Fractional Polynomials and Model Averaging

Paul C Lambert
Center for Biostatistics and Genetic Epidemiology
University of Leicester
UK
paul.lambert@le.ac.uk

Nordic and Baltic Stata Users Group Meeting, Stockholm
7th September 2007

Fractional Polynomials

- Fractional Polynomials are used in regression models to fit non-linear functions.
- Often preferable to cut-points.
- Functions from fractional polynomials more flexible than from 'standard' polynomials.
- See (Royston and Altman, 1994) or (Sauerbrei and Royston, 1999) for more details.
- Implemented in Stata with `fracpoly` and `mfp` commands.

Powers

- The linear predictor for a fractional polynomial of order M for covariate x can be defined as,

$$\beta_0 + \sum_{m=1}^{M} \beta_m x^{p_m}$$

- where each power p_m is chosen from a restricted set.
- The usual set of powers is

$$\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$$

- x^0 is taken as $\ln(x)$

Selecting the Best Fitting Model

- All combinations of powers are fitted and the 'best' fitting model obtained.
- Using the default set of powers for an FP2 model there are
  - 8 FP1 Models
  - 36 FP2 Models (including 8 repeated powers)
- The best fitting model for fractional polynomials of the same degree can be obtain by minimising the deviance.
- When comparing models of a different degree, e.g. FP2 and FP1 models, the model can be selected using a formal significance test or the Akaike Information Criterion (AIC).

Selecting the Best Fitting Model

- All combinations of powers are fitted and the 'best' fitting model obtained.
- Using the default set of powers for an FP2 model there are
  - 8 FP1 Models
  - 36 FP2 Models (including 8 repeated powers)
- The best fitting model for fractional polynomials of the same degree can be obtain by minimising the deviance.
- When comparing models of a different degree, e.g. FP2 and FP1 models, the model can be selected using a formal significance test or the Akaike Information Criterion (AIC).
- Model selection uncertainty is ignored.

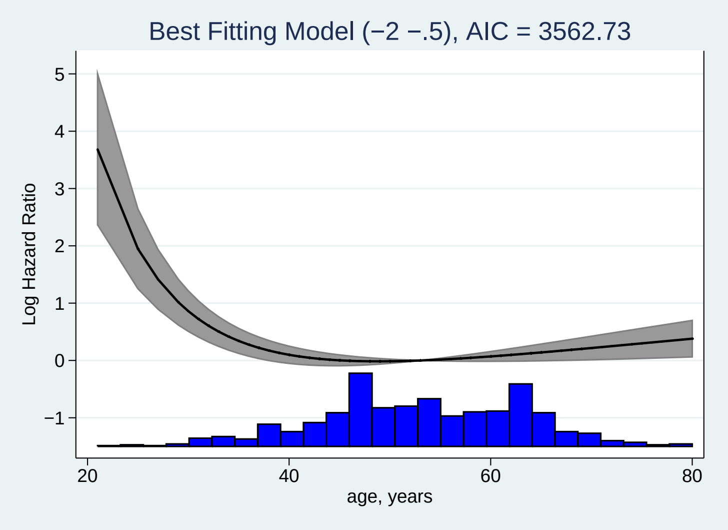
German Breast Cancer Study Group Data

- 686 women with primary node positive breast cancer (Sauerbrei and Royston, 1999).
- Time to recurrence or death (299 events).
- Covariates include,
  - Age (years)
  - Menopausal staus
  - Tumour Size (mm)
  - Tumour Grade
  - Number of positive lymph nodes
  - Progesterone Receptor (fmol)
  - Oestrogen Receptor (fmol)
  - Hormonal Therapy

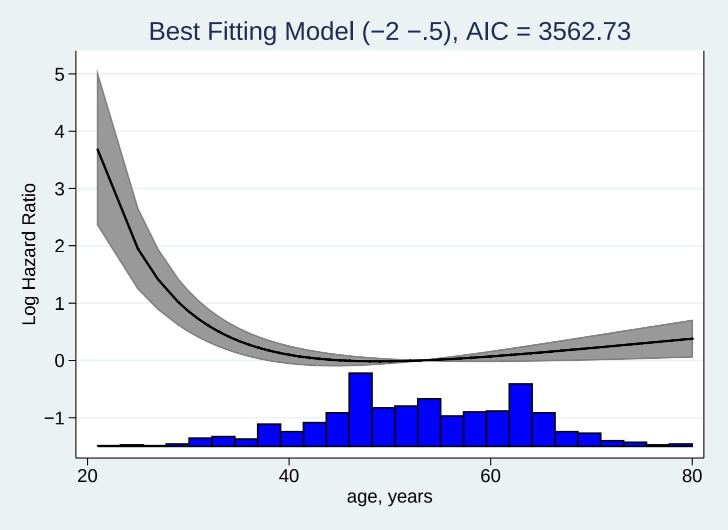
German Breast Cancer Study Group Data

- 686 women with primary node positive breast cancer (Sauerbrei and Royston, 1999).
- Time to recurrence or death (299 events).
- Covariates include,
    - Age (years)
    - Menopausal staus
    - Tumour Size (mm)
    - Tumour Grade
    - Number of positive lymph nodes
    - Progesterone Receptor (fmol)
    - Oestrogen Receptor (fmol)
    - Hormonal Therapy
- 5 covariates were selected using `mfp` command.

Breast Cancer - Best Fitting Model for Age



Breast Cancer - Best Fitting Model for Age

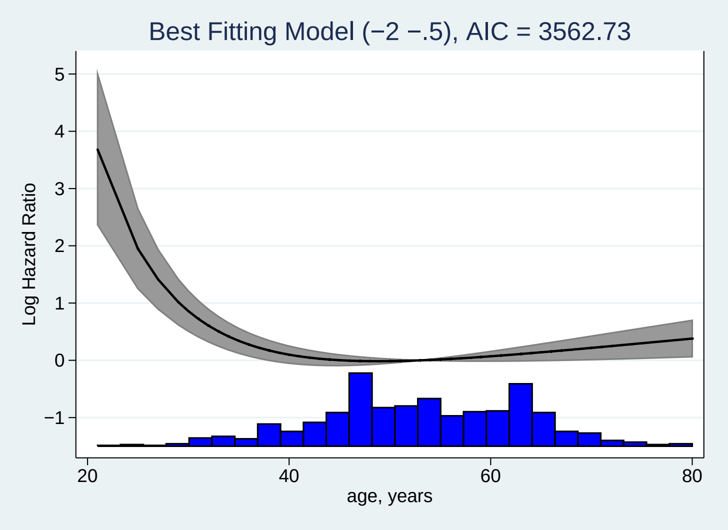


FP2(-2 -0.5): $\ln(h(t)) = \ln(h_0(t)) + \beta_1 Age_*^{-2} + \beta_2 Age_*^{-0.5}$

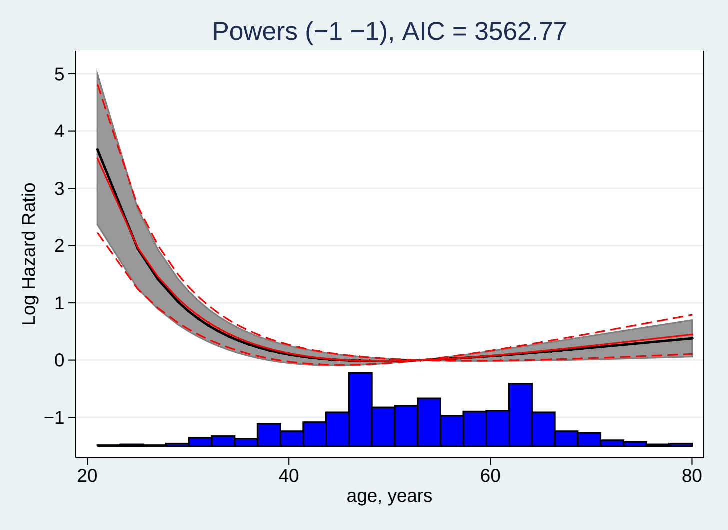
Breast Cancer - The 5 Best Fitting Model for Age

Powers	AIC
(-2,-0.5)	3562.73
(-1,-1)	3562.77
(-2,-1)	3562.78
(-2,0)	3562.83
(-2,0.5)	3563.05

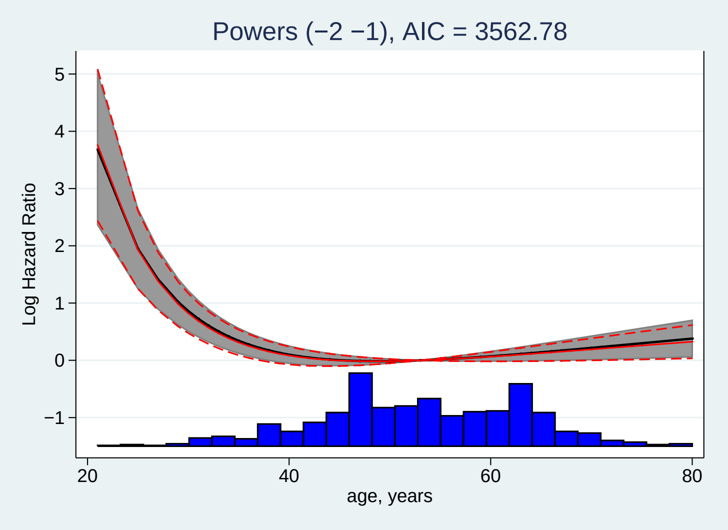
Breast Cancer - Age



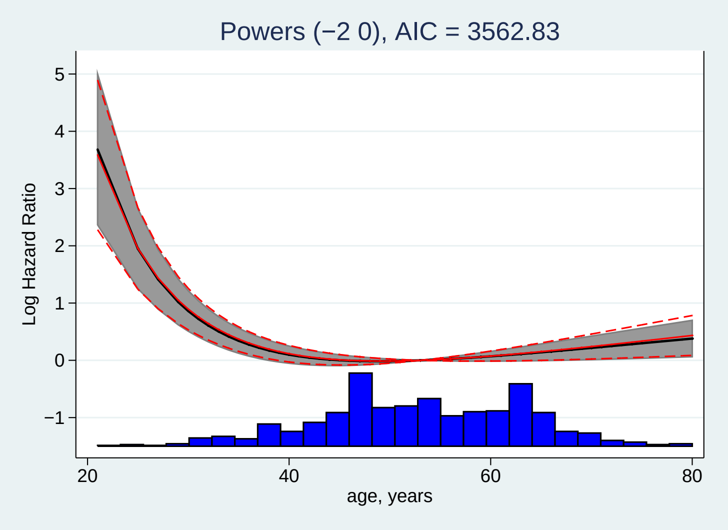
Breast Cancer - Age



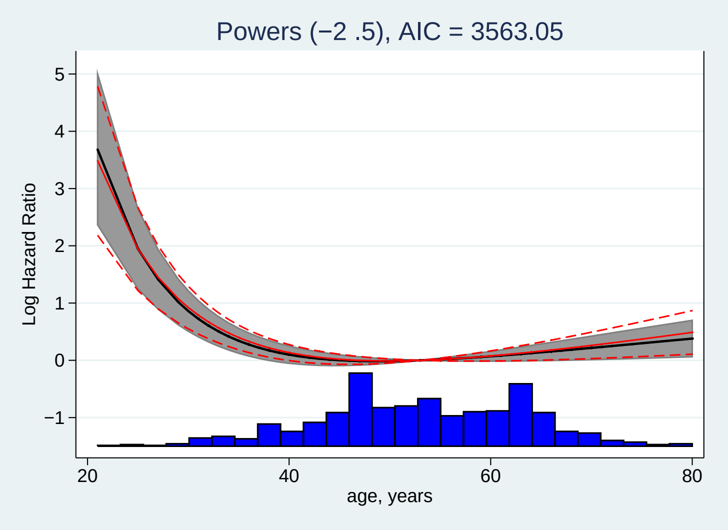
Breast Cancer - Age



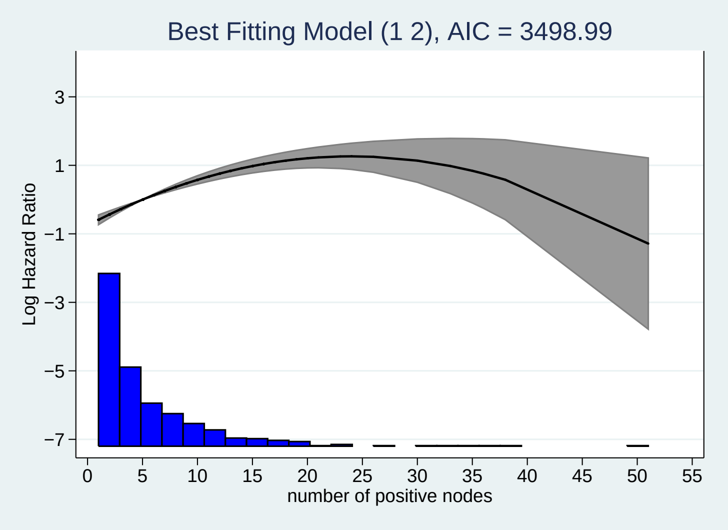
Breast Cancer - Age



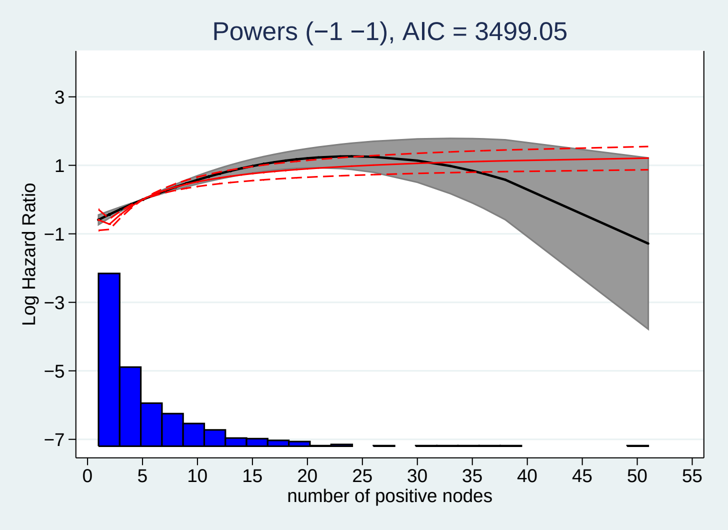
Breast Cancer - Age



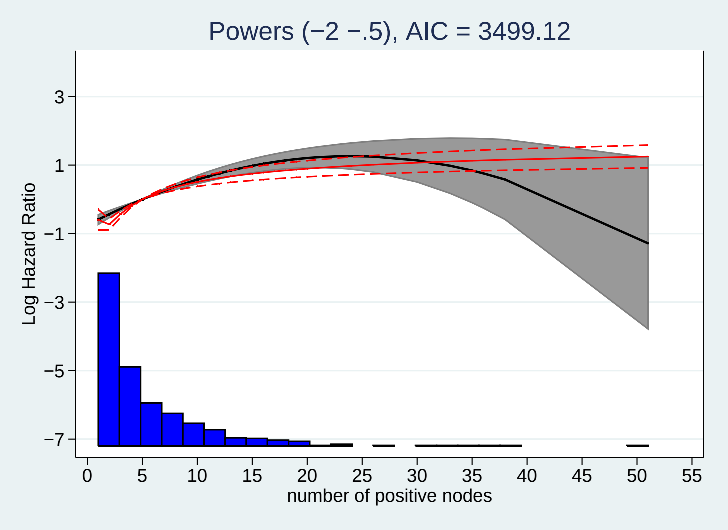
Breast Cancer - No. Positive Lymph Nodes



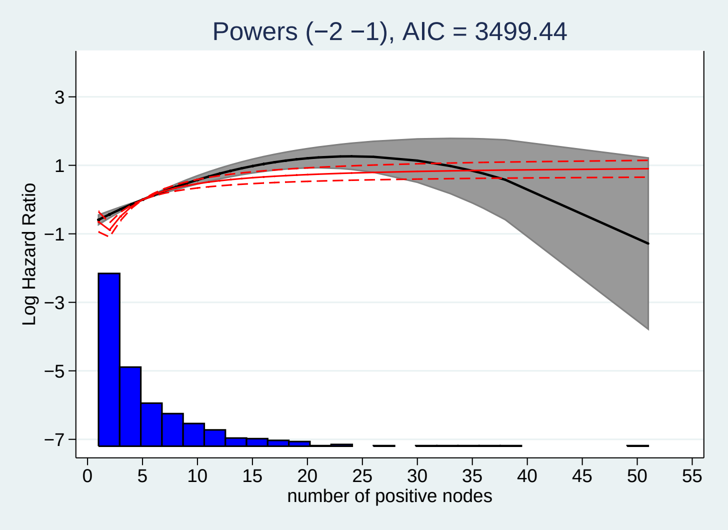
Breast Cancer - No. Positive Lymph Nodes



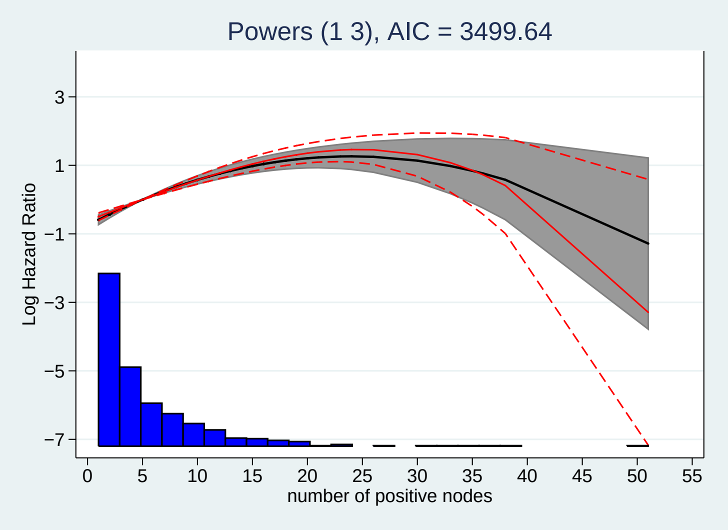
Breast Cancer - No. Positive Lymph Nodes



Breast Cancer - No. Positive Lymph Nodes



Breast Cancer - No. Positive Lymph Nodes



Model Averaging 1

- In FP models the model selection process is usually ignored when calculating fitted values and their associated confidence intervals.
- Model Averaging is popular Bayesian research area (Hoeting et al., 1999), (Congdon, 2007).
- Increasing interest from frequentist perspective (Burnham and Anderson, 2004) (Buckland et al., 2007) (Congdon, 2007) (Faes et al., 2007)
- Usually interest lies in model averaging for a parameter.
- Here we are interested in averaging over the functional form obtained from different models.

Model Averaging 2

- If there are K contending models, $M_k, k = 1, \ldots, K$ with weights, w_k, which are scaled so that $\sum w_k = 1$, then the estimate of a parameter or quantity, θ (assumed to be common to all models) is taken to be,

$$\widehat{\theta}_a = \sum_{k=1}^{K} w_k \widehat{\theta}_k$$

- The variance of $\widehat{\theta}_a$ is,

$$\text{var}\left(\widehat{\theta}_a\right) = \sum_{k=1}^{K} w_k^2 \left(\text{var}\left(\widehat{\theta}_k | M_k\right) + \left(\widehat{\theta}_k - \widehat{\theta}_a\right)^2\right)$$

Obtaining the Weights, w_k

- In a Bayesian context we want, $w_k = P(M_k|Data)$
- These probabilities are not trivial to calculate and various approximations are available.
- One such approximation is to use the Bayesian Information Criterion (BIC)

$$BIC_k = \ln(L_k) - \frac{1}{2} p \ln(n)$$

- The AIC can also be used to derive the model weights (Buckland et al., 2007)

$$AIC_k = \ln(L_k) - 2p$$

- Recently Faes used the AIC to derive model weights for fractional polynomial models (Faes et al., 2007).

Obtaining the Weights, w_k

- Let

$$\Delta_k = BIC_k - BIC_{min} \quad \text{or} \quad \Delta_k = AIC_k - AIC_{min}$$

- The weights, w_k, are then defined as,

$$w_k = \frac{\exp\left(\frac{1}{2}\Delta_k\right)}{\sum_{j=1}^{K} \exp\left(\frac{1}{2}\Delta_j\right)}$$

Using Bootstrapping to Obtain the Weights, w_k

- An alternative to using the AIC or BIC for the model weights, w_k, is to use bootstrapping (Holländer et al., 2006).
- For each bootstrap sample the best fitting fractional polynomial model is selected.
- The weights w_k, are simply obtained using the frequencies of the models selected over the B bootstrap samples.
- If comparing fractional polynomial models of different degrees then some selection process is needed. This is usually done by setting a value for α.

Using fpma

Using fpma

.

```
. fpma x1, ic(aic) xpredict: stcox x1
 Models Included (in order of weight)
```

	Powers		AIC	deltaAIC	weight	cum. weight
1	-2	-.5	3562.73	0.00	0.0802	0.0802
2	-1	-1	3562.77	0.03	0.0789	0.1591
3	-2	-1	3562.78	0.04	0.0785	0.2376
4	-2	0	3562.83	0.09	0.0766	0.3142
5	-2	.5	3563.05	0.31	0.0686	0.3827
6	-1	-.5	3563.05	0.32	0.0685	0.4512
7	-2	-2	3563.26	0.53	0.0616	0.5128
8	-2	1	3563.38	0.65	0.0580	0.5709
9	-1	0	3563.50	0.77	0.0546	0.6255
10	-.5	-.5	3563.52	0.79	0.0540	0.6795
(*output omitted*)						
43	2		3578.18	15.44	0.0000	1.0000
44	3		3578.32	15.58	0.0000	1.0000

- New variables created `xb_ma xb_ma_se xb_ma_lci xb_ma_uci`

Using `fpma` - Bootstrapping ($\alpha = 0.05$)

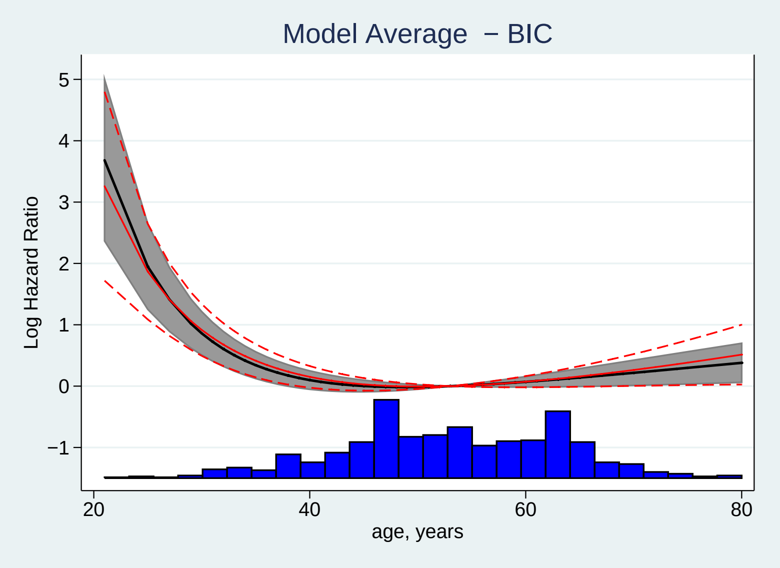
Using fpma with bootstrapping

.

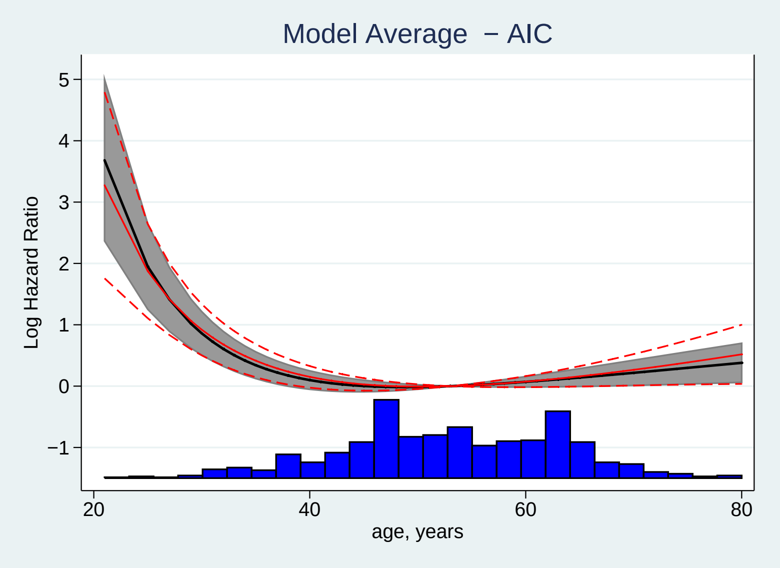
```
. fpma x1, ic(bootstrap) xpredict xpredname(x1_ma_boot1) reps(1000): stcox x1
Running 1000 bootstrap samples to determine model weights
(bootstrap: maboot)
```

	Powers		Freq.	weight	cum. weight
1	-2	-2	252	0.2520	0.2520
2	-2	-1	167	0.1670	0.4190
3	-1	-1	163	0.1630	0.5820
4	-1	-.5	88	0.0880	0.6700
5	1		71	0.0710	0.7410
6	-2		67	0.0670	0.8080
7	-.5	-.5	63	0.0630	0.8710
8	-2	-.5	51	0.0510	0.9220
9	-.5	0	30	0.0300	0.9520
10	0	0	19	0.0190	0.9710
11	0	.5	6	0.0060	0.9770
(output omitted)					
22	-1	0	1	0.0010	0.9990
23	-.5	.5	1	0.0010	1.0000

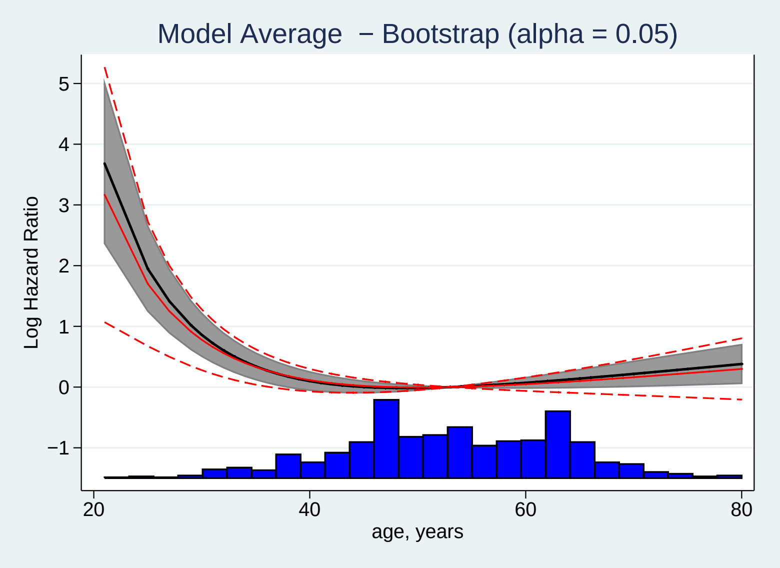
Breast Cancer - Age



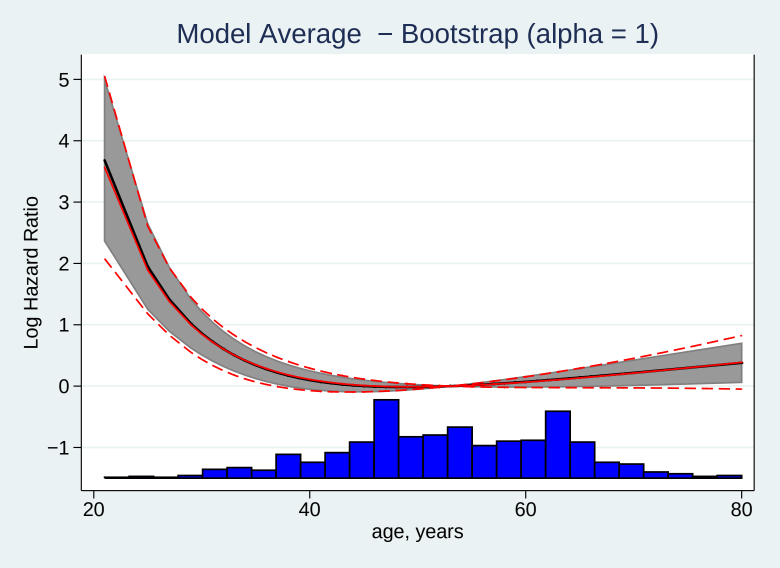
Breast Cancer - Age



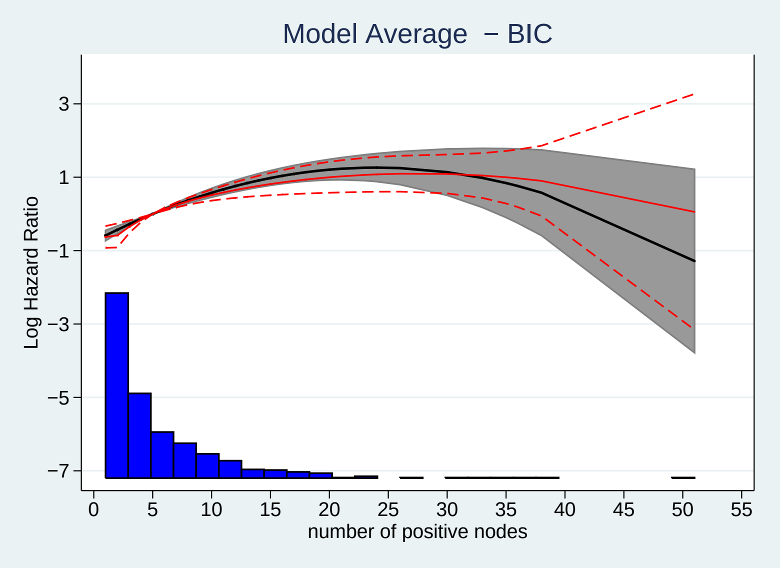
Breast Cancer - Age



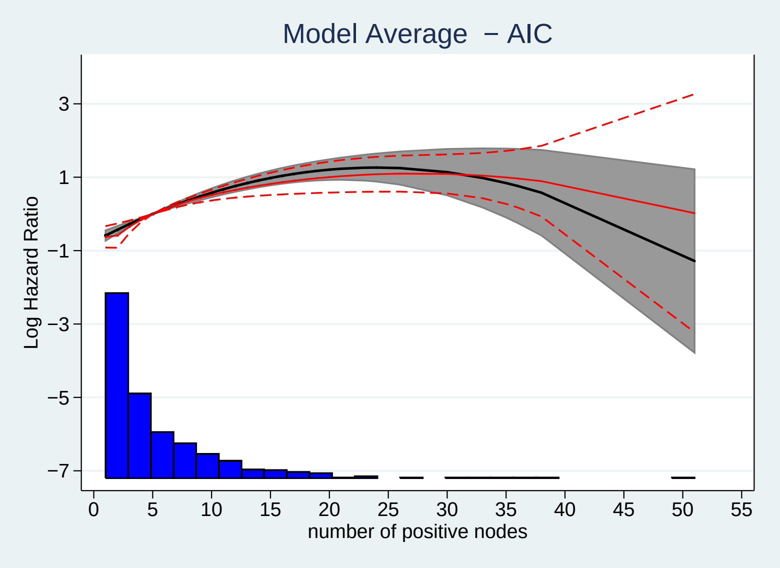
Breast Cancer - Age



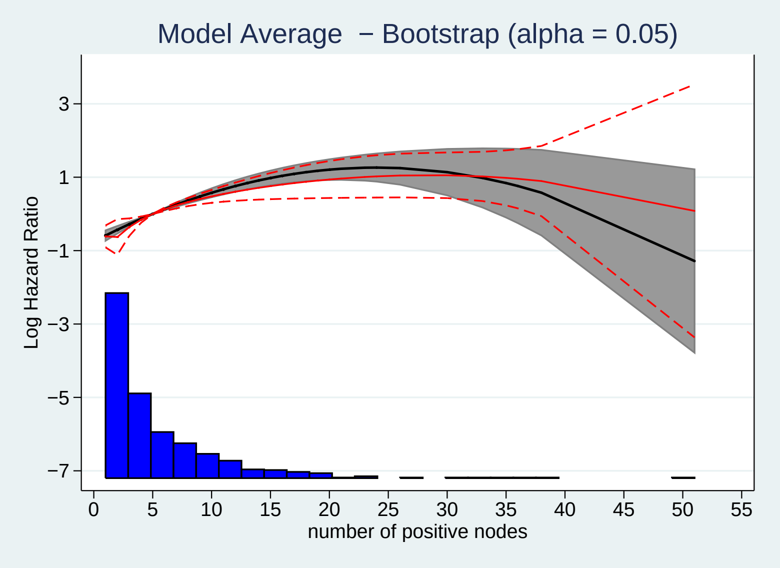
Breast Cancer - No. of Positive Lymph Nodes



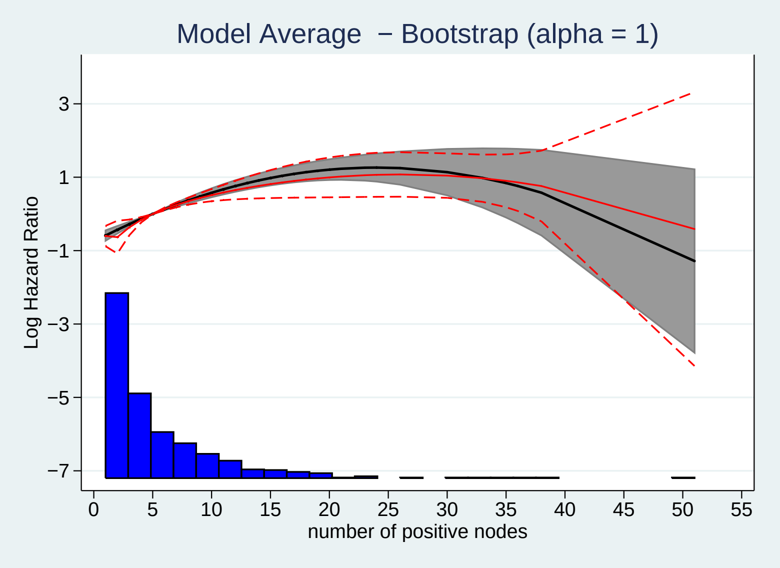
Breast Cancer - No. of Positive Lymph Nodes



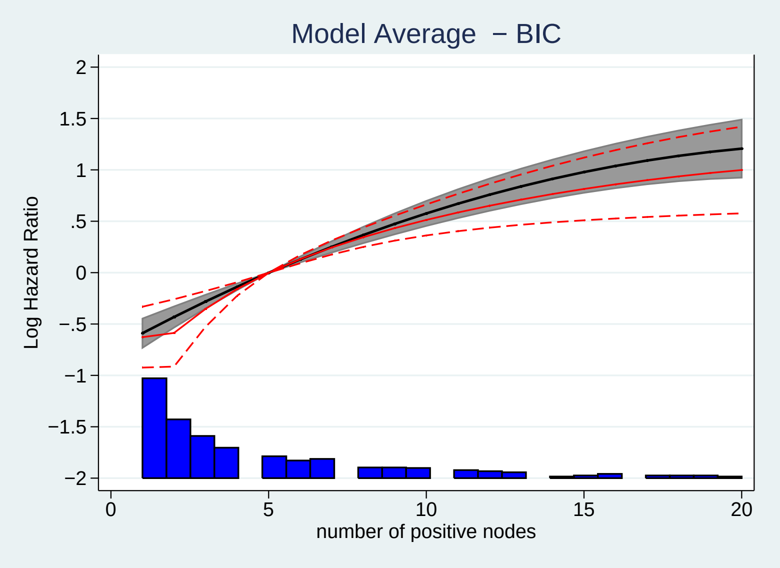
Breast Cancer - No. of Positive Lymph Nodes



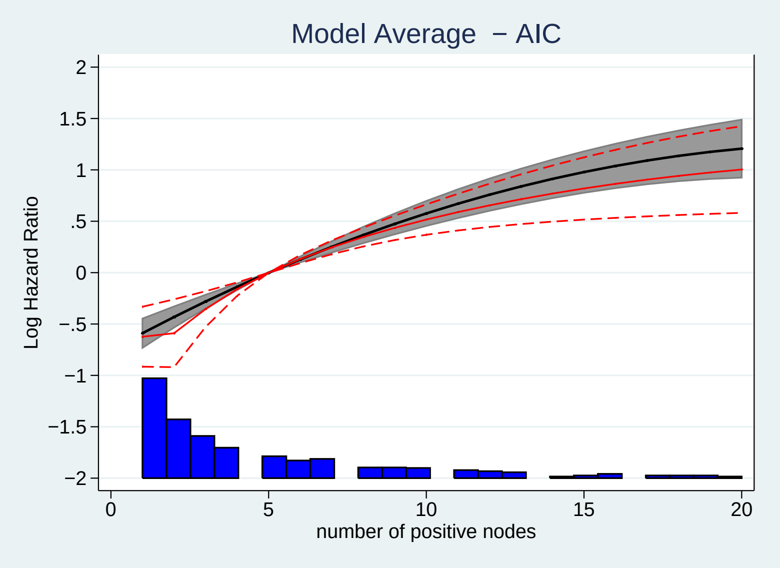
Breast Cancer - No. of Positive Lymph Nodes



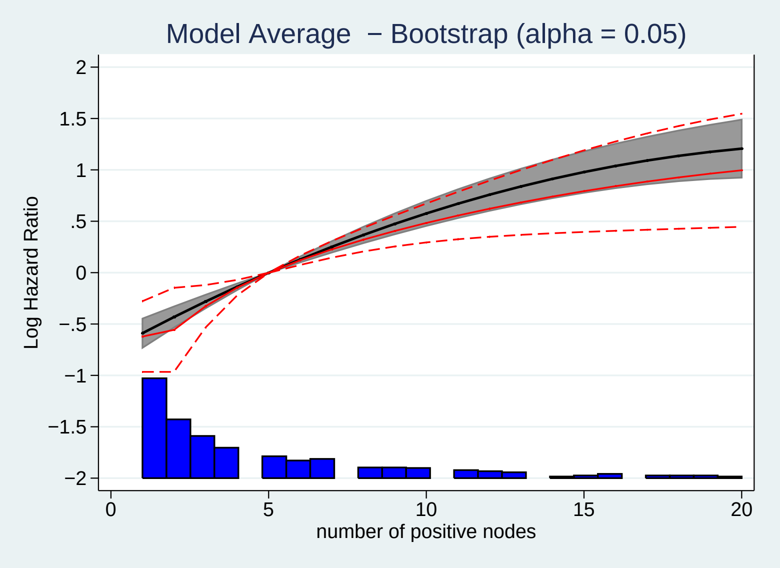
Breast Cancer - No. of Positive Lymph Nodes¡20



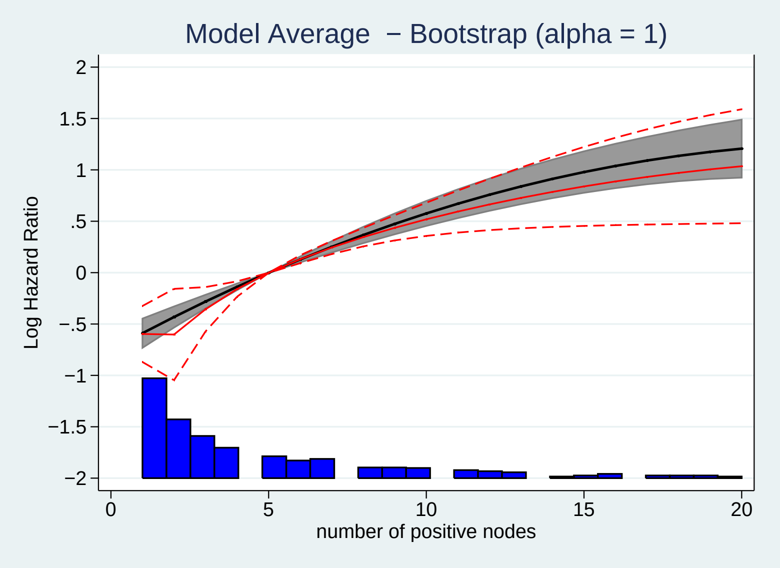
Breast Cancer - No. of Positive Lymph Nodes¡20



Breast Cancer - No. of Positive Lymph Nodes¡20



Breast Cancer - No. of Positive Lymph Nodes¡20



Multivariable Fractional Polynomials

- The above only really applies when using fractional polynomials for only one of the covariates in the model.
- However, is is common to use models with fractional polynomials for more than one covariate.
- A simple approach is to model average over various fractional polynomial models for the covariate of interest, while keeping the functional form of the remaining covariates constant.
- The `usemfp` option will do this for you.

Using mfp with Model Averaging

mfp

.

```
. mfp stcox x1 x2 x3 x4a x4b x5 x6 x7 hormon, nohr alpha(.05) select(0.05)
```

(*output omitted*)

```
Final multivariable fractional polynomial model for _t
```

Variable	df (Initial)	Select (Initial)	Alpha (Initial)	Status (Final)	df (Final)	Powers (Final)
x1	4	0.0500	0.0500	in	4	-2 -.5
x2	1	0.0500	0.0500	out	0	
x3	4	0.0500	0.0500	out	0	
x4a	1	0.0500	0.0500	in	1	1
x4b	1	0.0500	0.0500	out	0	
x5	4	0.0500	0.0500	in	4	-2 -1
x6	4	0.0500	0.0500	in	2	.5
x7	4	0.0500	0.0500	out	0	
hormon	1	0.0500	0.0500	in	1	1

```
Cox regression -- Breslow method for ties
```

(*output omitted*)

Using `fpma` after `mfp` - Age

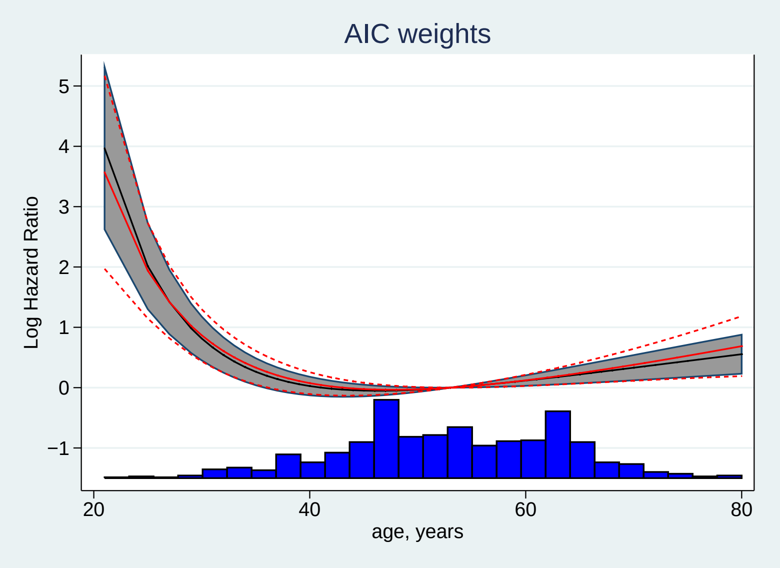
Using `fpma` after `mfp`

.

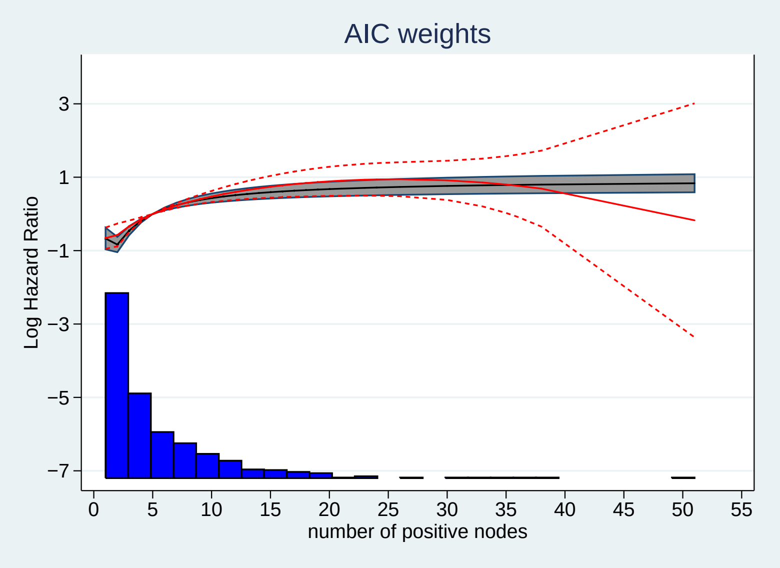
```
. fpma x1, ic(aic) xpredict: usemfp
 Models Included (in order of weight)
```

	Powers		AIC	deltaAIC	weight	cum. weight
1	-2	-.5	3434.72	0.00	0.0986	0.0986
2	-2	-1	3434.76	0.03	0.0969	0.1956
3	-1	-1	3434.85	0.13	0.0925	0.2881
4	-2	0	3434.89	0.17	0.0907	0.3788
5	-2	.5	3435.24	0.52	0.0760	0.4548
6	-1	-.5	3435.30	0.58	0.0740	0.5288
7	-2	-2	3435.43	0.70	0.0695	0.5982
8	-2	1	3435.75	1.03	0.0590	0.6572
9	-1	0	3435.96	1.24	0.0531	0.7103
10	-.5	-.5	3436.00	1.27	0.0522	0.7624
(*output omitted*)						
43	1		3452.04	17.31	0.0000	1.0000
44	.5		3452.05	17.32	0.0000	1.0000

Model Averaging after `mfp` - Age



Model Averaging after `mfp` - No. of Positive Lymph Nodes



Discussion

- Fractional Polynomials very useful for modelling non-linear functions.
- Model selection uncertainty is usually ignored after final model is obtained.
- Model averaging is easy to implement and incorporates FP model selection uncertainty.
- Still further work needed. For example,
  - Statistical properties (coverage etc).
  - Comparison with fully Bayesian model averaging.

References I

Buckland, S., Burnham, K., and Augustin, N. (2007). Model selection: An intergral part of inference. *Biometrics*, 53(2):603–618.

Burnham, K. P. and Anderson, D. R. (2004). Multimodal inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33(2):261–304.

Congdon, P. (2007). Model weights for model choice and averaging. *Statistical Methodology*, 4:143–157.

Faes, C., Aerts, M., H., G., and Molenberghs, G. (2007). Model averaging using fractional polynomials to estimate a safe level of exposure. *Risk Analysis*, 27(1):111–123.

Hoeting, J. A., Madigan, D., E., R. A., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–417.

Holländer, N., Augustin, N., and Sauerbrei, W. (2006). Investigation on the improvement of prediction by bootstrap model averaging. *Methods of Information in Medicine*, 45:44–50.

Royston, P. and Altman, D. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *JRSSA*, 43(3):429–467.

Sauerbrei, W. and Royston, P. (1999). Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *JRSSA*, 162(1):71–94.