Refined estimation of parametric models by functional approximation

Gareth Liu-Evans

December 15th, 2013

Abstract

A computational method for closed-form refinement in estimation of parametric models.

Presentation plan

- Introduction and the general idea
- Applications so far
- Prior beliefs
- Confidence Interval estimation (in progress)
- Planned applications and extensions
- Conclusions

The general idea

- Phillips and Yu (2005a,b, 2009a,b), Tang and Chen (2009), Wang, Phillips and Yu (2011), Ait-Sahalia and Kimmel (2007)

Given an original estimator $\hat{\Theta} = (\hat{\theta}, \hat{\Theta}_1)$ of $\Theta = (\theta, \Theta_1)$ consider

$$\tilde{\theta} = \hat{\theta} + G(\hat{\Theta}, n) \quad (1)$$

where $G : \Re^{d+1} \to \Re$ is to be chosen appropriately somehow, and unchanged across n. One may also consider a more general functional estimator:

$$\tilde{\theta}_n = \hat{\theta} + G_n(\hat{\Theta}, n) \quad (2)$$

where $G_n : \Re^{d+1} \to \Re$ is possibly different for each particular sample size.

Minimal requirements

(R1) $plim(\tilde{\theta}_n) = \theta$ under the same conditions as $\hat{\theta}$

(R2) If $n^s(\hat{\theta} - \theta) \xrightarrow{L} N(0, v)$ for some $s > 0$ and $0 < v < \infty$

then $n^s(\tilde{\theta}_n - \theta) \xrightarrow{L} N(0, v')$ where $v' \leq v$

Choice of G or G_n

Regularisation

$$G(\hat{\Theta}, n) = \sum_{j=1}^{r} \frac{1}{n^{s'j}} f_j(\hat{\Theta}) \tag{3}$$

$$G_n(\hat{\Theta}, n) = \sum_{j=1}^{r} \frac{1}{n^{s'j}} f_{j,n}(\hat{\Theta}) \tag{4}$$

where $s' > s$ and the functions f_j and $f_{j,n}$ are continuous for all j and n.

- In practice f_j and $f_{j,n}$ will be chosen according to a finite training set $\bar{C}_\Theta$ comprised of different parameterisations $\Theta \in C_\Theta$
- but this choice ensures consistency and equal convergence rate at all points including those outside the training set.

Consistency

Consistency requires

$$plim\left(\sum_{j=1}^{r}\frac{1}{n^{s'}j}f_{j,n}(\hat{\Theta})\right) = 0, \tag{5}$$

and the terms $plim f_{j,n}(\hat{\Theta})$ do not exist in general. But for all j and n,

$$\frac{f^L(\hat{\Theta})}{n^{s'}j} \leq \frac{f_{j,n}(\hat{\Theta})}{n^{s'}j} \leq \frac{f^U(\hat{\Theta})}{n^{s'}j}$$

where $f^L = \bigwedge_{n=1}^{\infty}\bigwedge_{j=1}^{r} f_{j,n}$ and $f^U = \bigvee_{n=1}^{\infty}\bigvee_{j=1}^{r} f_{j,n}$. In particular,

$$\left|\frac{f_{j,n}(\hat{\Theta})}{n^{s'}j}\right| \leq \max\left(\left|\frac{f^L(\hat{\Theta})}{n^{s'}j}\right|, \left|\frac{f^U(\hat{\Theta})}{n^{s'}j}\right|\right) \equiv \frac{1}{n^{s'}j}(|f^L| \vee |f^U|)(\hat{\Theta}) \tag{6}$$

If we require f^L and f^U to be continuous, then so is $\frac{1}{n^{s'}j}(|f^L| \vee |f^U|)(\hat{\Theta})$, and therefore, for any $\epsilon > 0$,

$$P\left[\left|\frac{1}{n^{s'}j} f_{j,n}(\hat{\Theta})\right| > \epsilon\right] \le P\left[\frac{1}{n^{s'}j}(|f^L| \vee |f^U|)(\hat{\Theta}) > \epsilon\right] \to 0 \tag{7}$$

which proves (5). Alternatively, if we also assume C_Θ is compact, the functions $f^L(\hat{\Theta})$ and $f^U(\hat{\Theta})$ are bounded, so $f_{j,n}(\hat{\Theta})$ are bounded and $\frac{1}{n^{s'}j} f_{j,n}(\hat{\Theta}) \to 0$ pointwise.

An idealised form of the objective

$$\underset{G_n}{inf} \; ||E|| \text{ s.t. } G_n \in S \tag{8}$$

$E =$ the set of values of some chosen functional of $\tilde{\theta}$

- where each value corresponds to a parameterisation $\Theta \in C_\Theta$
- and where the minimisation is subject to G_n having certain other properties in terms of $\tilde{\theta}$ and therefore belonging to some constraint set S

Examples of E

1. the set of all bias values $E[\tilde{\theta} - \theta]$ across C_Θ

2. the set of all RMSE values $\sqrt{E[(\tilde{\theta} - \theta)^2]}$ across C_Θ

3. the set of all expected values of an asymmetric loss function

The members G_n may correspond to estimators $\tilde{\theta}$ that do no worse than $\hat{\theta}$ at each $\Theta \in C_\Theta$ according to certain criteria e.g. bias and RMSE.

In practice:

- restrict G_n to a family parameterised by $w \in \Re^m$ so that choice of G_n becomes a choice of w.
- restrict C_Θ to a finite training set $\bar{C}_\Theta$.
- in the case where $\tilde{\theta}$ is considered rather than $\tilde{\theta}_n$, the training set is $\bar{C}_\Theta \times N$ where N is a finite set of sample sizes.

CIR/Feller square root model

The CIR model takes the following form:

$$dX(t) = \kappa(\alpha - X(t))dt + \sigma\sqrt{X(t)}dB(t) \tag{9}$$

where we require $2\kappa\alpha/\sigma^2 > 1$.

We consider estimation of κ, the mean reversion parameter, refining the Nowman estimator[1]. The training sets are

$$\bar{C}_{\kappa,\sigma,\alpha} = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.97\} \times \{0.05\} \times \{0.05\}$$
$$\bar{C}_{\kappa,\sigma,\alpha} \times N = \{0.1, 0.5, 0.97\} \times \{0.05\} \times \{0.05\} \times \{120, 500\}$$

[1] See e.g. Tang and Chen(2009), and Wang, Phillips and Yu (2011)

The chosen approximant: Pade(m_1,m_2)

- We set $r = 2$ and restrict $f_{1,n}$ and $f_{2,n}$ to Pade(10,10) rational approximants, and optimise over the $m = 40$ parameters to obtain $\tilde{\kappa}_n$.
- Set $s' = 1$, and consider $G_n(\hat{\kappa}, n) = \frac{1}{n} f_{1,n}(\hat{\kappa}) + \frac{1}{n^2} f_{2,n}(\hat{\kappa})$
- Similar for $\tilde{\kappa}$
- In general if a Pade(m_1, m_2) approximant is used, $f_{j,n}$ takes the following form where the n and j subscripts have been suppressed:

$$f(\hat{\theta}) = \frac{\sum_{i=1}^{m_1} \alpha_i \hat{\theta}^i}{1 + \sum_{i=1}^{m_2} \beta_i \hat{\theta}^{i-1}} \tag{10}$$

Pade approximants are known to have relatively good convergence properties. Used by Phillips P.C.B. (1983) for approximating pdfs.

$\kappa = 0.15,\ n = 36,\ h = 1/12$

		Nowman		Refined	
α	σ	$b(\hat{\kappa})$	$MSE(\hat{\kappa})$	$b(\tilde{\kappa})$	$MSE(\tilde{\kappa})$
0.01	0.01	1.955923	7.169248	-0.0009253894	3.804886
	0.05	1.94968	6.991571	-0.01248947	3.373257
	0.1	1.967965	7.248981	0.006235249	3.852657
0.05	0.01	1.965342	7.08418	-0.002129663	3.575042
	0.05	1.963529	7.209255	0.005000181	3.750329
	0.1	1.94966	7.090192	-0.007638642	3.676049
0.1	0.01	1.959761	8.080639	0.02671395	6.940536
	0.05	1.956805	6.97331	-0.006054655	3.430174
	0.1	1.957084	7.04932	-0.003045995	3.595293
0.2	0.01	1.946749	7.085481	-0.01015642	3.652696
	0.05	1.952635	7.106731	-0.01044673	3.534738
	0.1	1.950137	7.043505	-0.009663761	3.580629
0.3	0.01	1.9651	7.187168	0.01302755	5.64097
	0.05	1.959559	7.202771	0.0003169344	3.92183
	0.1	1.967364	7.144145	0.007461471	3.633588

$\kappa = 0.35,\ n = 36,\ h = 1/12$

		Nowman		Refined	
α	σ	$b(\hat{\kappa})$	$MSE(\hat{\kappa})$	$b(\tilde{\kappa})$	$MSE(\tilde{\kappa})$
0.01	0.01	1.983664	7.488337	0.008819736	4.011953
	0.05	1.989131	7.548144	0.01650167	4.346404
	0.1	1.980552	7.430053	0.004846383	4.087748
0.05	0.01	1.986126	7.407562	0.001889549	3.696335
	0.05	1.983739	7.464124	0.006651923	4.004538
	0.1	1.983811	7.514946	0.01065478	4.342401
0.1	0.01	2.004479	8.13131	0.0232761	6.442626
	0.05	1.985256	7.350423	0.003368224	3.711434
	0.1	1.983122	7.428735	0.00432952	3.982858
0.2	0.01	1.977339	7.45868	-0.003091943	3.903862
	0.05	1.980451	7.493577	0.001688011	4.067215
	0.1	1.997003	7.663253	0.01893795	4.18272
0.3	0.01	1.998676	7.656647	0.0267062	4.443887
	0.05	1.968926	7.393619	-0.007874103	3.893664
	0.1	2.000894	7.696298	0.02968754	4.340119

$\kappa = 0.55,\ n = 36,\ h = 1/12$

		Nowman		Refined	
α	σ	$b(\hat{\kappa})$	$MSE(\hat{\kappa})$	$b(\tilde{\kappa})$	$MSE(\tilde{\kappa})$
0.01	0.01	1.994771	7.83145	0.008819736	4.011953
	0.05	1.97313	7.72698	0.01650167	4.346404
	0.1	1.97855	7.825871	0.004846383	4.087748
0.05	0.01	2.004429	7.982552	0.001889549	3.696335
	0.05	1.981432	7.724879	0.006651923	4.004538
	0.1	1.988605	7.78989	0.01065478	4.342401
0.1	0.01	1.999134	8.185766	0.0232761	6.442626
	0.05	1.992435	7.78604	0.003368224	3.711434
	0.1	1.978867	7.801302	0.00432952	3.982858
0.2	0.01	1.978195	7.748837	-0.005949311	4.558871
	0.05	1.993419	7.788061	0.006466423	4.431221
	0.1	1.985322	7.990336	-0.006763512	6.340474
0.3	0.01	1.993156	7.806859	0.008332399	4.600078
	0.05	1.976505	7.584684	-0.01480571	4.15327
	0.1	1.971401	7.70486	-0.01226849	4.646064

$\kappa = 0.75,\ n = 36,\ h = 1/12$

		Nowman		Refined	
α	σ	$b(\hat{\kappa})$	$MSE(\hat{\kappa})$	$b(\tilde{\kappa})$	$MSE(\tilde{\kappa})$
0.01	0.01	1.976798	8.149878	-0.01676605	5.187773
	0.05	1.962864	8.0007	-0.02868564	5.683793
	0.1	1.948445	7.764359	-0.05087107	4.669562
0.05	0.01	1.968832	7.945917	-0.02326145	5.128554
	0.05	1.971753	8.051387	-0.0256602	4.927505
	0.1	1.963903	7.996199	-0.02740599	5.125489
0.1	0.01	1.986503	8.303805	-0.0005676753	9.148857
	0.05	1.971518	8.058796	-0.02124003	5.151163
	0.1	1.974364	8.060977	-0.01425365	5.299895
0.2	0.01	1.982436	8.144416	-0.008158088	5.287912
	0.05	1.962116	8.019692	-0.03205122	5.104205
	0.1	1.970349	8.075008	-0.02217264	5.272301
0.3	0.01	1.984953	8.075709	-0.01123994	4.970435
	0.05	1.984781	8.132679	-0.01137226	5.374882
	0.1	1.979274	8.260835	-0.008547272	5.614236

$\kappa = 0.935,\ n = 36,\ h = 1/12$

		Nowman		Refined	
α	σ	$b(\hat{\kappa})$	$MSE(\hat{\kappa})$	$b(\tilde{\kappa})$	$MSE(\tilde{\kappa})$
0.01	0.01	1.98552	8.475223	-0.004320048	7.01467
	0.05	1.968306	8.331661	-0.02744672	5.778012
	0.1	1.975928	8.45688	-0.01497658	5.982235
0.05	0.01	1.967622	8.497981	-0.03286456	5.977841
	0.05	1.97202	8.334162	-0.01984813	6.173778
	0.1	1.961013	8.190679	-0.03605932	5.607397
0.1	0.01	1.974811	8.409136	-0.01990532	5.85914
	0.05	1.976558	8.344273	-0.0226322	5.65616
	0.1	1.962007	8.209547	-0.04117223	5.331304
0.2	0.01	1.943206	8.284214	-0.05855973	6.605743
	0.05	1.974205	8.44561	-0.01258418	10.73915
	0.1	1.960668	8.272798	-0.03286031	5.802976
0.3	0.01	1.981484	8.518452	-0.01691518	7.076362
	0.05	1.973076	8.322509	-0.024432	5.558417
	0.1	1.952751	8.364731	-0.05249336	6.690607

Model 1, Bootstrap[2], Indirect Inference[3], FAR-n, FAR, FAR (RMSE)

$(\kappa, \alpha, \sigma^2) = (0.892, 0.09, 0.033)$

	B	I	FAR-n	FAR	FAR (RMSE)
$n = 120$					
$\%bias$	*0.178*	*2.677*	-0.1039791592	-0.0782726009	-30.4565919283
$RMSE$	*0.651*	*0.603*	0.8062292478	0.8141654623	0.3027610444
$n = 300$					
$\%bias$	*0.447*	*3.79*	0.0891370516	-0.1200183857	-2.4909966368
$RMSE$	*0.326*	*0.328*	0.3369998516	0.3614515182	0.2409639392
$n = 500$					
$\%bias$	*0.826*	*0.258*	0.1828207399	-0.4885730942	-0.1876699552
$RMSE$	*0.245*	*0.248*	0.2544836733	0.2593140567	0.2043264545

[2] Tang and Chen (2009). Results in italics also from here.

[3] Gourieroux and Montford (1993), Phillips and Yu (2005b)

Model 2, Bootstrap, Indirect Inference, FAR-n, FAR, FAR (RMSE)

$(\kappa, \alpha, \sigma^2) = (0.223, 0.09, 0.008)$

	B	I	FAR-n	FAR	FAR (RMSE)
$n = 120$					
%bias	*13.579*	*43.497*	-1.217073991	-1.2241278027	22.9708520179
RMSE	*0.502*	*0.495*	0.5958538411	0.5855602446	0.2747289391
$n = 300$					
%bias	*3.461*	*14.92*	0.4952165919	-5.9895863229	16.1722869955
RMSE	*0.226*	*0.208*	0.2419427825	0.2347528062	0.2351027222
$n = 500$					
%bias	*1.325*	*6.728*	-0.0329141794	-4.2331289238	9.9017679372
RMSE	*0.15*	*0.14*	0.1564493528	0.1420734669	0.1609439654

Model 3, Bootstrap, Indirect Inference, FAR-n, FAR, FAR (RMSE)

$(\kappa, \alpha, \sigma^2) = (0.148, 0.09, 0.005)$

	B	I	FAR-n	FAR	FAR (RMSE)
$n = 120$					
%bias	*39.597*	*19.17*	3.0938790541	0.0668109529	28.3788116592
RMSE	*0.507*	*0.484*	0.5773518858	0.5949663856	0.3141120978
$n = 300$					
%bias	*4.459*	*17.67*	1.2119135135	-3.4512286996	18.3980605381
RMSE	*0.214*	*0.209*	0.2367381887	0.2023811503	0.2397829018
$n = 500$					
%bias	*1.83*	*8.45*	-1.0885439189	-2.1390190583	11.3480829596
RMSE	*0.133*	*0.122*	0.1482933916	0.1157120132	0.1599322982

INARP(1) model

The INARP(1) model takes the following form:

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim Poisson(\lambda)$

where $\alpha \in [0, 1]$ and $\alpha \circ X = \sum_{i=1}^{X} Y_i$

Y_i is an $i.i.d$ sequence of random variables independent of X such that

$$P(Y_i = 1) = 1 - P(Y_i = 0) = \alpha$$

We consider estimation of α, the binomial thinning parameter, refining the CLS estimator[4] $\hat{\alpha}$.

$\bar{C}_{\alpha,\lambda} = \{0.1, 0.5, 0.97\} \times \{0.5, 4\}$. $f_{1,n}(\hat{\alpha})$ and $f_{2,n}(\hat{\alpha})$ Pade(10,10) again.

[4] See Al-Osh and Alzaid (1987).

INARP(1). CLS, FAR-n, $n = 100$

		CLS	$\tilde{\alpha}_n$
$\alpha = 0.2$	$\%bias$	-8.670360	-0.057438
	$RMSE$	0.103945	0.112290
$\alpha = 0.4$	$\%bias$	-6.185498	1.968260
	$RMSE$	0.101787	0.101540
$\alpha = 0.6$	$\%bias$	-4.949088	-0.449997
	$RMSE$	0.092264	0.090598
$\alpha = 0.8$	$\%bias$	-4.678163	0.258547
	$RMSE$	0.079087	0.074628
$\alpha = 0.99$	$\%bias$	-5.335047	-0.238859
	$RMSE$	0.069347	0.047539

Vasicek/OU model

The Vasicek model takes the following form:

$$dX(t) = \kappa(\alpha - X(t))dt + \sigma dB(t) \tag{11}$$

We consider estimation of κ, the mean reversion parameter, refining the CML estimator[5] $\hat{\kappa}$. The training set is

$$\bar{C}_{\kappa,\sigma,\alpha} = \{0.1, 0.5, 0.97\} \times \{0.05\} \times \{0.05\}$$

[5] See e.g. Tang and Chen(2009)

The chosen approximant: neural net

- Based on a subset of the model parameter estimates $\hat{\Theta}_S \subset \hat{\Theta}$, $f_{j,n}$ takes the following form:

$$f(\hat{\Theta}_S)= \sum_{i=1}^{m'} \beta_i F(w_i \cdot \hat{\Theta}_S + b_i)$$

$$\text{with} \quad F(v)= (1+e^{-v})^{-1}. \qquad (12)$$

- This is a "single hidden-layer feedforward neural network".[6]

- Our refined estimator is $\tilde{\kappa} = \hat{\kappa} + \frac{1}{n} f_{1,n} = \hat{\kappa} + \frac{1}{n} \sum_{i=1}^{m'} \beta_i (1 + e^{-w_i \hat{\kappa} + b_i})^{-1}$, where $m' = 10$ is chosen. Optimisation is over $m = 30$ parameters.

[6] Hornik, Stinchcombe and White (1989), Leshno, Lin, Pinkus and Schocken (1993)

Vasicek Models, $n = 500$. Bootstrap, Indirect Inference, FAR-n.

Models 1-3 $(\kappa, \alpha, \sigma^2) = (0.858, 0.0891, 0.00219)$, $(0.215, 0.0891, 0.0005)$, $(0.140, 0.0891, 0.0003)$

	CML	J^7	B	I	FAR-n
Model 1					
%bias	*12.88*	*0.586*	*0.073*	*0.72*	-0.2974424242
RMSE	*0.265*	*0.25*	*0.235*	*0.24*	0.237044827
Model 2					
%bias	*53.033*	*5.23*	*0.861*	*7.612*	-1.9507706977
RMSE	*0.189*	*0.171*	*0.147*	*0.14*	0.1540890976
Model 3					
%bias	*76.593*	*7.7*	*2*	*10.624*	-0.6403240714
RMSE	*0.17*	*0.159*	*0.147*	*0.116*	0.1417119614

Prior beliefs

Prior beliefs introduced by choice of weight function in the norm:

$$||E||_{C_\Theta} = \left(\int_{C_\Theta} w(x) E(x)^2 d\mu(x) \right)^{\frac{1}{2}}$$

where $w(\Theta) \propto p(\Theta)$, the prior joint density function. In practice given a training set $\bar{C}$ with K points:

$$||E||_{C_\Theta/\bar{C}} = \left(\sum_{i=1}^{K} w(\Theta_i) E(\Theta_i)^2 \right)^{\frac{1}{2}}$$

Example Vasicek beliefs, $Prob(\kappa_i \in (0, 0.3]) = 1$

$$w(\Theta_i) = 1 \text{ if } \kappa_i \in (0, 0.3]$$

$$w(\Theta_i) = 0 \text{ otherwise}$$

Two training sets are considered:

$$\bar{C}^{(1)}_{\kappa,\sigma,\alpha} = \{0.01, 0.3\} \times \{0.05\} \times \{0.05\}$$

$$\bar{C}^{(2)}_{\kappa,\sigma,\alpha} = \{0.001, 0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25, 0.275, 0.3\} \times \{0.05\} \times \{0.05\}$$

and using $w(\Theta_i) = 1$ for all training points in (0,0.3].

Vasicek Models, $n = 500$. Bootstrap, Indirect Inference, FAR-n, FARb-n, FARb-n (RMSE).
Models 2-3 $(\kappa, \alpha, \sigma^2) = (0.215, 0.0891, 0.0005), (0.140, 0.0891, 0.0003)$

	CML	J^8	B	I	FARb-n (1)	FARb-n (RMSE) (1)	(2)	(2)
Model 2								
$\%bias$	*53.033*	*5.23*	*0.861*	*7.612*	-2.397	-5.682	-2.907	-19.545
$RMSE$	*0.189*	*0.171*	*0.147*	*0.14*	0.162	0.080	0.147	0.065
Model 3								
$\%bias$	*76.593*	*7.7*	*2*	*10.624*	-4.905	13.886	-1.806	7.423
$RMSE$	*0.17*	*0.159*	*0.147*	*0.116*	0.148	0.084	0.164	0.049

- Alternative methods for incorporating beliefs

Planned applications and extensions

- Confidence Interval estimation
- Refining a simulation method e.g. indirect inference.
- Option price estimation following Phillips and Yu (2009)

Conclusions

- A new simulation/computational method, but yielding closed-form refinements
- Simple reusable formulas to correct existing estimators
- Type of refinement is flexible
- Prior beliefs can be utilised
- Has worked well so far.