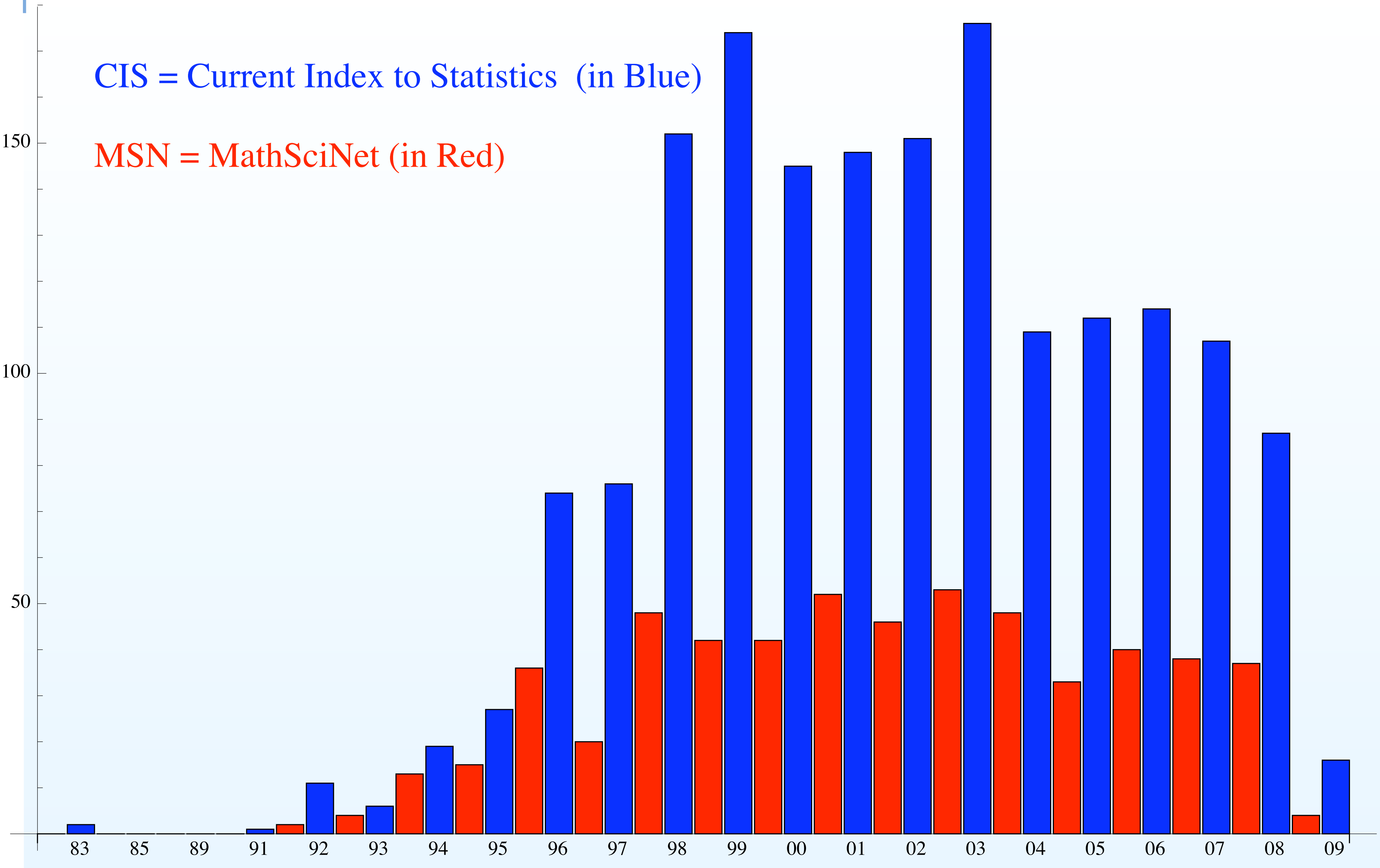
Numbers of papers by year, Semiparametric Models



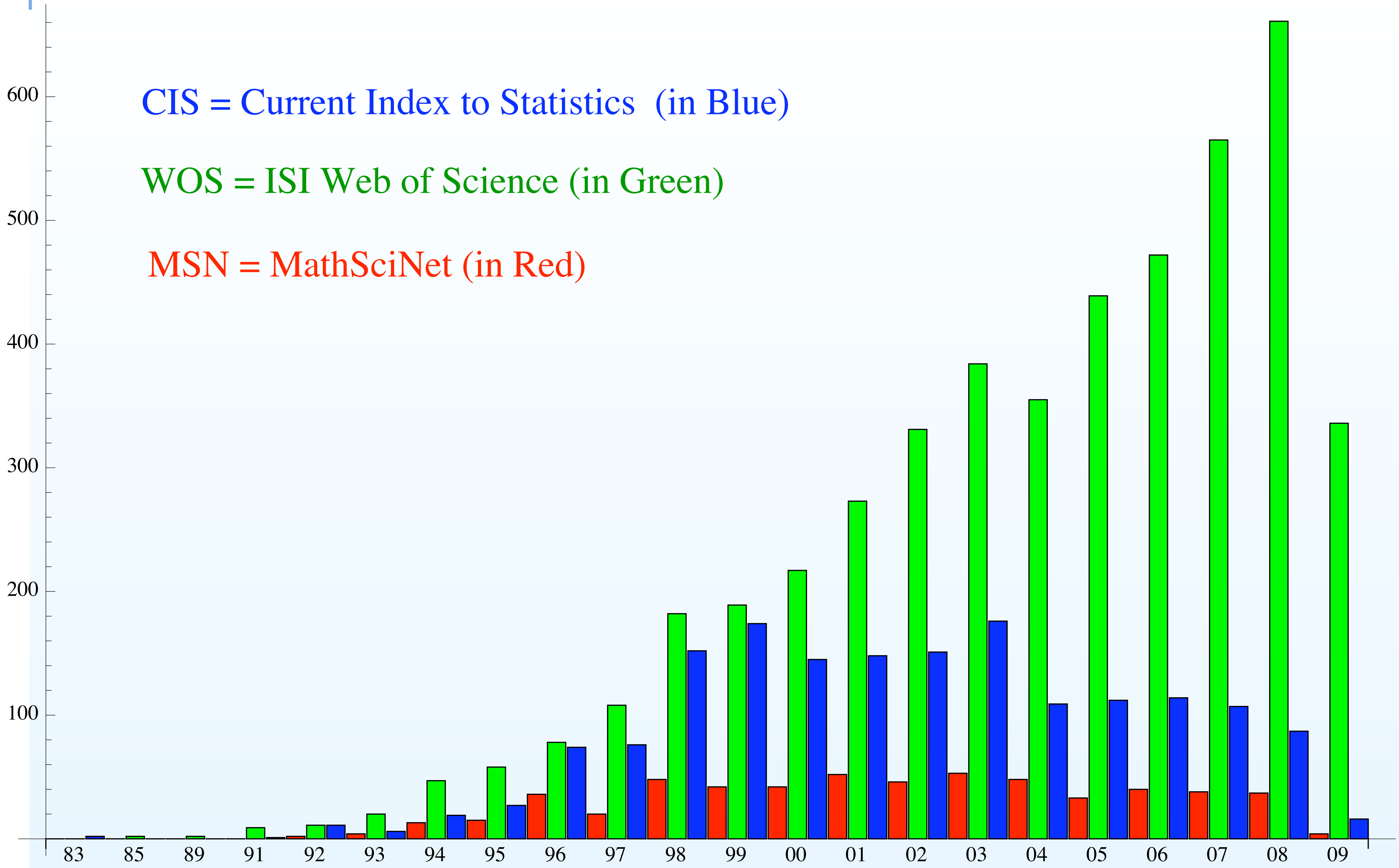
Numbers of papers by year, Bootstrap Methods



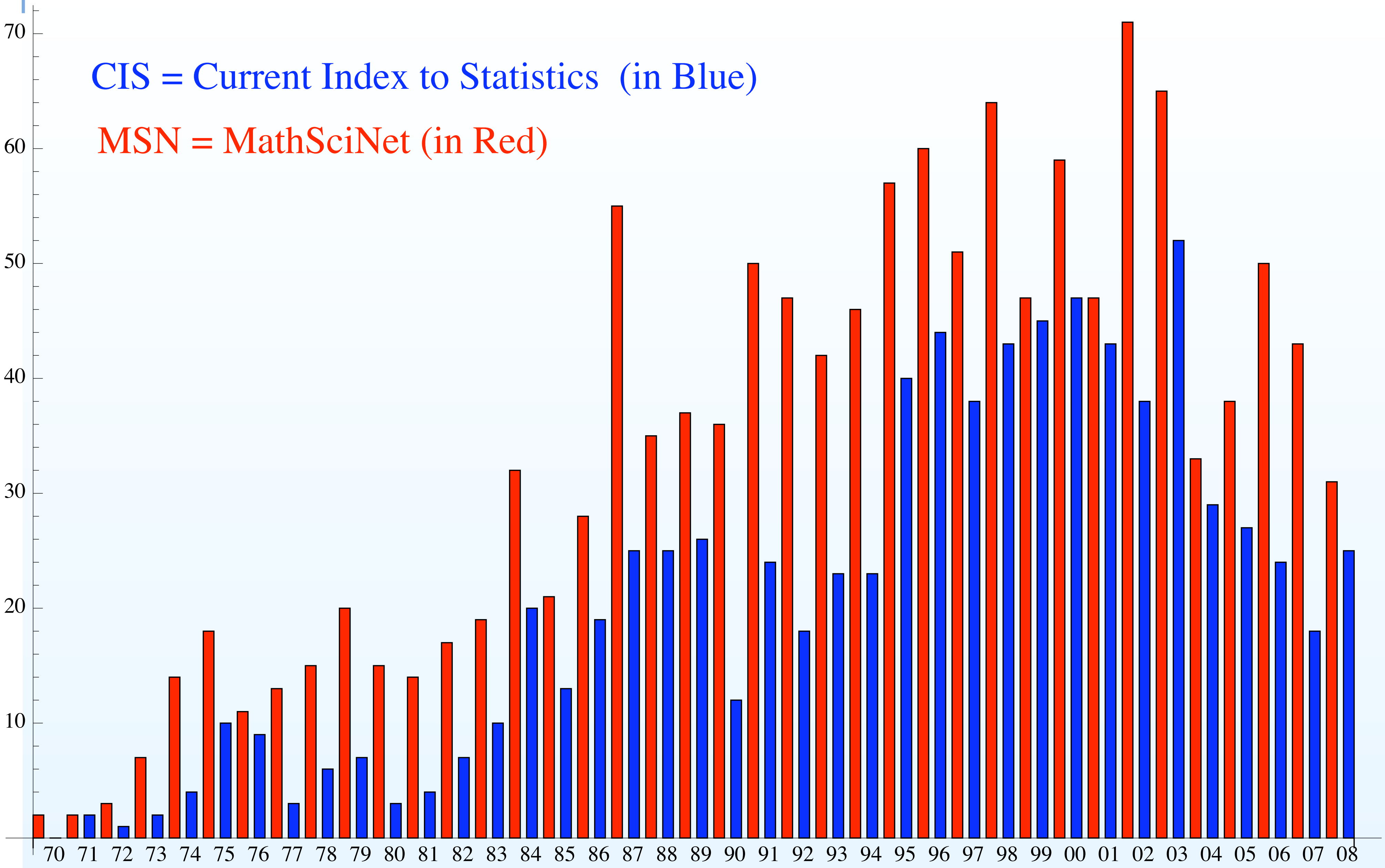
Numbers of papers by year, Bootstrap Methods, all of MSN, CIS, and WOS



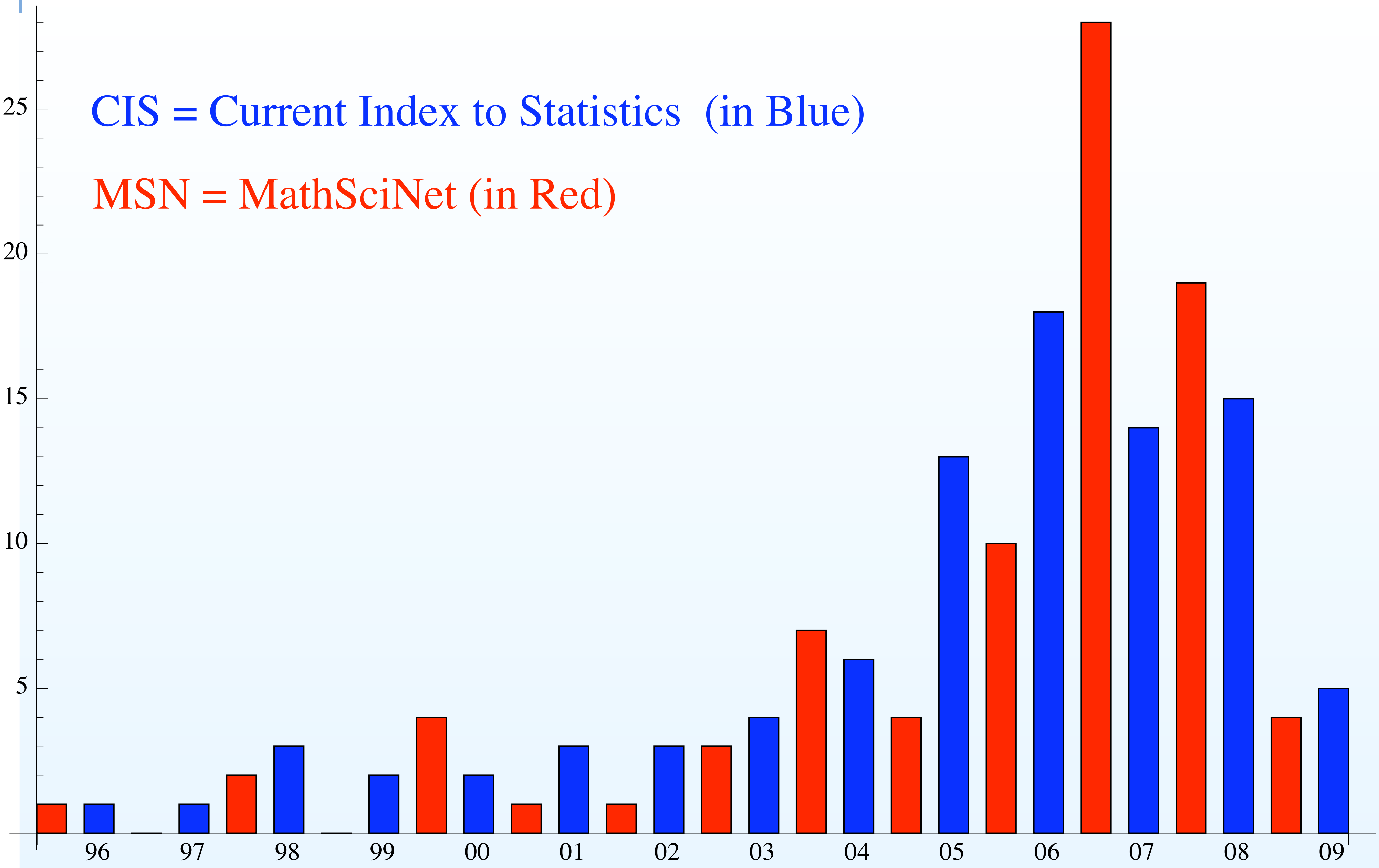
Numbers of papers by year, Markov chain monte carlo



Numbers of papers by year, Markov chain monte carlo



Numbers of papers by year, Empirical Processes



Numbers of papers by year, Lasso + Regression

3. Possible future directions?

The best way to predict the future is to invent it. (Alan Kay)

- Inference for networks, trees, and graphs. (Non-Euclidean data and/or parameters)

- Inference for networks, trees, and graphs. (Non-Euclidean data and/or parameters)
  - Recent talks by Bickel and Chen, Fienberg, and Marron (IMS-China meeting)

- Inference for networks, trees, and graphs. (Non-Euclidean data and/or parameters)
  - Recent talks by Bickel and Chen, Fienberg, and Marron (IMS-China meeting)
  - More math connections possible and needed: probability models, combinatorics, computational algebra development (and limit theory) ahead of statisticians?

- Inference for networks, trees, and graphs. (Non-Euclidean data and/or parameters)
  - Recent talks by Bickel and Chen, Fienberg, and Marron (IMS-China meeting)
  - More math connections possible and needed: probability models, combinatorics, computational algebra development (and limit theory) ahead of statisticians?
    - Probability models; combinatorics, computational algebra .

- Inference for networks, trees, and graphs. (Non-Euclidean data and/or parameters)
  - Recent talks by Bickel and Chen, Fienberg, and Marron (IMS-China meeting)
  - More math connections possible and needed: probability models, combinatorics, computational algebra development (and limit theory) ahead of statisticians?
    - Probability models; combinatorics, computational algebra .
    - Algorithms.

- Inference for networks, trees, and graphs. (Non-Euclidean data and/or parameters)
  - Recent talks by Bickel and Chen, Fienberg, and Marron (IMS-China meeting)
  - More math connections possible and needed: probability models, combinatorics, computational algebra development (and limit theory) ahead of statisticians?
    - Probability models; combinatorics, computational algebra .
    - Algorithms.
    - One basis for work by Bickel and Chen: theory developed by Aldous (1985) and Kallenberg

- Inference for networks, trees, and graphs. (Non-Euclidean data and/or parameters)
  - Recent talks by Bickel and Chen, Fienberg, and Marron (IMS-China meeting)
  - More math connections possible and needed: probability models, combinatorics, computational algebra development (and limit theory) ahead of statisticians?
    - Probability models; combinatorics, computational algebra .
    - Algorithms.
    - One basis for work by Bickel and Chen: theory developed by Aldous (1985) and Kallenberg
  - Networks ubiquitous: internet, social networks, citation networks, ...

- Efficiency theory incorporating computational expense or effort required

- Efficiency theory incorporating computational expense or effort required
    - Meinshausen, Bickel, and Rice (2009).

- Efficiency theory incorporating computational expense or effort required
    - Meinshausen, Bickel, and Rice (2009).
    - Wainwright (2006).

- Efficiency theory incorporating computational expense or effort required
  - Meinshausen, Bickel, and Rice (2009).
  - Wainwright (2006).
  - Stark (2008): “controlled computational cost” in inverse problems.

- Efficiency theory incorporating computational expense or effort required
  - Meinshausen, Bickel, and Rice (2009).
  - Wainwright (2006).
  - Stark (2008): “controlled computational cost” in inverse problems.
- Interactions between computational methods convex analysis, optimization theory, and statistics

- Efficiency theory incorporating computational expense or effort required
  - Meinshausen, Bickel, and Rice (2009).
  - Wainwright (2006).
  - Stark (2008): "controlled computational cost" in inverse problems.
- Interactions between computational methods convex analysis, optimization theory, and statistics
  - Variational approximations: Jaakkola and Jordan (1999), Wainright and Jordan (2008)

- Efficiency theory incorporating computational expense or effort required
  - Meinshausen, Bickel, and Rice (2009).
  - Wainwright (2006).
  - Stark (2008): “controlled computational cost” in inverse problems.
- Interactions between computational methods convex analysis, optimization theory, and statistics
  - Variational approximations: Jaakkola and Jordan (1999), Wainright and Jordan (2008)
  - Convex relaxations: Donoho & ...

- Efficiency theory incorporating computational expense or effort required
  - Meinshausen, Bickel, and Rice (2009).
  - Wainwright (2006).
  - Stark (2008): “controlled computational cost” in inverse problems.
- Interactions between computational methods convex analysis, optimization theory, and statistics
  - Variational approximations: Jaakkola and Jordan (1999), Wainright and Jordan (2008)
  - Convex relaxations: Donoho & ...
  - Stability of algorithms (or lack of stability), e.g. inverse Laplace transform

- Efficiency theory incorporating computational expense or effort required
  - Meinshausen, Bickel, and Rice (2009).
  - Wainwright (2006).
  - Stark (2008): "controlled computational cost" in inverse problems.
- Interactions between computational methods convex analysis, optimization theory, and statistics
  - Variational approximations: Jaakkola and Jordan (1999), Wainright and Jordan (2008)
  - Convex relaxations: Donoho & ...
  - Stability of algorithms (or lack of stability), e.g. inverse Laplace transform
  - Scalability of algorithms (or lack of scalability).

- Efficiency theory incorporating computational expense or effort required
    - Meinshausen, Bickel, and Rice (2009).
    - Wainwright (2006).
    - Stark (2008): “controlled computational cost” in inverse problems.
- Interactions between computational methods convex analysis, optimization theory, and statistics
    - Variational approximations: Jaakkola and Jordan (1999), Wainright and Jordan (2008)
    - Convex relaxations: Donoho & ...
    - Stability of algorithms (or lack of stability), e.g. inverse Laplace transform
    - Scalability of algorithms (or lack of scalability).
- Replacements for Markov chain monte carlo in special classes of models

- Efficiency theory incorporating computational expense or effort required
  - Meinshausen, Bickel, and Rice (2009).
  - Wainwright (2006).
  - Stark (2008): “controlled computational cost” in inverse problems.
- Interactions between computational methods convex analysis, optimization theory, and statistics
  - Variational approximations: Jaakkola and Jordan (1999), Wainright and Jordan (2008)
  - Convex relaxations: Donoho & ...
  - Stability of algorithms (or lack of stability), e.g. inverse Laplace transform
  - Scalability of algorithms (or lack of scalability).
- Replacements for Markov chain monte carlo in special classes of models
  - Nested Laplace approximatons: Rue, Martino, Chopin (2008)

- Sparsity, sparse representations, compressed sensing
  - Ingster (1993a,1993b, 1997)
  - Searching for sparse signals: Donoho and Jin (2004), Johnstone and Silverman (2004)
  - Estimation of the proportion of sparse signals: Meinshausen and Rice (2006), Cai, Jin and Low (2007), Jin and Cai (2007)

- Sparsity, sparse representations, compressed sensing
  - Ingster (1993a,1993b, 1997)
  - Searching for sparse signals: Donoho and Jin (2004), Johnstone and Silverman (2004)
  - Estimation of the proportion of sparse signals: Meinshausen and Rice (2006), Cai, Jin and Low (2007), Jin and Cai (2007)
- More on model selection
  - Enormous qualitative changes in past 10 years:
  - Changing perspectives: often no one “true” model.
  - Replace with specified goals: prediction or variable/feature selection
  - Often now based on “model averaging”, or “weighting”, or “aggregation” methods.
  - Need for much more work on inference following model selection (e.g. H. Leeb, B. Pötscher)

4. Some specific problems (of special interest to me)

I never think of the future - it comes soon enough. (Albert Einstein)

- Shape restrictions and mixture models

- Shape restrictions and mixture models
  - Nonparametric mixture models with smooth kernels: limiting distribution of MLE at fixed points? Example: Completely monotone densities (i.e. scale mixtures of exponential); consistency of MLE due to Jewell (1982).

- Shape restrictions and mixture models
    - Nonparametric mixture models with smooth kernels: limiting distribution of MLE at fixed points? Example: Completely monotone densities (i.e. scale mixtures of exponential); consistency of MLE due to Jewell (1982).
    - Multivariate shape restrictions: e.g. log-concave densities on $\mathbb{R}^d$
      R. Samworth (Cambridge), A. Seregin (U. Washington).

- Shape restrictions and mixture models
    - Nonparametric mixture models with smooth kernels: limiting distribution of MLE at fixed points? Example: Completely monotone densities (i.e. scale mixtures of exponential); consistency of MLE due to Jewell (1982).
    - Multivariate shape restrictions: e.g. log-concave densities on $\mathbb{R}^d$
      R. Samworth (Cambridge), A. Seregin (U. Washington).
    - Inefficiency of MLE for shape restricted families when $d \geq d_0$?!
      How to penalize or further constrain to find rate-efficient estimators in the classes when $d \geq d_0$?

- Shape restrictions and mixture models
  - Nonparametric mixture models with smooth kernels: limiting distribution of MLE at fixed points? Example: Completely monotone densities (i.e. scale mixtures of exponential); consistency of MLE due to Jewell (1982).
  - Multivariate shape restrictions: e.g. log-concave densities on $\mathbb{R}^d$
    R. Samworth (Cambridge), A. Seregin (U. Washington).
  - Inefficiency of MLE for shape restricted families when $d \geq d_0$?!
    How to penalize or further constrain to find rate-efficient estimators in the classes when $d \geq d_0$?
  - Interactions between (nonparametric) mixture models and Empirical Bayes methods: C-H Zhang (2009), Jiang and Zhang (2009).

- Sampling based (Horvitz - Thompson) empirical processes

- Sampling based (Horvitz - Thompson) empirical processes
  - $\{\xi_i\}_{i=1}^N$, sampling indicators;
    $\{\pi_i\}_{i=1}^N$, marginal inclusion probabilities;
    $\{X_i\}_{i=1}^N$, population to be sampled.

$$\mathbb{P}_N^\pi = N^{-1}\sum_{i=1}^N \frac{\xi_i}{\pi_i}\delta_{X_i} = \text{Horvitz - Thompson empirical measure,}$$

$$\mathbb{G}_N^\pi = \sqrt{N}(\mathbb{P}_N^\pi - P)$$

- Sampling based (Horvitz - Thompson) empirical processes
  - $\{\xi_i\}_{i=1}^N$, sampling indicators;
    $\{\pi_i\}_{i=1}^N$, marginal inclusion probabilities;
    $\{X_i\}_{i=1}^N$, population to be sampled.

$$\mathbb{P}_N^\pi = N^{-1}\sum_{i=1}^N \frac{\xi_i}{\pi_i}\delta_{X_i} = \text{Horvitz - Thompson empirical measure,}$$

$$\mathbb{G}_N^\pi = \sqrt{N}(\mathbb{P}_N^\pi - P)$$

  - Finite - dimensional central limit theory known: Hájek (1960, 1964); Rosén (1972).

- Sampling based (Horvitz - Thompson) empirical processes
    - $\{\xi_i\}_{i=1}^N$, sampling indicators;
      $\{\pi_i\}_{i=1}^N$, marginal inclusion probabilities;
      $\{X_i\}_{i=1}^N$, population to be sampled.

$$\mathbb{P}_N^\pi = N^{-1}\sum_{i=1}^N \frac{\xi_i}{\pi_i}\delta_{X_i} = \text{Horvitz - Thompson empirical measure},$$

$$\mathbb{G}_N^\pi = \sqrt{N}(\mathbb{P}_N^\pi - P)$$

    - Finite - dimensional central limit theory known: Hájek (1960, 1964); Rosén (1972).
    - Weak convergence theory of $\mathbb{G}_N^\pi$ known so far only in special cases:
      Bernoulli sampling; two phase sampling without replacement in strata:
      Breslow and Wellner (2007, 2008)

- Sampling based (Horvitz - Thompson) empirical processes
  - $\{\xi_i\}_{i=1}^N$, sampling indicators;
    $\{\pi_i\}_{i=1}^N$, marginal inclusion probabilities;
    $\{X_i\}_{i=1}^N$, population to be sampled.

$$\mathbb{P}_N^\pi = N^{-1}\sum_{i=1}^N \frac{\xi_i}{\pi_i}\delta_{X_i} = \text{Horvitz - Thompson empirical measure},$$

$$\mathbb{G}_N^\pi = \sqrt{N}(\mathbb{P}_N^\pi - P)$$

  - Finite - dimensional central limit theory known: Hájek (1960, 1964); Rosén (1972).
  - Weak convergence theory of $\mathbb{G}_N^\pi$ known so far only in special cases:
    Bernoulli sampling; two phase sampling without replacement in strata:
    Breslow and Wellner (2007, 2008)
  - Weak convergence theory for $\mathbb{G}_N^\pi$ not known in general.

- Sampling based (Horvitz - Thompson) empirical processes
  - $\{\xi_i\}_{i=1}^N$, sampling indicators;
    $\{\pi_i\}_{i=1}^N$, marginal inclusion probabilities;
    $\{X_i\}_{i=1}^N$, population to be sampled.

$$\mathbb{P}_N^\pi = N^{-1}\sum_{i=1}^N \frac{\xi_i}{\pi_i}\delta_{X_i} = \text{Horvitz - Thompson empirical measure,}$$

$$\mathbb{G}_N^\pi = \sqrt{N}(\mathbb{P}_N^\pi - P)$$

  - Finite - dimensional central limit theory known: Hájek (1960, 1964); Rosén (1972).
  - Weak convergence theory of $\mathbb{G}_N^\pi$ known so far only in special cases:
    Bernoulli sampling; two phase sampling without replacement in strata:
    Breslow and Wellner (2007, 2008)
  - Weak convergence theory for $\mathbb{G}_N^\pi$ not known in general.

- Sampling based (Horvitz - Thompson) empirical processes, continued

- Sampling based (Horvitz - Thompson) empirical processes, continued
    - Unifications with “model assisted survey sampling” and survey design: e.g. Saarndal, Swensson, and Wretman (1992); Lumley (2009).

- Sampling based (Horvitz - Thompson) empirical processes, continued
  - Unifications with “model assisted survey sampling” and survey design: e.g. Saarndal, Swensson, and Wretman (1992); Lumley (2009).
- Estimators for (semiparametric) models under miss-specification

- Sampling based (Horvitz - Thompson) empirical processes, continued
    - Unifications with “model assisted survey sampling” and survey design: e.g. Saarndal, Swensson, and Wretman (1992); Lumley (2009).
- Estimators for (semiparametric) models under miss-specification
    - Behavior of estimators when $P \notin \mathcal{P}$?

- Sampling based (Horvitz - Thompson) empirical processes, continued
    - Unifications with “model assisted survey sampling” and survey design: e.g. Saarndal, Swensson, and Wretman (1992); Lumley (2009).
- Estimators for (semiparametric) models under miss-specification
    - Behavior of estimators when $P \notin \mathcal{P}$?
    - Bickel and Kwon (2001); Kleijn and van der Vaart (2006).

- Sampling based (Horvitz - Thompson) empirical processes, continued
    - Unifications with “model assisted survey sampling” and survey design: e.g. Saarndal, Swensson, and Wretman (1992); Lumley (2009).
- Estimators for (semiparametric) models under miss-specification
    - Behavior of estimators when $P \notin \mathcal{P}$?
    - Bickel and Kwon (2001); Kleijn and van der Vaart (2006).
    - Kosorok, Lee, and Fine (2004), Newey (2004).

The Quotes, Again

- Prediction is very difficult, especially when it involves the future.
  (Niels Bohr, quoted in Alpher and Hermann (2001), p. 159).

The Quotes, Again

- Prediction is very difficult, especially when it involves the future.
  (Niels Bohr, quoted in Alpher and Hermann (2001), p. 159).
- The future ain't what it used to be. (Yogi Berra)

The Quotes, Again

- Prediction is very difficult, especially when it involves the future.
  (Niels Bohr, quoted in Alpher and Hermann (2001), p. 159).
- The future ain't what it used to be. (Yogi Berra)
- The best way to predict the future is to invent it. (Alan Kay)

The Quotes, Again

- Prediction is very difficult, especially when it involves the future.
  (Niels Bohr, quoted in Alpher and Hermann (2001), p. 159).
- The future ain't what it used to be. (Yogi Berra)
- The best way to predict the future is to invent it. (Alan Kay)
- I never think of the future - it comes soon enough. (Albert Einstein)

Some facts:

- In 1967 Leonard J. Savage (1917-1971) was 50.

Some facts:

- In 1967 Leonard J. Savage (1917-1971) was 50.
- In 1967 Lucien Le Cam (1924-2000) was 43.

Some facts:

- In 1967 Leonard J. Savage (1917-1971) was 50.
- In 1967 Lucien Le Cam (1924-2000) was 43.
- In 1967 Jack Kiefer (1924-1981) was 43.

Some facts:

- In 1967 Leonard J. Savage (1917-1971) was 50.
- In 1967 Lucien Le Cam (1924-2000) was 43.
- In 1967 Jack Kiefer (1924-1981) was 43.
- In 1967 David Freedman (1938-2008) was 29.

Some facts:

- In 1967 Leonard J. Savage (1917-1971) was 50.
- In 1967 Lucien Le Cam (1924-2000) was 43.
- In 1967 Jack Kiefer (1924-1981) was 43.
- In 1967 David Freedman (1938-2008) was 29.
- In 1969 (time of the first moon landing) less than half of the current population of the U.S. had been born. (The national median age in the U.S. was 36.7 years in February 2009.)

A few references

- Lindsay, B. G., Kettenring, J., and Siegmund, D. O. (2004). A report on the future of statistics. With comments. *Statist. Sci.* **3**, 387 - 413
- Lindsay, B. G., Kettenring, J., and Siegmund, D. O. (2004). Statistics: Challenges and Opportunities for the Twenty-First Century. http://www.stat.psu.edu/ bgl/nsf report.pdf .
- Raftery, A. E., Tanner, M. A., and Wells, M. T.(2002). *Statistics in the 21st Century*. Chapman and Hall/CRC. Boca Raton.
- Watts, D. G. (1968). *The Future of Statistics.* Academic Press. New York

- Bickel, P. J., Chen, A. (2009). Block models and community detection. *Proc. Nat. Acad. Sci.* **xx**, yy-zz.
- Meinshausen, N., Bickel, P. J., and Rice, J. (2009). Efficient blind search: optimal power of detection under computational cost constraints. *Ann. Appl. Statist.* **3**, 38 - 60.
- Wainright, M. J. (2006). Estimating the “wrong” graphical model: benefits in the computation-limited setting. *J. Mach. Learning. Res.* **7**, 1829 - 1859.
- Stark, P. B. (2008). Generalizing resolution. *Inverse Problems* **24**, 1 - 17.
- Zhang, C.-H. (2009). Generalized maximum likelihood estimation of normal mixture densities. *Statistica Sinica* **19**, 1297 - 1318.
- Jiang, W. and Zhang, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.* **37**, 1647 - 1684.

- Rue, H., Martino, S., and Chopin, N. (2008). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. Technical Report.
- Jaakkola, T. S. and Jordan, M.I. (1999). Improving the mean field approximation via the use of mixture distributions. In *Learning in Graphical Models*, (M. I. Jordan, ed.), 105-161, MIT Press, 1999.
- Wainright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning* **1**, 1-305.
- Lin, D. Y. (2000). On fitting Cox's proportional hazards models to survey data. *Biometrika* **87**, 37 - 47.
- Davies, P. L. (1995). Data features. *Statistica Neerl.* **49**, 185-245.
- Davies, P. L. (2008). Approximating data. *J. Korean Statist. Soc.* **37**, 191- 240. (with discussion).