Team led by Fulvia Mecatti, Professor of Statistics at University of Milano-Bicocca
Dr. Gaia Bertarelli, Post-doc at University of Pisa
Dr. Maeregu Woldeyes Arisido, Post-doc at University of Milano-Bicocca

Final Report

Background

With the purpose to produce district-wise TB prevalence estimates for Pakistan, based on the 2010-2011 TB prevalence survey and other NTP data shared by the Pakistan NTP, we applied an advanced state-of-the-art methodology, dubbed SAE-LM, which integrates model-based *Small Area Estimation* (SAE) and *Latent Markov* (LM) modeling.

SAE applies since 2010-2011 Pakistan survey was not powered to estimate prevalent cases at district level. Instead, the survey was designed to produce a national estimate with controlled precision (20% max error with 95% probability) upon cluster-level data, precisely a sample of 133,000 adult individuals from 95 thensils (sub-district areas) selected according to the RedBook guidelines. Consequently, the sample size for districs is not under control: it is random, possibly too small to estimate a very small number, and limited to 68 districts which either intersect or include at least one sampled thensil, otherwise no district sample data are available.

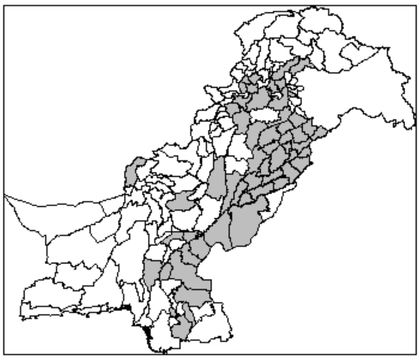


Figure 1: Sampled (gray) and non-sampled districts (white)

Direct district estimates, computed on available district-wise sample data, may or may not be accurate, as measured for instance by Coefficient of Variation (CV) possibly greater than a conventional acceptable threshold. Figure 2 shows direct district estimates (left panel) based on available district samples, and the distribution of their CVs (right panel) classified according to official statistics guidelines (for instance Statistics Canada 2016, p.35): sufficiently accurate (CV $\leqslant$ 16.6%), must be released with caveats because of large sampling error (CV between 16.6% and 33.3%), unreliable and release restrictions apply (CV $\geqslant$ 33.3%). No direct estimate exists for districts with zero sample size.

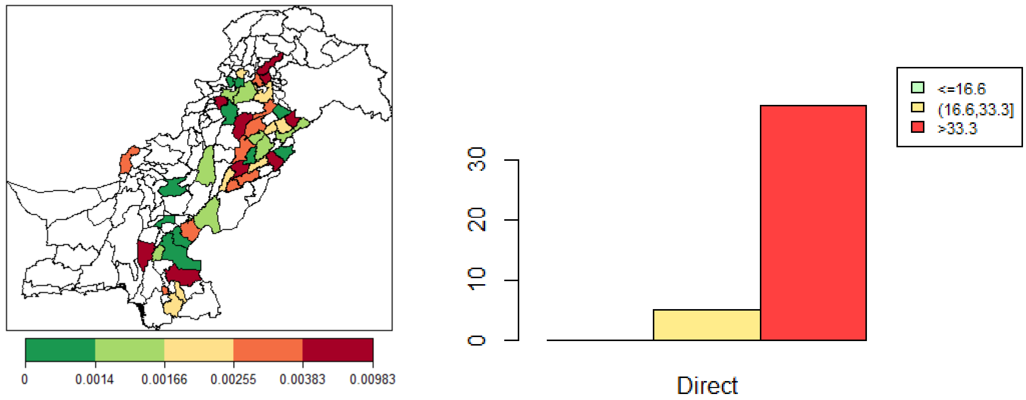
The basic idea of model-based SAE is to borrow straight from available out-of-sample auxiliary information and from related sources to form *indirect* estimates, either with reduced sampling error than the corresponding direct estimate or as a *prediction* of prevalence for non-sampled districts.

Table 1 lists all auxiliary variables available for at least 96 districts after check, editing and matching of data shared by the Pakistan NTP. Figure 3 shows the collection of 96 districs with complete auxiliary variables available to be used as covariates for our analysis, which includes 47 sampled districs (with computed direct estimates) and 49 non-sampled districts (with zero sample size thus no direct estimate).

Brief description of our SAE-LM model

A SAE model has two components: 1) the *sampling model*, essentially a model for the sampling error, concerns the probability distribution of the direct estimators conditioned on the true values of the district prevalence; and 2) the *linking model* which relates district prevalence (to be estimated) with the available covariates.

Our SAE-LM method uses a Latent Markov (LM) model as linking model.



(a) (b)

Figure 2: Direct estimates for sampled districts (a) and their CVs (b)

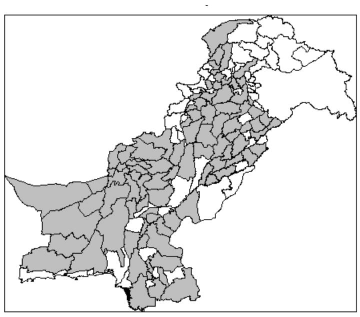


Figure 3: 96 Districts (grey) with complete covariates available for the analysis

LM models have been developed in the 70's for the analysis of longitudinal data related to a *latent* response variable, that is a variable of interest that can be measured only indirectly (for a comprehensive account see Bartolucci et al. 2013, *Latent Markov Models for Longitudinal Data* CRC Press.).

The rational of our SAE-LM methodology can be summarized in the following 3 key concepts:

1. we want to fully exploit all available data, both cross-sectional sample data form the 2011 prevalence survey, and longitudinal out-of-sample data available at district level for the time period 2011-2016;

2. we look at the true values of district prevalence as a latent response variable, partially and indirectly measured (via sample data and longitudinal covariates) at successive time points, *i.e.* an underlying *latent process*;

3. of such latent process we model both the distribution in space and its evolution in time, by means of a Markov chain, that is a popular probability distribution for a process.

At its turn a LM model comprises two parts: 1) the *measurement model* which concerns the probability distribution of the true values of district prevalence given the measurable part of the underlying latent process, *i.e.* the available covariates; and 2) the *latent model* which concerns the probability distribution of the latent process itself, that is the mentioned Markov chain .

The approach is *Hierarchical Bayes* (borrowed from time series analysis). The SAE sampling model is introduced at the top level of hierarchy and uses direct distric-wise estimates as initial input. At the lower level the LM model is introduced as linking model, split in its two components. Available covariates are included in the measurement model, in order to enforce prediction and improved indirect estimates. Finally, the latent model will catch all the residual un-observed and un-explained heterogeneity. Consequently, the set of model parameters to be estimated comprises three groups: *small area, measurement* and *latent* parameters. Essential details of SAE-LM model specification are given at the end of this report. For complete details see Bertarelli et al. 2018, Small area estimation for unemployment using latent Markov models, *Survey Methodology*, 44, 167-1928.

Variable	Nature	Period
Notified Bacteriological positive cases	total # and proportion	2011 to 2016
Notified Bacteriological negative cases	total #	2011 to 2016
Notified Bacteriological extrapulmonary cases	total #	2011 to 2016
Notified Grand Total positive, negative and other cases	total #	2017
Households	total #	2017
Male population	total #	2017
Female population	total #	2017
Transgender population	total #	2017
All sex Population	total #	2017
Total sex ratio	ratio	2017
Population	total #	2017
Overall growth rate	ratio	2017
Urban area households	total #	2017
Rural area households	total #	2017
Urban male population	total #	2017
Rural male population	total #	2017
Urban female population	total #	2017
Rural female population	total #	2017
Urban transgender population	total #	2017
Rural transgender population	total #	2017
All sex urban population	total #	2017
All sex rural population	total #	2017
Urban sex ratio	ratio	2017
Rural sex ratio	ratio	2017
Rural growth rate	ratio	2017
Urban growth rate	ratio	2017
Notified HIV cases	total #	2011 to 2018
Facilities	total #	2013 to 2016
Slides used	total #	2011 to 2016
Total error slides committed	total #	2011 to 2016

Table 1: Auxiliary variables available for at least 96 districts

Our SAE-LM model has been fitted by means of a Computational Statistics Algorithm, under a Bayesian approach to estimation of model parameters consistent with the model hierarchy. The fitting algorithm is a Data Augmentation Markov Chain Monte Carlo (MCMC), based on a Gibbs sampler, which requires to run Monte Carlo simulations and to discard a burn-in period.

Model selection, as it is usual the case for complex statistical models with several parameters, is defined in terms of an appropriate information criterion, *i.e.* a mechanism that uses data to give each candidate model a certain score, usually based on (some version of) maximum likelihood. Our best model has been selected by means of both a marginal likelihood measure of goodness of fit (the Chib's estimator, the most reliable for SAE-LM model selection), and the Bayesian Information Criterion (BIC) for hidden markov models (R-package *depmixS4*), which had to be in agreement.

The 68 available direct district-wise estimates entered the sampling error model as initial values of the small area parameters (top level of the of SAE-LM model hierarchy). For non sampled districts, thus with no direct estimates, missing values imputation has been applied, still by means of a Gibbs sampler directly upon the MCMC (posterior conditional) distribution.

Covariates entering the measurement model (intermediate level of hierarchy) underwent a careful selection from the collection of auxiliary variable in Table 1. First, hightly correlated auxiliary variables have been detected and reduced. Second, step-wise regression has been applied to identify sub-sets of relevant covariates.

Model fitting was completed by a grid search to select the value for the main parameter of the probibility distribution of the latent model (bottom level of hierarchy), i.e. the number k of states for the Markov Chain.

Our best model has been selected after running a MCMC simulation for combinations of sets of covariates and k values, each with 60,000 iterations on top of a burn-in of 30,000 iterations.

Limits and Strenghts of our SAE-LM method

SAE-LM methodology has the main limit to be resource consuming. The model fitting and selection algorithm requires adequate statistical skills, computational power and time.

It is also important to emphasise that even the most elaborated model cannot produce sufficiently accurate indirect estimates when no sample data and no covariates with good prediction power are available. Our analysis is limited to 96 Districts for which a complete set of covariates was available for the period 2011-2016 (including constant census 2017 variables). 24 districts (mostly concentrated in the north-eastern area of the country) were not associated to a complete set of covariates (and often not sampled too). Some sort of extrapolation/imputation for such missing districts might be considered, on the basis of our model results and under the assumption that all model covariates have sufficiently similar spatial distribution across districts either with or without known complete covariates. Such an assumtpion is rather restrictive, though, and for this reason we preferred to limit our analysis by excluding such missing districts.

In addition, we reckon that results and discussion presented in the following may suffer of our limited knowledge of Pakistan, its geo-administration as well as of NTP, TB epidemiology and risk factors. We cannot esclude misunderstandings and mislookings in the pre-processing of data, editing and matching. On the other hand, all the trajectories of MCMC estimation of model parameters (trace plots available on demand) show high stability after the 60,000 runs. This suggests that no significant differences should be observed were our SAE-LM method applied to a better pre-processed dataset.

A potential limit could be the Bayesian approach to estimation, chosen over the more familiar frequentist, fixed parameter paradigm. However, in such a choice there are some advantages, both methodological and practical. Two clear ones are the prediction of prevalent cases for all districts, including non-sampled districts, via point indirect estimation, and the assessment of the prediction accuracy is automatically solved by means of uncertainty intervals, directly upon the MCMC (full posterior) probability distribution provided by the fitting.

A major asset of SAE-LM metodology is its quite reach output, that comprises indicators and predictors beyond the mere district-wise decomposition of the national estimate, as a result of the estimation of all three groups of model parameters. Moreover, despite being a complex advanced statistical model, SAE-LM output is easily readable and ready-to-use.

Finally, with respect to this specific analysis, we think that both the validity of our modelling and the reliability of results are supported by the good performance of all the diagnostics provided in this report as well as further validation results available on demand.

SAE-LM Results

Primary output are district-wise TB prevalence (indirect) estimates (Bacteriologically positive among adults $\geqslant 15$ years) (Fig. 4).

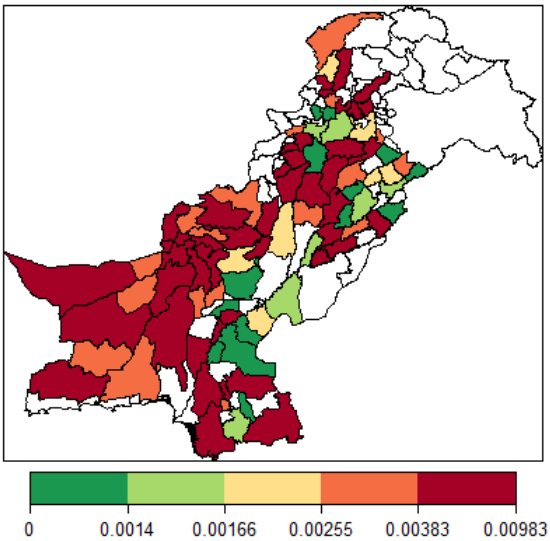


Figure 4: District-wise TB prevalence (indirect) estimates from 2011 survey

The map provided illustrates the decomposition district-wise of the national estimate offered by the 2010-2011 survey. Notice that SAE-LM, for combining both cross-sectional survey data and longitudinal covariates, in fact provides indirect district estimates for every time point, thus showing the evolution in time of TB prevalence, besides its spatial distribution at district level. The list of all indirect district estimates for the entire period 2011-2016 is provided at the end of this report. They show little differences along the 7-years period, which suggests a slow evolution in time of district-wise TB burden distribution.

Further output of SAE-LM methodology is the spatial classification of Pakistan districts into three classes of increasing TB burden (the $k = 3$ latent states of the selected best model). It is still a map, where homogeneous districts are represented by the same colour and intra-area variability is highlighted by different colours. This district-wise mapping shows the residual areal variability of TB prevalence in the country, *i.e.* all heterogeneity not caught by the observed covariates.

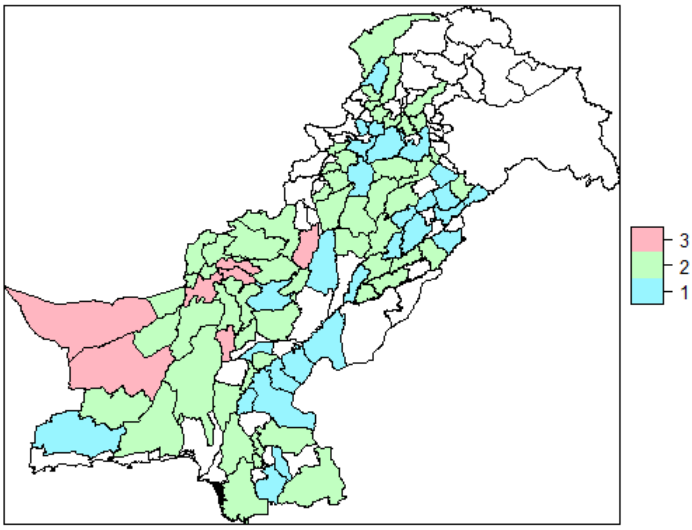


Figure 5: (Latent) Classification of districts in 3 classes of increasing TB burden

Notice that if districts were ordered with respect to increasing values of estimated prevalence (as eported in Figure 4), the result would widely reflect the classification above, though not entirely. This suggests that some relevant determinant(s) of the district-distribution of TB burden remain(s) un-measured, not included into the covariates entering the measurement model, and hence it is caught by the latent model as residual heterogeneity.

Finally, estimated latent parameters offer powerful prediction tools.

A vector of $k = 3$ estimated *initial probabilities*

$$(0.322, 0.587, 0.091)$$

informs us about the overall average probability, for a given district, to be classified into each one of the three increasing levels of TB burden at time 2011 of national survey data.

A matrix of *transition probabilities*

$$\begin{pmatrix} 0.975 & 0.012 & 0.014 \\ 0.005 & 0.984 & 0.010 \\ 0.021 & 0.033 & 0.946 \end{pmatrix}.$$

for any district classified in a given class of TB burden (by pointing a given row of the matrix) will predict the chance for that area to change class in the future, that is to either have improved (moving backward across the matrix columns) or worsened (moving forward). The matrix clearly shows a general stability (the largest probabilities, close to unity, lay along the diagonal) suggesting, again, a slow evolution in time of the district-wise distribuiton of TB prevalence. Note also that the probabilities to improve in the future (moving backward) are greater for districts classified as highest risk (third row), almost doubled than the probability to worsen in the future (moving forward) for districts classified as lowest risk (first row).

SAE-LM model Validation

There are basic properties that we require for the district-wise indirect estimates provided by our SAE-LM methodology. First the most part should have acceptable coefficient of variation and, where a corresponding direct estimate exists, they should provide precision gains upon smaller CV. Figure 6 shows CVs of all our district-wise indirect estimates (provided in Figure 4). CVs of direct (left) and indirect (right) estimates are compared (limited to sampled districts) in Figure 7 with an appreciable improvement.

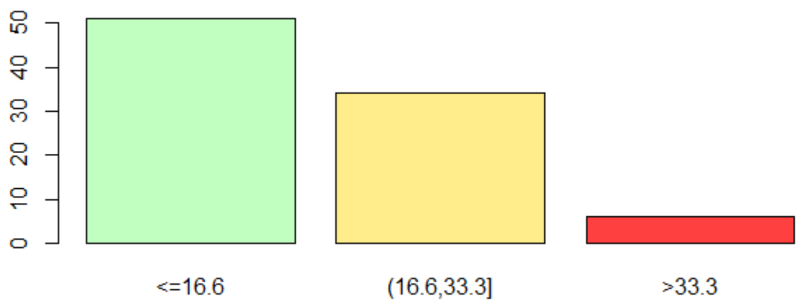


Figure 6: CV's of all district-wise indirect estimates from our best SAE-LM model

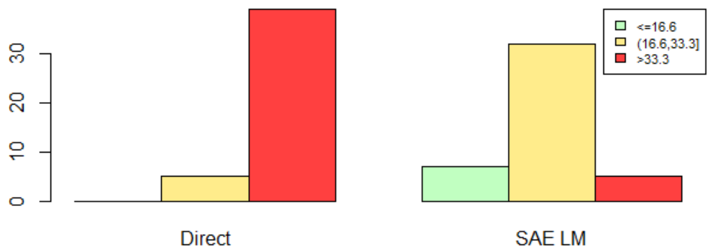


Figure 7: CVs of direct (left) and indirect (right) estimates for sampled districts

A second crucial requirement is that "the parts fit the whole", *i.e.* by aggregating all district-wise estimates should give a national estimate consistent with a reliable national benchmark. By averaging the district-wise indirect esimates provided by our best SAE-LM model for 96 Pakistan districs, leads to 0.00361 as national TB prevalence, significanlty close to the 0.00364 national estimate provided (with controlled precision) by the 2010-2011 TB prevalence survey (see for instance WHO report at *https://www.who.int/tb/advisory_bodies/impact_measurement_taskforce/meetings/gf5presentation09_pakistan2011prevalencesurve*).

Further assessment of the precision of district estimates of prevalent cases are uncertainty intervals with acceptable short length.

Interval estimates, with chosen confidence level, are given directly by posterior distribution of small area parameters of our best SAE-LM model, as Bayesian credible interval. Familiar 95% Confidence Intervals around the area prevalence estimate $\pm 1.96 \times SE$ (SE= Standard Error) can be computed as well. The complete list of the district 95% interval estimates is provided at the end of this report.

Further useful diagnostic is given via visualization of the differences (either positive or negative) between direct district estimates and indirect district estimates produced by our best SAE-LM model, for the subset of sampled districts. In Figure 8 such differences are plotted (on the vertical axe) versus all domains (on the horizontal axe) ordered with respect to variability of direct estimate.

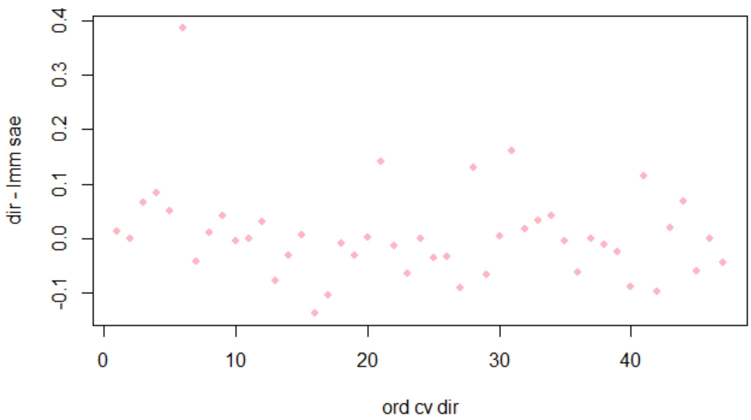


Figure 8: Differences between direct and indirect estimates for sampled districts ordered by CV of direct estimates

The distribution appears markedly symmetric around 0, with smaller differences for more precise direct estimates and larger differences as the variability of direct estimate increases. This suggests that our SAE-LM model has no systematic tendency to either over- or under- estimation.

(Further diagnostics has been performed and is available on demand)

Traditional (frequentist) goodness of fit diagnostics can be used to test whether or not there is significant difference between direct and indirect estimates provided by our best SAE-LM model, e.g. Wald's Chi Square. However it is not advisable: frequentist diagnostics to test Bayesian estimation would give rise to consistency concerns.

Finally, we remark that cross-validation methods, for instance the recommended leave-one-out cross-validation, are not practical options for SAE-LM model validation, essentially due to its resource consuming nature. However, there is a theoretical connection between leave-one-out cross-validation and BIC, as information criterion methods based on maximum likelihood which we used (for technical details see Klaenskens, 2008, *Model selection and Model averaging* Cambridge University Press, Section 2.9)

Essential details on SAE-LM model specification

- List of Notation

  True Prevalence = # of TB cases/# population

P	national	
P_d	for district d	
P_{dt}	for district d at time point t	
D	# of districts with complete covariates availabe	$d = 1, 2 \dots D$

  Estimates

  $\hat{P}_d$ direct prevalence estimate for district d

  computed upon district sample data of size $n_d > 0$ at time of the national prevalence survey; no direct estimate available for non-sampled districts with $n_d = 0$

Σ_d	sampling variance of direct estimate	
$\check{P}_d$	indirect estimate for district d	primary output of SAE-LM fitting & selection algorithm

  U_{dt} underlying un-observable (latent) process with k latent states (classes of TB burden ultimately provided by the SAE-LM selected model)

- Cross-sectional data: sample data at time of the national prevalence survey (time 1= 2011), constant at all time points

- Longitudinal data: auxiliary data available at district level at different points in time, starting from the time of the national prevalence survey, to be used as model covariate

  Time period: 2011-2016 $\rightarrow$ time points: $t = 1, 2 \ldots 6$

  c # of covariates (entering the model)

  $\underline{x}_{dt}$ $c \times 1$ vector of covariates for district d at time point t

- SAE-LM main assumptions:

  P_{dt} are conditionally independent given U_{dt}, that is the true values of district prevalence depend only on the underlying latent process.

  The latent state to which a district belongs at a given time point only depends on the latent state at the previous point in time.

- Model hierarchy and probability distributions:

  Sampling model: direct district prevalence estimates given (conditioned on) true district prevalences

$$\hat{P}_d | P_d \sim N\left(P_d,\ \Sigma_d\right)$$

  as Linking model, the popular SAE (area level) regression linking model $P_{dt} = \underline{x}_{dt}\underline{\beta}\ +\ N(0, \sigma^2)$ is replaced by the LM model, split into its two parts:

  Measurement model: true district prevalence given covariates, i.e. the measurable part of the latent process

$$P_{dt}|(U_{dt} = u) \sim N\left(\underline{x}_{dt}\underline{\beta}_u,\ \sigma_u^2\right)$$

  Latent Model: probability distribution (discrete and dynamic in time) of the residual un-observed part of latent process, not explained by the upper hierarchy

$$\begin{array}{ll} U_{dt} = u \sim & \text{(1st order) Markov chain with} \\ \quad k & \text{latent states } u = 1, 2 \ldots k \\ \quad [\pi_u] & k \times 1 \text{ vector of initial probabilities} \\ \quad [\pi_{u|u'}] & k \times k \text{ matrix of transition probabilities, constant in time} \end{array}$$

- The overall set of model parameters comprises three groups of parameters to be estimated by the SAE-LM fitting & selection algorithm

  small areas parameters: $[P_{dt}]$ $D \times 6$ matrix of indirect area estimates (the primary output of the modeling)

  measurement parameters: $\left[\underline{\beta}_u,\ \sigma_u^2\right]$ $2k \times 1$ vector of regression coefficients and error variances

  latent process parameters: $[\pi_u]$ k vector of initial probabilities, and $\left[\pi_{u|u'}\right]$ $k \times k$ matrix of transition probabilities.

List of indirect (point and interval) district estimates

District	point estimates	95% Bayes Credible Intervals	95% Confidence Intervals
Abbottabad	0.00550	[0.00399, 0.00699]	[0.00401 0.00699]
Attock	0.00144	[0.00070, 0.00211]	[0.00073, 0.00214]
Awaran	0.00347	[0.00251, 0.00442]	[0.00252, 0.00443]
Badin	0.00156	[0.00066, 0.00238]	[0.00068, 0.00243]
Bahawal Nagar	0.00400	[0.00265, 0.00542]	[0.00262, 0.00538]
Bannu	0.00646	[0.00620, 0.00672]	[0.00620, 0.00672]
Barkhan	0.00406	[0.00332, 0.00478]	[0.00333, 0.00478]
Bhakkar	0.00478	[0.00457, 0.00499]	[0.00457, 0.00499]
Chagai	0.00881	[0.00801, 0.00960]	[0.00801, 0.00960]
Chakwal	0.00409	[0.00388, 0.00430]	[0.00388, 0.00430]
Chiniot	0.00097	[0.00033, 0.00161]	[0.00033, 0.00161]
Chitral	0.00353	[0.00320, 0.00386]	[0.00320, 0.00386]
Dadu	0.00597	[0.00444, 0.00751]	[0.00443, 0.00751]
Dera bugti	0.00008	[−0.00012, 0.00028]	[−0.00012, 0.00028]
Dera ghazi khan	0.00193	[0.00108, 0.00270]	[0.00112, 0.00273]
Dera ismail khan	0.00531	[0.00505, 0.00557]	[0.00505, 0.00557]
Faisalabad	0.00152	[0.00088, 0.00216]	[0.00087, 0.00216]
Ghotki	0.00247	[0.00144, 0.00348]	[0.00129, 0.00366]
Gujranwala	0.00193	[0.00119, 0.00268]	[0.00118, 0.00268]
Gujrat	0.00000		
Hafizabad	0.00184	[0.00105, 0.00253]	[0.00107, 0.00261]
Hangu	0.00360	[0.00329, 0.00391]	[0.00329, 0.00391]
Haripur	0.00443	[0.00298, 0.00587]	[0.00296, 0.00590]
Harnai	0.00749	[0.00530, 0.00887]	[0.00585, 0.00912]
Hyderabad	0.00355	[0.00206, 0.00490]	[0.00214, 0.00496]
Jacobabad	0.00092	[0.00038, 0.00147]	[0.00038, 0.00147]
Jamshoro	0.00560	[0.00543, 0.00576]	[0.00543, 0.00576]
Jhal magsi	0.00273	[0.00180, 0.00364]	[0.00180, 0.00365]
Jhang	0.00454	[0.00324, 0.00580]	[0.00326, 0.00581]
Jhelum	0.00416	[0.00278, 0.00555]	[0.00279, 0.00554]
Kachhi	0.00498	[0.00491, 0.00506]	[0.00491, 0.00506]
Kalat	0.00491	[0.00357, 0.00640]	[0.00351, 0.00631]
Karak	0.00416	[0.00253, 0.00563]	[0.00263, 0.00570]
Kasur	0.00104	[0.00041, 0.00165]	[0.00042, 0.00166]
Kech	0.00420	[0.00342, 0.00509]	[0.00333, 0.00506]
Khairpur	0.00071	[0.00008, 0.00133]	[0.00009, 0.00133]
Khanewal	0.00401	[0.00248, 0.00556]	[0.00247, 0.00556]
Kharan	0.00345	[0.00250, 0.00432]	[0.00254, 0.00436]
Khushab	0.00492	[0.00345, 0.00639]	[0.00345, 0.00638]
Khuzdar	0.00604	[0.00561, 0.00648]	[0.00561, 0.00648]
Killa Abdullah	0.00422	[0.00254, 0.00571]	[0.00264, 0.00581]
Killa Saifullah	0.00437	[0.00390, 0.00485]	[0.00390, 0.00485]
Kohat	0.00153	[0.00082, 0.00217]	[0.00086, 0.00220]
Kohlu	0.00176	[0.00096, 0.00245]	[0.00102, 0.00251]
Lakki marwat	0.00429	[0.00401, 0.00456]	[0.00401, 0.00457]
Larkana	0.00554	[0.00533, 0.00575]	[0.00533, 0.00574]
Lasbela	0.00255	[0.00144, 0.00356]	[0.00139, 0.00371]
Layyah	0.00374	[0.00353, 0.00395]	[0.00353, 0.00395]
Lodhran	0.00405	[0.00261, 0.00549]	[0.00258, 0.00552]
Loralai	0.00487	[0.00441, 0.00534]	[0.00441, 0.00533]
Lower dir	0.00405	[0.00379, 0.00431]	[0.00379, 0.00431]
Malakand	0.00520	[0.00486, 0.00555]	[0.00486, 0.00555]
Mansehra	0.00513	[0.00341, 0.00667]	[0.00353, 0.00673]
Mardan	0.00315	[0.00164, 0.00440]	[0.00181, 0.00449]

Mastung	0.00424	[0.00339,	0.00511]	[0.00338,	0.00510]
Matiari	0.00585	[0.00551,	0.00619]	[0.00551,	0.00619]
Mianwali	0.00066	[0.00005,	0.00127]	[0.00005,	0.00126]
Mirpur	0.00316	[0.00275,	0.00356]	[0.00276,	0.00356]
Mirpur khas	0.00058	[−0.00010,	0.00142]	[−0.00017,	0.00133]
Multan	0.00152	[0.00083,	0.00234]	[0.00075,	0.00228]
Musakhel	0.00763	[0.00638,	0.00877]	[0.00644,	0.00881]
Narowal	0.00110	[0.00041,	0.00174]	[0.00044,	0.00176]
Nasirabad	0.00367	[0.00317,	0.00415]	[0.00318,	0.00417]
Naushahro Feroze	0.00124	[0.00053,	0.00189]	[0.00055,	0.00192]
North waziristan	0.00416	[0.00382,	0.00450]	[0.00382,	0.00449]
Nowshera	0.00076	[0.00009,	0.00152]	[0.00005,	0.00147]
Nushki	0.00277	[0.00208,	0.00350]	[0.00204,	0.00350]
Okara	0.00455	[0.00320,	0.00589]	[0.00321,	0.00590]
Panjgur	0.00367	[0.00290,	0.00443]	[0.00290,	0.00444]
Peshawar	0.00000				
Pishin	0.00331	[0.00302,	0.00360]	[0.00301,	0.00360]
Quetta	0.00444	[0.00413,	0.00476]	[0.00413,	0.00476]
Rahim yar khan	0.00161	[0.00161,	0.00088]	]0.00087,	0.00235]
Rawalpindi	0.00247	[0.00170,	0.00322]	[0.00171,	0.00322]
Sahiwal	0.00308	[0.00188,	0.00426]	[0.00190,	0.00427]
Sanghar	0.00391	[0.00244,	0.00543]	[0.00241,	0.00541]
Sargodha	0.00356	[0.00230,	0.00480]	[0.00231,	0.00482]
Shaheed	0.01000	[0.01000,	0.01000]	[0.01000,	0.01000]
Sheikhupura	0.00143	[0.00087,	0.00200]	[0.00087,	0.00199]
Shikarpur	0.00656	[0.00626,	0.00684]	[0.00627,	0.00685]
Sialkot	0.00319	[0.00178,	0.00465]	[0.00176,	0.00463]
Sibi	0.00413	[0.00349,	0.00475]	[0.00350,	0.00476]
Sohbatpur	0.00654	[0.00625,	0.00683]	[0.00625,	0.00683]
Sujawal	0.00000				
Sukkur	0.00130	[0.00064,	0.00197]	[0.00064,	0.00196]
Swabi	0.00524	[0.00501,	0.00547]	[0.00501,	0.00547]
Swat	0.00395	[0.00375,	0.00416]	[0.00375,	0.00416]
Tank	0.00427	[0.00385,	0.00469]	[0.00385,	0.00469]
Tharparkar	0.00688	[0.00661,	0.00715]	[0.00661,	0.00715]
Thatta	0.00698	[0.00675,	0.00721]	[0.00675,	0.00721]
Toba Tek singh	0.00000				
Upper Dir	0.00236	[0.00201,	0.00269]	[0.00202,	0.00269]
Vehari	0.00400	[0.00246,	0.00555]	[0.00248,	0.00553]
Washuk	0.00621	[0.00513,	0.00722]	[0.00517,	0.00724]
Zhob	0.00284	[0.00256,	0.00313]	[0.00255,	0.00313]
Ziarat	0.00269	[0.00156,	0.00382]	[0.00157,	0.00381]

List of all district-wise estimates computed and time evolution of indirect estimates

District	Direct Estimates	Indirect Estimates 2011	2012	2013	2014	2015	2016
abbottabad	0.00562	0.00550	0.00522	0.00516	0.00476	0.00467	0.00480
attock	0.00143	0.00144	0.00143	0.00143	0.00145	0.00147	0.00151
awaran		0.00347	0.00307	0.00400	0.00303	0.00357	0.00436
badin	0.00222	0.00156	0.00156	0.00153	0.00152	0.00153	0.00157
bahawal nagar	0.00484	0.00400	0.00377	0.00374	0.00469	0.00410	0.00408
bannu		0.00646	0.00622	0.00557	0.00587	0.00520	0.00446
barkhan		0.00406	0.00424	0.00472	0.00515	0.00577	0.00476
bhakkar		0.00478	0.00486	0.00530	0.00516	0.00509	0.00499
chagai		0.00881	0.00747	0.00876	0.00847	0.00587	0.00783
chakwal		0.00409	0.00420	0.00436	0.00446	0.00434	0.00411
chiniot	0.00147	0.00097	0.00097	0.00095	0.00096	0.00095	0.00100
chitral		0.00353	0.00266	0.00284	0.00308	0.00315	0.00385
dadu	0.00983	0.00597	0.00597	0.00557	0.00486	0.00443	0.00454
dera bugti		0.00008	0.00442	0.00849	0.01000	0.00792	0.00886
dera ghazi khan	0.00151	0.00193	0.00193	0.00186	0.00185	0.00180	0.00197
dera ismail khan		0.00531	0.00531	0.00514	0.00487	0.00481	0.00460
faisalabad	0.00163	0.00152	0.00153	0.00141	0.00135	0.00128	0.00165
ghotki	0.00290	0.00247	0.00246	0.00243	0.00242	0.00243	0.00247
gujranwala	0.00189	0.00193	0.00176	0.00170	0.00168	0.00174	0.00216
gujrat	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
hafizabad	0.00214	0.00184	0.00183	0.00181	0.00177	0.00179	0.00183
hangu		0.00360	0.00275	0.00270	0.00302	0.00338	0.00354
haripur	0.00366	0.00443	0.00416	0.00423	0.00378	0.00310	0.00359
harnai		0.00749	0.00859	0.00834	0.00854	0.00735	0.00648
hyderabad	0.00324	0.00355	0.00343	0.00337	0.00326	0.00315	0.00339
jacobabad	0.00099	0.00092	0.00093	0.00093	0.00093	0.00094	0.00094
jamshoro		0.00560	0.00598	0.00526	0.00563	0.00466	0.00397
jhal magsi		0.00273	0.00661	0.00631	0.00571	0.00634	0.00687
jhang	0.00316	0.00454	0.00433	0.00424	0.00453	0.00477	0.00386
jhelum	0.00311	0.00416	0.00442	0.00429	0.00450	0.00454	0.00420
kachhi		0.00498	0.00510	0.00464	0.00510	0.00514	0.00472
kalat		0.00491	0.00540	0.00517	0.00534	0.00416	0.00391
karak	0.00408	0.00416	0.00422	0.00415	0.00380	0.00449	0.00440
kasur	0.00072	0.00104	0.00098	0.00097	0.00092	0.00091	0.00106
kech		0.00420	0.00427	0.00370	0.00374	0.00343	0.00359
khairpur	0.00073	0.00071	0.00070	0.00069	0.00066	0.00062	0.00070
khanewal	0.00542	0.00401	0.00422	0.00321	0.00510	0.00508	0.00413
kharan		0.00345	0.00525	0.00423	0.00562	0.00713	0.00706
khushab	0.00478	0.00492	0.00492	0.00468	0.00472	0.00454	0.00448
khuzdar		0.00604	0.00636	0.00678	0.00619	0.00520	0.00387
killa abdullah	0.00358	0.00422	0.00430	0.00430	0.00399	0.00395	0.00411
killa saifullah		0.00437	0.00418	0.00416	0.00496	0.00547	0.00530
kohat	0.00153	0.00153	0.00154	0.00155	0.00153	0.00153	0.00157
kohlu	0.00140	0.00176	0.00177	0.00177	0.00177	0.00177	0.00177
lakki marwat		0.00429	0.00380	0.00416	0.00405	0.00374	0.00335
larkana		0.00554	0.00534	0.00458	0.00510	0.00539	0.00460
lasbela	0.00222	0.00255	0.00256	0.00257	0.00256	0.00257	0.00258
layyah		0.00374	0.00724	0.00404	0.00456	0.00423	0.00394
lodhran	0.00314	0.00405	0.00329	0.00378	0.00278	0.00274	0.00306
loralai		0.00487	0.00612	0.00446	0.00583	0.00607	0.00564
lower dir		0.00405	0.00421	0.00443	0.00446	0.00245	0.00169
malakand		0.00520	0.00506	0.00493	0.00435	0.00310	0.00294
mansehra	0.00643	0.00513	0.00458	0.00424	0.00404	0.00420	0.00427
mardan	0.00249	0.00315	0.00285	0.00273	0.00182	0.00150	0.00219

mastung		0.00424	0.00752	0.00399	0.00370	0.00350	0.00680
matiari		0.00585	0.00580	0.00519	0.00521	0.00465	0.00436
mianwali	0.00070	0.00066	0.00062	0.00059	0.00062	0.00066	0.00073
mirpur		0.00316	0.00366	0.00364	0.00386	0.00383	0.00362
mirpur khas	0.00219	0.00058	0.00058	0.00057	0.00058	0.00056	0.00058
multan	0.00169	0.00152	0.00146	0.00143	0.00135	0.00130	0.00144
musakhel		0.00763	0.01000	0.00843	0.00726	0.00680	0.00736
narowal	0.00143	0.00110	0.00107	0.00107	0.00108	0.00107	0.00113
nasirabad		0.00367	0.00417	0.00460	0.00403	0.00453	0.00350
naushahro feroze	0.00165	0.00124	0.00125	0.00124	0.00123	0.00123	0.00133
north waziristan		0.00416	0.00338	0.00355	0.00243	0.00220	0.00243
nowshera	0.00071	0.00076	0.00075	0.00073	0.00072	0.00074	0.00081
nushki		0.00277	0.00242	0.00432	0.00327	0.00394	0.00313
okara	0.00394	0.00455	0.00377	0.00378	0.00371	0.00325	0.00328
panjgur		0.00367	0.00356	0.00532	0.00484	0.00540	0.00437
peshawar	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
pishin		0.00331	0.00388	0.00271	0.00266	0.00278	0.00334
quetta		0.00444	0.00444	0.00403	0.00460	0.00441	0.00336
rahim yar khan	0.00149	0.00161	0.00157	0.00159	0.00155	0.00152	0.00162
rawalpindi	0.00223	0.00247	0.00240	0.00241	0.00257	0.00256	0.00295
sahiwal	0.00219	0.00308	0.00290	0.00246	0.00272	0.00279	0.00287
sanghar	0.00505	0.00391	0.00454	0.00329	0.00393	0.00425	0.00428
sargodha	0.00259	0.00356	0.00322	0.00291	0.00308	0.00329	0.00310
shaheed		0.01000	0.01000	0.01000	0.01000	0.01000	0.01000
sheikhupura	0.00164	0.00143	0.00138	0.00138	0.00136	0.00126	0.00140
shikarpur		0.00656	0.00661	0.00594	0.00640	0.00672	0.00572
sialkot	0.00388	0.00319	0.00320	0.00333	0.00284	0.00235	0.00227
sibi		0.00413	0.00597	0.00721	0.00555	0.00413	0.00441
sohbatpur		0.00654	0.00444	0.00444	0.00443	0.00443	0.00443
sujawal		0.00000	0.00002	0.00002	0.00620	0.00620	0.00620
sukkur	0.00071	0.00130	0.00130	0.00129	0.00129	0.00129	0.00132
swabi		0.00524	0.00536	0.00536	0.00535	0.00497	0.00429
swat		0.00395	0.00422	0.00427	0.00312	0.00199	0.00207
tank		0.00427	0.00469	0.00592	0.00508	0.00448	0.00368
tharparkar		0.00688	0.00680	0.00666	0.00679	0.00675	0.00615
thatta		0.00698	0.00688	0.00675	0.00679	0.00672	0.00641
toba tek singh	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
upper dir		0.00236	0.00261	0.00287	0.00283	0.00268	0.00253
vehari	0.00355	0.00400	0.00407	0.00397	0.00365	0.00336	0.00335
washuk		0.00621	0.00764	0.00790	0.00817	0.00913	0.00877
zhob		0.00284	0.00414	0.00371	0.00345	0.00344	0.00301
ziarat		0.00269	0.00320	0.00361	0.00383	0.00146	0.00168