

On the Evolution of the United Kingdom Price Distributions

BM Chu¹ KP Huynh² DT Jacho-Chávez³ O Kryvtsov²

¹Carleton University

²Bank of Canada

³Emory University

18 June 2018

10th ECB Workshop on Forecasting Techniques: Economic
Forecasting with Large Datasets

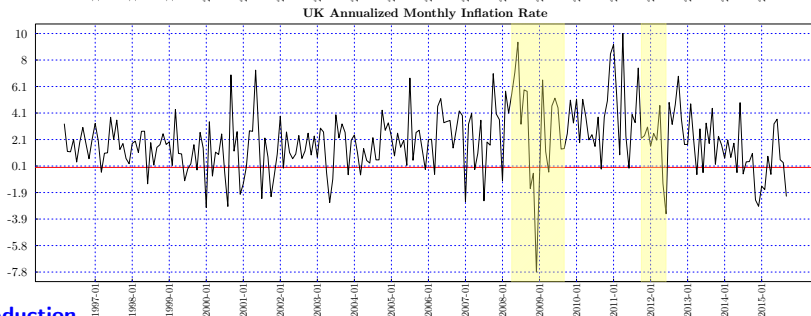
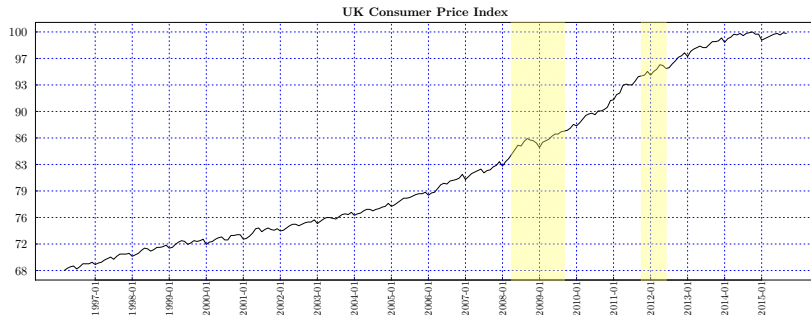
*The views expressed are those of the authors and should not be attributed to
the Bank of Canada.*

Contributions

- The paper provides a *distributional analysis* of a *publicly available monthly survey of prices* that a governmental statistical agency collects to construct the Consumer Price Index (CPI) for the United Kingdom.
- An adaptation of Kneip & Utikal's (2001, *JASA*) *Functional Principal Component Analysis* of density families is proposed utilizing the *Sampling Weights Kernel Density* estimator of Buskirk & Lohr (2007, *JSPI*) to take into account the complex survey nature of the data set.
- Develop an algorithm to conduct out-of-sample density forecasts.

forthcoming in the Annals of Applied Statistics (accepted on 20 April 2018).

UK Consumer Price Index & Inflation



JOURNAL
OF THE ROYAL STATISTICAL SOCIETY.

MARCH, 1924.

THE INTERRELATION AND DISTRIBUTION OF PRICES AND THEIR
INCIDENCE UPON PRICE STABILIZATION.

By NORMAN CRUMP.

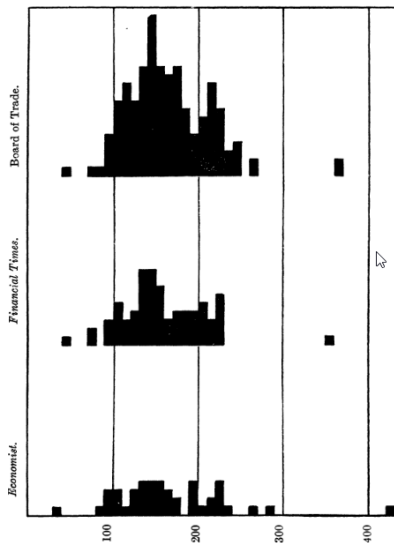
APPENDIX 5.—*Comparison of Four Index-Numbers.*

(1913 = 100.)

	Board of Trade.	<i>Financial Times.</i>	<i>Economist.</i>	Statistique Générale de la France.
Upper quartile (Q ₁)	194	200	195	479
Median (M)	160	152	156	386
Lower quartile (Q ₂)	132	130	129	329
Arithmetic mean (A)	164·7	158·6	166·3	421·2
Geometric mean (G)	157·9	150·9	154·8	390·7
Standard deviation (σ)	48·2	47·4	64·6	168·5
Coefficient of variation (V)	0·293	0·299	0·388	0·400
Angle of deviation (α)	17° 4'	17° 24'	22° 50'	23° 36'

Note.—Period covered: British figures, average prices for first half of 1923; French figures, end of September, 1923.

DISTRIBUTION OF BRITISH INDEX-NUMBERS.



Office for National Statistics (ONS)

- Data has been placed on the ONS website.
- Download the data and documentation.
- Thanks to the ONS for their assistance in this project.

The screenshot shows the ONS website interface. At the top, there is a navigation bar with the ONS logo and the text 'Office for National Statistics'. To the right, there are links for 'English (EN) | Cymraeg (CY)', 'Release calendar', 'Methodology', 'Media', 'About', and 'Blog'. Below this is a secondary navigation bar with categories: 'Home', 'Business, industry and trade', 'Economy', 'Employment and labour market', 'People, population and community', and 'Taking part in a survey?'. A search bar is present with the placeholder text 'Search for a keyword(s) or time series ID'. The main content area shows the breadcrumb 'Home > Economy > Inflation and price indices > Consumer price inflation item indices and price quotes'. The dataset title is 'Consumer price inflation item indices and price quotes'. Below the title, there is a contact information section with the ONS logo, 'Contact: Philip Gooding', 'Release date: 23 May 2018', and 'Next release: 13 June 2018'. The 'About this dataset' section explains that price quote data and item indices underpin consumer price inflation statistics. The 'Your download options' section lists four items: 'Item indices: April 2018' (with a 'zip (19.1 kB)' download button), 'Price quote: April 2018', 'item indices: March 2018', and 'Price quote: March 2018'. On the right side, there are three boxes: 'View all data related to inflation and price indices', 'Contact details for this dataset' (with contact information for Philip Gooding), and 'Publications that use this data' (with a link to 'Consumer price inflation, UK: April 2018').

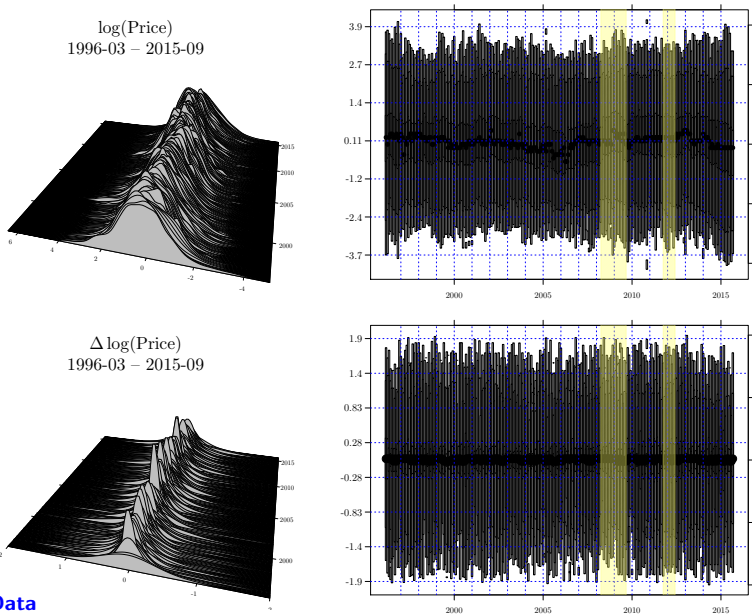
UK CPI

- Monthly collection, except for some services and seasonal items.
- 03/1996-09/2015 or 235 months.
- 110,000 units per month.
- 26 million observations.
- Stratified sampling by:
 - (1) shop type,
 - (2) region,
 - (3) shop type \times region.Shop type: multiple and independents (less than 10 outlets).

Figure 1: UK Regions



Figure 2: $\log Price - \Delta \log Price$, Demeaned & Standardized



Functional Principal Components

Karhunen-Loève decomposition by Kneip and Utikal (JASA, 2001):

$$f_t = f_\mu + \sum_{j=1}^J \theta_{t;j} g_j,$$

$f_\mu = \sum_{t=1}^T f_t / T$ is the common mean.

$\theta_{t;j}$ are the j -th components at time t .

g_j is the time-invariant profile for the j -th component.

Singular-value decomposition to $[M_{ts}]_{T \times T} = \langle f_t - f_\mu, f_s - f_\mu \rangle$:

$$\theta_{tr} = \lambda_r^{1/2} p_{tr}, \quad g_r = \sum_{t=1}^T \theta_{t;r} f_t / \sum_{t=1}^T \theta_{t;r}^2.$$

$$\sum_{t=1}^T \theta_{t;j} = 0, \quad \sum_{t=1}^T \theta_{t;j} \theta_{t;l} = 0 \text{ if } j \neq l, \quad \sum_{t=1}^T \theta_{t;j}^2 = \lambda_j, \quad j = 1, \dots, J.$$

$$f_t = f_\mu + \theta_{t;1} g_1,$$

FPCA with Complex Survey Data

Buskirk & Lohr (2007): Sample-Weighted Kernel Density

Finite population of size N is divided into L distinct strata with respective sizes N_1, \dots, N_L , i.e. $N = \sum_{k=1}^L N_k$.

The **strata density function**, $f(x) = \sum_{k=1}^L W_k f_k(x)$ where $W_k = N_k/N$, $k = 1, \dots, L$, can be estimated by

$$\hat{f}_{S_{N_n}}(x) = \frac{1}{wh} \sum_{i \in S_{N_n}} K\left(\frac{x - x_i}{h}\right) w_i$$

where S_{N_n} is any sample of size n taken from this strata density, $w_i = \pi_i^{-1}$ with $\pi_i = \Pr_D\{i \in S_{N_n}\}$ and $w = \sum_{i \in S_{N_n}} w_i$. For each i , w_i is called the unit's *sampling weight*.

FPCA with Complex Survey Data

Allow for weak-dependence in longitudinal data: T stratified samples,

$$\left\{ \left\{ (X_{it}, w_{it})^\top \right\}_{i=1}^{n_t} \right\}_{t=1}^T \Rightarrow \tilde{M}_{ts} = \langle \hat{f}_{S_t, h} - \hat{f}_{S, \mu}, \hat{f}_{S_s, h} - \hat{f}_{S, \mu} \rangle,$$

where $\hat{f}_{S, \mu} = T^{-1} \sum_{t=1}^T \hat{f}_{S_t, h}$, and $\langle \xi_1, \xi_2 \rangle = \int \xi_1(x) \xi_2(x) \varpi(x) dx$, for some ϖ continuous, uniformly bounded weight function.

Let \hat{M}_{ts} be the *biased corrected* version of \tilde{M}_{ts} with $\text{svd}(\tilde{M}_{ts}) \Rightarrow \hat{\lambda}_r$ and $\hat{\rho}_r$ of \hat{M} :

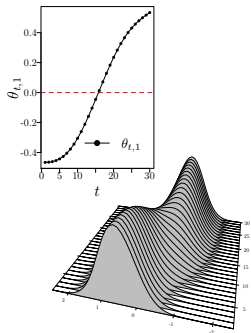
$$\hat{\theta}_{t,r} = \hat{\lambda}_r^{1/2} \hat{\rho}_{t,r}, \quad \hat{g}_r = \frac{\sum_{t=1}^T \hat{\theta}_{t,r} \hat{f}_{S_t, h}}{\sum_{t=1}^T \hat{\theta}_{t,r}^2}. \quad (1)$$

$$\sum_{t=1}^T \theta_{t,j} = 0, \quad \sum_{t=1}^T \theta_{t,j} \theta_{t,l} = 0 \text{ if } j \neq l, \quad \sum_{t=1}^T \theta_{t,j}^2 = \lambda_j, \quad j = 1, \dots, J, \quad (2)$$

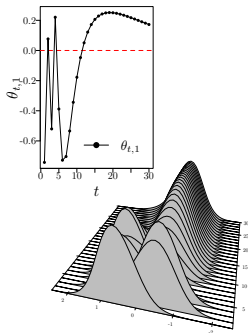
Computational Implementation

- ① Cross-validate 235 bandwidths for each cross-sectional density.
24 hours with 474 cores on EDITH using RSQLite and npRmpi.
- ② Numerical integrate $[\widehat{M}_{ts}]_{T \times T} = \langle \widehat{f}_t - \widehat{f}_\mu, \widehat{f}_s - \widehat{f}_\mu \rangle$
 $T \times (T - 1)/2$ or 27,495 integrals.
1 hour with 474 cores on EDITH using RSQLite and RSnow.
- ③ Singular value decomposition of $[\widehat{M}_{ts}]_{T \times T}$ to retrieve T -eigenvalues/vectors.
- ④ Scree plot (rank-order eigenvalues) and compute $\widehat{\theta}_{t;j}$ and \widehat{g}_j .
The first four components of $\log Price$ and $\Delta \log Price$ account for 5% and 7%, respectively. Compared to the rule-of-thumb of $1/T$ for significance.

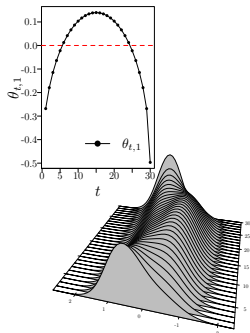
Monte Carlo Designs



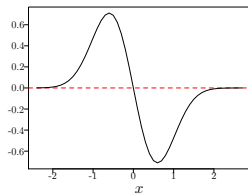
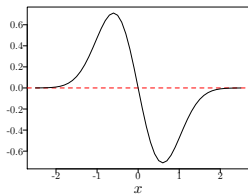
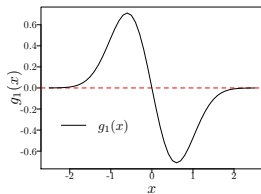
Scenario 1



Scenario 2



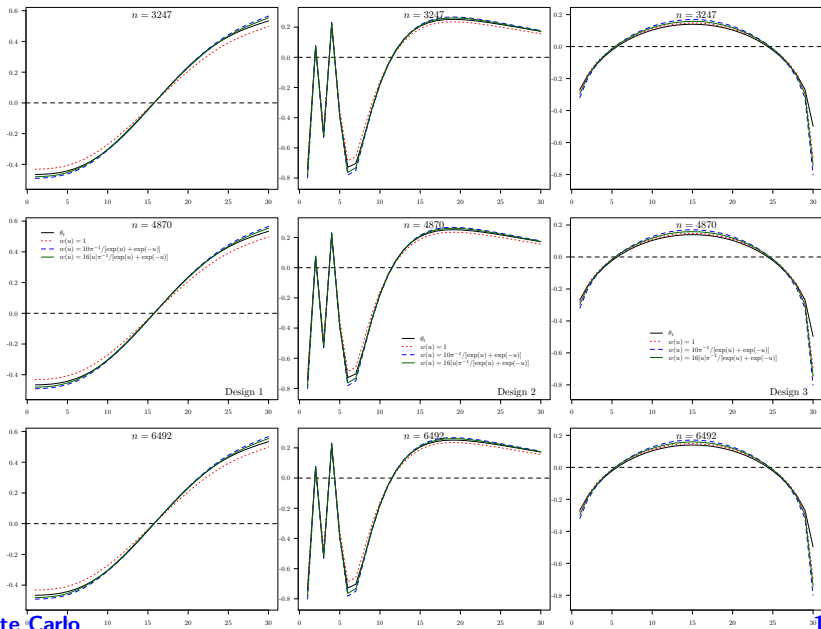
Scenario 3



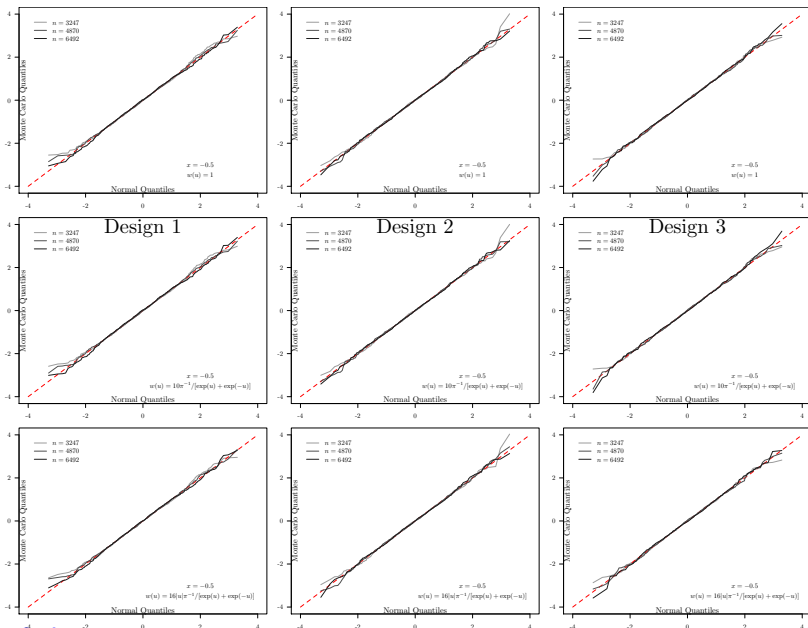
Monte Carlo Details

- 1 For each $t = 1, \dots, 30$, we set $N = 20,000$ distributed over $L = 6$ sub-populations, $N_1 = 2,223$, $N_2 = 4,445$, $N_3 = 6,666$, $N_4 = 2,219$, $N_5 = 2,223$, and $N_6 = 2,224$.
- 2 We draw samples $n_1 = c \times 667$, $n_2 = c \times 712$, $n_3 = c \times 1,000$, $n_4 = c \times 111$, $n_5 = c \times 556$, $n_6 = c \times 201$ ($n = c \times 3247$) from each f_t in the designs for $c \in \{1, 1.5, 2\}$ along with their inclusion probabilities.
- 3 In each of 1,000 Monte Carlo replications, Silverman's rule-of-thumb bandwidths and second-order gaussian kernels are used when calculating the 435 numerical integrals.
- 4 Three different weighting functions were used: $w(u) = 1$; $w(u) = 10\pi^{-1}/[\exp(u) + \exp(-u)]$; and $w(u) = 16|u|\pi^{-1}/[\exp(u) + \exp(-u)]$.

Monte Carlo: θ_t Performance



Monte Carlo: $g(x)$ Performance



$\Delta \log Price$: $\hat{\theta}_{t,1} \times \hat{g}_1$, $\hat{\theta}_{t,1}$, and UK Monthly inflation rate (blue)

Density Forecasting with FPCA

$$f_t = f_\mu + \sum_{j=1}^J \theta_{t,j} g_j,$$

Step 1. Using the first T^* estimated densities, $\{\widehat{f}_t\}_{t=1}^{T^*}$, calculate $\{\{\widehat{\theta}_{t,r}^*\}_{r=1}^{L^*}\}_{t=1}^{T^*}$ and $\{\widehat{g}_r^*\}_{r=1}^{L^*}$, where $L^* \leq T^*$ represents the number of the first non-zero eigenvalues of the $T^* \times T^*$ -matrix \widehat{M}^* . Also set $\widehat{f}_\mu^* = (1/T^*) \sum_{t=1}^{T^*} \widehat{f}_t$.

Step 2. Exploiting the orthonormal features of $\{\{\widehat{\theta}_{t,r}^*\}_{r=1}^{L^*}\}_{t=1}^{T^*}$, we utilize the algorithm in Hyndman & Khandakar (2008) to automatically identify the best-fitted ARMA model for each generated series, $\{\widehat{\theta}_{t,r}^*\}_{t=1}^{T^*}$, $r = 1, \dots, L^*$, and then proceed to obtain an automatic forecast for period $T^* + \ell$, i.e., $\{\widehat{\theta}_{T^*+\ell|T^*,r}^*\}_{r=1}^{L^*}$.

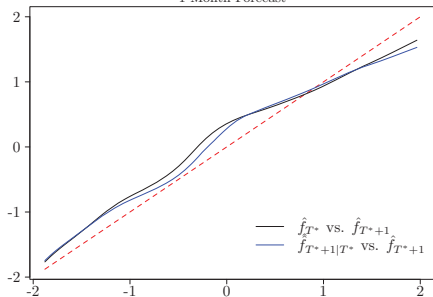
Density Forecasting with FPCA (cont.)

Step 3. Set

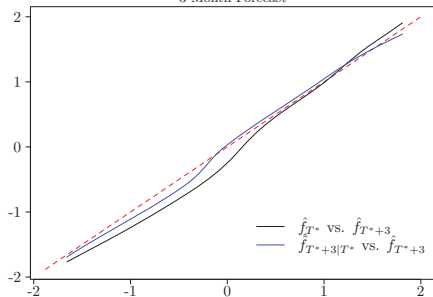
$$\hat{J} = \arg \min_{l \in \{1, \dots, L^*\}} \int \left(\hat{f}_{T^*+l}(x) - \hat{f}_\mu^*(x) - \sum_{r=1}^l \hat{\theta}_{T^*+l|T^*, r} \hat{g}_r^*(x) \right)^2 dx,$$
$$\hat{f}_{T^*+l|T^*} = \hat{f}_\mu^* + \sum_{r=1}^{\hat{J}} \hat{\theta}_{T^*+l|T^*, r} \hat{g}_r^*. \quad (3)$$

QQ Plots Forecast for $\Delta \log Price$

$\Delta \log(\text{Price})$ - July 2015
1-Month Forecast



$\Delta \log(\text{Price})$ - September 2015
3-Month Forecast



$\hat{J}=4$ with $\hat{\theta}_{t,1} \sim \text{SARMA}(0,0)(2,0)_{12}$, $\hat{\theta}_{t,2} \sim \text{SARMA}(1,1)(0,0)_{12}$,
 $\hat{\theta}_{t,3} \sim \text{SARMA}(0,2)(1,1)_{12}$, and $\hat{\theta}_{t,4} \sim \text{SARMA}(1,1)(0,0)_{12}$.

Summary

- Develop methodology to account for sampling weights in nonparametric estimation of FPCA.
- Demonstrate the efficacy of FPCA to visualize the dynamics of cross-sectional distributions.
- Application to UK Consumer Price Distributions via ONS.
- Conduct an out-of-sample forecasting exercise.

Thanks/Merci/Danke!