

Patrick Manning, Yun Zhang, and Bowen Yi

Volume and Direction of the Atlantic Slave Trade, 1650-1870: Estimates by Markov Chain Carlo Analysis

Abstract

This article presents methods and results in the application of the Markov Chain Monte Carlo (MCMC) analysis to a problem in missing data. The data used here are from The Atlantic Slave Trade Database (TASTDB), 2010 version, available online. Of the 35,000 transatlantic slaving voyages known at that time, original data on the size of the cargo of captives exist for some 25 percent of voyage embarkations in Africa and for about 50 percent of arrivals in the Americas. Previous efforts to estimate the missing data (and project the total number of captives who made the transatlantic migration) have proceeded through eclectic projections of maximum likelihood estimates of captives per voyage, without error margins. This paper creates new estimates of total migrant flow through two methods: one is a formally frequentist set of multiple methods, and the other is through Markov Chain Monte Carlo methodology. Comparison of the three methods, all based on the same raw data, show that the results of our two new methods are fairly close to one another and they yield total flows of migrant captives of more than 20 percent higher than the previous estimates. Quantitative results, presented in simplified graphs and tables within the text and in detailed spreadsheets available online, provide a new estimate of the volume of African embarkations and American arrivals in the transatlantic slave trade for the period from 1650 to 1870, by decade, for eleven African regions of embarkation and seven American and European regions of arrival.

Patrick Manning, Yun Zhang, Bowen Yi

Volume and Direction of the Atlantic Slave Trade, 1650-1870: Estimates by Markov Chain Carlo Analysis

The volume and direction of the Atlantic slave trade has long been a subject of importance and controversy. Detailed research has explored the numbers embarked on slave vessels at various African ports, the numbers who lost their lives in the course of the voyages of two months or more across the Atlantic, and the numbers who disembarked at the end of the voyage, mostly in the Americas. While a great deal of information has been gathered, especially within the past fifty years, the problem of missing data remains serious.¹

Markov Chain Monte Carlo (MCMC) analysis provides a systematic and comprehensive method for estimating missing data. We use it to present estimates of the total number of captives embarked, by decade (from the 1650s to 1860s), for eleven regions of the African coast. In the same analysis, we estimate the number of captives who arrived in the same decades for six American regions and for Europe. The raw but incomplete data from which the missing data are estimated come from The Atlantic Slave Trade Database (Eltis, et al., 2010: hereafter TASTDB 2010). In addition, the article discusses two other procedures for estimating the levels of African slave trade: one based on the combination of multiple methods carried out by our group; the other being the projections produced and presented along with the TASTDB.

Studies of Atlantic Slave Trade: from Synthesis to Voyage-Based Data

Two major steps forward, and many smaller steps, have characterized the quantitative study of the Atlantic slave trade. First, the work of Philip Curtin, *The Atlantic Slave Trade: A Census*, presented comprehensive estimates of the volume of slave trade (Curtin 1969). He offered a total of 9.5 million persons—for the number of arrivals (that is, disembarkations) of captive Africans in the Americas from the fifth century to 1870. Curtin carried out his research through secondary works and with a wide range of methods. His estimated total, which was smaller than previously thought, brought an outpouring of research into primary documents. This research brought distinction among such related variables as numbers of captures, embarkations, arrivals, and mortality, including distinction among the processes of migration in Africa, on the Atlantic, and in the Americas. Joseph Inikori sought to show that the estimate should be increased, and a debate ensued (Inikori 1976; Curtin 1976). Increasingly, the analysis focused on documentation of Atlantic slaving voyages, though there remained numerous other methods for estimating aspects of the slave trade. After twenty years of debate, the estimated volume of the slave trade had crept up by a million or so, and clearer distinctions were made on estimates of embarkations, arrivals, and mortality at various stages in the trade (Lovejoy 1982, 1989).

The second major step in documenting the Atlantic slave trade was a project to combine the many separate research projects on slaving voyages into a comprehensive database. David Eltis led this work, in association with several colleagues, and sustained it over more than twenty years. The initial report of this research was a 2000 CD-ROM with data on some 27,000 slave voyages, organized according to a systematic codebook (Eltis et al., 2000). The publication is also known by the database itself as The Atlantic Slave Trade Database or TASTDB (Eltis et al., 1999: hereafter TASTDB 1999). This aggregative research clarified distinctions between (1) known voyages and unknown voyages and (2) between known numbers of captive migrants—documented for known voyages—and missing values of captive migrants on known voyages. The succeeding versions of the database included large and expanding numbers of variables regarding the ships, ownership, crew, times and places of the voyages, and details on cargoes of captives. The database includes raw data, recorded directly from the sources, and imputed variables, calculated and estimated by the editors from the raw data. Most cells for the full database are empty for lack of specific data, yet the full collection of data is immensely rich and varied.

The 1999 CD-ROM was a great step forward and received wide attention (William and Mary Quarterly 2001). At first the editors emphasized the relative completeness of their data, which were indeed effective in identifying the great majority of voyages in the British, Dutch, and French slave trades. With time, however, and as submissions of new data continued, the editorial team recognized that there were many missing voyages—especially in Portuguese, Spanish, and Brazilian vessels—and energetically added the new information to their database. By 2008 the editors had published an expanded database in an online venue, “Slave Voyages” (Eltis et al. 2008a, 2008b). Since the online publication of the 2008 database, a subsequent and expanded 2010 database was added to the Slave Voyages website (TASTDB 2010). By 2010, the number of voyages included had risen to nearly 35,000. Significant additional modifications were published on the Slave Voyages website in 2015.

Estimates of Missing Migration Data: Alternative Methods

The purpose of this analysis is to develop new and useful estimates of the volume of the Atlantic slave trade. The effort to document the overall volume of the Atlantic slave trade—in time and space and with attention to gender, ethnicity, age, and mortality—has been the principal topic of interest in the overall analysis of the slave trade. The synthesis-based analyses, most notably by Curtin and Lovejoy, assembled an eclectic range of data and analytical techniques. In the discussion, distinctions among raw data, direct and indirect estimates were necessarily fuzzy. With time and expanded research, voyage-based data became increasingly important in slave-trade studies: the remarkably detailed archival studies of Jean Mettas on the French slaving voyages stood out in this regard as an exemplary summary of raw data (Mettas 1978-84).

Eltis, from his earliest work on British reports on slave trade in the 1840s, began a practice of estimating “imputed” numbers of captives for known voyages with missing data on numbers of captives (Eltis 1979). In later years he expanded this practice with a second procedure for estimating additional missing values (Eltis 1987a, Eltis 1987b).

The 1999 CD-ROM included raw data on captive migrants and also included “imputed” figures for missing data on captive migrants. Along with the database, presented in SPSS format, the editors made estimates of missing values and proposed overall totals of slave trade by region and by time period, in an extension of Eltis’s earlier procedures. The procedure, however, did not include any estimates of error margins or tolerances in the estimates. These methods, using various approximations for various situations, were developed in the course of work on the

Transatlantic Slave Trade Database and the Slave Voyages project: types of voyages, average number of captives per voyage, etc. (Eltis and Lachance 2010).

The present analysis goes beyond previous estimates to develop a second and third opinion on missing data and totals in the volume of Atlantic slave trade, based on two sets of statistical methods. That is, working with a single set of data on known voyages and known captive migration (from the 2010 TASTDB database), this analysis presents three estimates of missing values and total volume of the transatlantic slave trade: the 2010 Slave Voyages estimates, our estimates through multi-method frequentist analysis, and our estimates through a Bayesian statistical approach, Markov Chain Monte Carlo analysis.

The variables and values reported in the 2010 TASTDB database are of two broad categories: variables coded from original sources (what we will call “raw data”) and variables created through transformation of the raw data (what we call “imputed variables,” following the terminology of the database’s editors.) We rely on both raw data and imputed variables from the TASTDB database. Data were drawn from the TASTDB 2010 database for the decades from the 1650s to the 1860s for the following variables:

voyageid: arbitrary voyage identification number, imputed by editors
majbyimp: region of embarkation (R_E); imputed, based on port data
mjselimp1: region of initial arrival (R_A); imputed, based on port data
year10: decade, imputed and numbered serially from 1650s to 1860s.
voy2imp: imputed days of transatlantic voyage (L)
tslavesd: recorded captive embarkations (E)
slaarriv: recorded captive arrivals (A)

Two additional imputed variables from the TASTDB 2010 database are important for comparison:

slaximp: imputed captive embarkations
slamimp: imputed captive arrivals

These are the overall estimations of embarkation and arrival computed in TASTDB 2010, to which our estimates are compared in the concluding sections of this study.

As a simplified statement of the strategy of comparing these analyses, one may propose two simple equations for the estimation of captive embarkations in Africa, one summarizing the unknown reality of the past, and the other summarizing our best estimates:

known embarkations + actual missing data = actual total embarkations
known embarkations + estimated missing data = estimated total embarkations

where the underscored terms represent the estimates made by various methods. The “known embarkations” are given in each case through the variable “tslavesd” from TASTDB 2010. The “estimated total embarkations” are given by the variable “slaximp” from TASTDB 2010 and by the results of multiple-method analysis and MCMC analysis from our analysis.

Here are additional comments on the handling of raw data and imputed variables in our analysis. First is spatial information. For all of the estimates of captive transatlantic migration, it is necessary to specify regional dimensions of embarkation and arrival. The original data in Slave Voyages are coded geographically by port—the various ports

of slave trade on the African coast and the ports of slave disembarkation or arrival in the Americas. For the estimation of missing values in numbers of captives, the ports are grouped into regions, both for the African and American coasts. The regions of African slave trade are known by a relatively consistent terminology, though the boundaries among regions are not precise: the regions of Western Africa run from Senegambia in the north to Angola in the south. Regions in the TASTDB database were based on groupings of the ports listed in the original data. We accepted the regions from Senegambia through the Bight of Benin, as defined in TASTDB, but recoded the data for the region from the Bight of Biafra through Angola. Where TASTDB reports only on Bight of Biafra and West Central Africa, we reclassified ports to yield data on the Bight of Biafra, Gabon, Loango, and Angola. Our choice was not for statistical reasons but to fit more tightly with the historical literature, which has distinguished the slave trade of the Congo basin from that of Angola more fully than is permitted by the very large West Central Africa region of the TASTDB database. We wanted to estimate slave exports for the regions we believe to be most commonly and usefully employed in analyzing export slave trade from the Bight of Biafra through Angola. (The totals should be unchanged by this reclassification of ports, but the distribution among regions is different.) For regions of arrival, we used the regions as documented in TASTDB 2010.

Second, in addition to regional analysis, the volume of slave trade has been calculated in terms of the flag or national affiliation of each slaving voyage. This approach permits assigning the flow of captive migrants to the British, French, Dutch and other carriers: it provides more information on the number of captives delivered to various imperial units in the Americas than it provides on the African origin of captives, so we will give little attention to this approach in the present study.

Third, analysis of the volume of African slave trade is habitually carried out in terms of decades. (Decades may be defined with different starting points: the literature on Atlantic slave trade includes both decades starting at year 0 and at year 1—thus, 1740–1749 and 1741–1750 may both be used. In fact, the TASTDB defines data as 1740–1750, etc.) We have chosen to label our analysis as beginning in 1650—it was at that point that the Atlantic slave trade exceeded the trans-Saharan slave trade in volume, and documentation becomes steadily more dependable from 1650 forward.

Fourth, our analysis includes the length of voyage as a variable: we have used the imputed variable from TASTDB (*voy2imp*), which calculates the number of days between departure and landfall on each transatlantic voyage. As will be noted later in this study, the results of our analysis of voyage length were somewhat problematic, and require further investigation.

In sum, our analysis according to all three methods—imputed figures from TASTDB 2010, our multiple-methods analysis, and our MCMC analysis—relies most heavily on the raw data on captive embarkations and arrivals, as reported in the 2010 TASTDB database, but also relies on a number of imputed variables from the same database. The data are explored over the course of 22 decades. Within each decade there are 11 regions for embarkation and 7 regions for arrivals. Thus there are 242 cells in the output for embarkation and 154 cells in the output for arrivals.

In addition, we note that the editors of the Slave Voyages project have recently published a new set of estimates of embarkations, arrivals, and total flow of captives in the transatlantic slave trade.² These estimates, based on projections of undocumented slave voyages according to various techniques, propose decennial and total volumes of the Atlantic slave trade that are considerably larger than their previous estimates. While we recognize that these estimates may be an important addition to knowledge about the Atlantic slave trade, we have not

attempted to include them in the present study. Our objective is to compare the estimates of slave-trade volume using the same data but different methods, to clarify the consistencies, inconsistencies, strengths and weaknesses of the various methods.

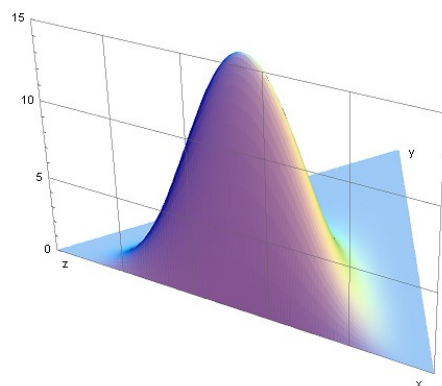
The two statistical frameworks applied here—multiple-method and MCMC—each draw on various theoretical distributions of data in their analyses. The various distributions depend on the known and unknown aspects of each. As an indication of their distinctions, formulas and graphic representations are shown here for the Dirichlet, Poisson, Binomial, and Gamma distributions. Figure 1 presents views of the various probability distributions used in the multiple models, along with their algebraic representation. (In addition, the multinomial distribution is also used in the analysis but is not displayed as an image.)

Most of the rest of this paper consists of two detailed descriptions of estimates of missing values: for the multiple-method analysis and for the MCMC analysis. Concluding sections and the Appendix present a comparison of the various estimates and suggestions for further investigation of this issue.

New Method 1: multiple-method analysis of voyage-based data

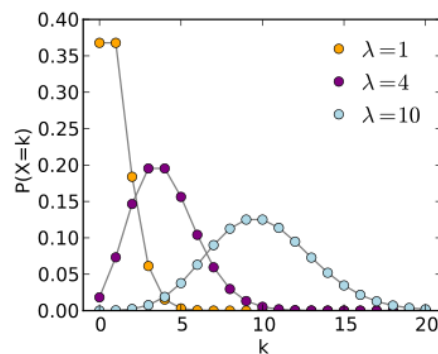
In multiple-method analysis, we identify different types of missing data, then project missing data for each type according to an appropriate algorithm. The primary goal is to impute the embarkation count for a given region within a certain decade. The voyages that may contribute to the total embarkation count within a given region and decade can be partitioned into sixteen categories based on the existence of records in the variables: embarkation region, decade, embarkation count and arrival count. The strategy is to compute the total embarkation count for each subset of data partitioned by missing type. Table 1 outlines the methods that will be employed in the treatment of each subset. Then the individual contributions are added to the current total and the estimated variance is updated by summing the variances. This implicitly assumes that these totals are independent, but also provides lower bounds of the estimate of variance.³

In the data preparation stage, we check for known inconsistencies in the database. One is the situation in which the arrival count is much greater than the embarkation count. We treat these cases by keeping the arrival count but labeling the embarkation count as missing. In fact, records with arrival count to embarkation count ratio between 0.9 and 1 are also suspicious, but we have left them unedited for now.



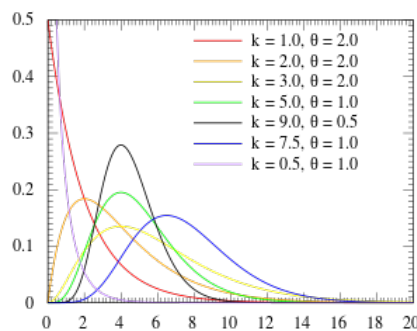
$$p(x_1, \dots, x_K) = \frac{1}{B(\alpha)} \prod_{i=1}^K x_i^{\alpha_i - 1}, \quad x_i \in (0, 1)$$

Dirichlet



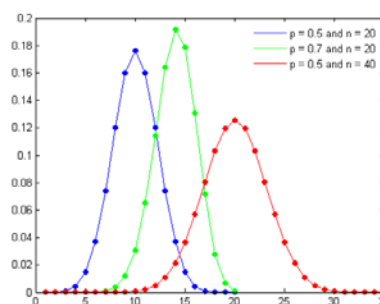
$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 1, \dots, \infty$$

Poisson



$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x \in (0, 1)$$

Gamma



$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 1, \dots, n$$

Binomial

Figure 1: Illustration of four distributions used in the analysis, where p is a continuous probability density function (PDF) and P is a discrete probability mass function (PMF).

	Region w/ Decade	Decade w/o Region	Region w/o Decade	Neither
Embark.	Direct Estimate	Multin.	Multin.	Multin.
Arriv. w/o Embark.	Ratio & SAE	Ratio & Multin.	Ratio & Multin.	Ratio & Multin.
Neither	Mean & SAE	Mean & Multin.	Mean & Multin.	Mean & Multin.

Table 1: Multiple methods by region, decade, and embarkation/arrival

Table 1 shows the type of estimates applied for voyages under three sets of embarkation/arrival conditions and four sets of combinations of region and decade. On the vertical axis, the table shows voyages with data for embarkations, voyages with data for arrivals without embarkations, and voyages with neither. On the horizontal axis, the table shows voyages documented for region and decade, for decade without region, for region without decade, and with neither. The types of estimates shown in the table are abbreviated as follows: SAE - small area estimation; Ratio - ratio estimate; Multin. - propagation through the multinomial model; Mean - using the mean as an estimate.

Note that all the models we introduce below are built for each decade separately except for the decade assignment model.

Ratio Estimate: To illustrate the method, we perform a ratio estimate for the embarkation count (E_i) against the arrival count (A_i). We seek to estimate the ratio β if the assumed model is

$$E_i = \beta A_i + \epsilon_i \quad (1)$$

where $\epsilon_i \sim n(0, \sigma^2)$. The estimate for β with variance is given by,

$$\hat{\beta} = \frac{\sum_{i \in T} E_i}{\sum_{i \in T} A_i} \quad (2)$$

$$\widehat{\mathbf{V}}(\hat{\beta}) = \left(1 - \frac{n}{N}\right) \frac{1}{n \bar{A}_N^2} \frac{\sum_{i \in T} (E_i - \hat{\beta} A_i)^2}{n - 1} \quad (3)$$

where T is our training sample (those with both records for embarkation and arrival count); n and N are the size of the training sample and all the samples with the arrival count respectively; \bar{A}_N is the mean of the arrival counts amongst all those voyages with the count. The covariance between the individual regional ratio estimate and the pooled ratio estimate will be needed in the shrinkage procedure, and is given by,

$$\widehat{Cov}(\beta_p, \beta) = \widehat{V}(\beta_i) \frac{\sum_{i \in p} A_i}{\sum_{i \in \cup_j p_j} A_i} \quad (4)$$

Notice that the implicit assumption with this method is that the missingness of the embarkation count can be considered to be randomized, hence, making the training set a random sample of all the voyages with the arrival count.

In the cases with both the region and decade records, small area estimation will be used, yielding a modified estimate of β . We are now able to estimate the net embarkation count (E_M), for voyages containing the arrival count and missing the embarkation count, with estimated variance by,

$$\hat{E}_M = \hat{\beta} \sum_{i \in M} A_i \quad (5)$$

$$\widehat{V}(E_M) = \left(\sum_{i \in M} A_i^2 \right) \widehat{V}(\hat{\beta}) \quad (6)$$

Mean Estimate: The mean estimate is a very straightforward standard calculation. For a given embarkation region, r , we will denote the mean of the embarkation count in our training sample (those with records for embarkation) as \bar{E}_r , and size of training sample as $|r|$

$$\widehat{V}(\bar{E}_r) = \frac{\sum_{i \in r} (E_i - \bar{E}_r)^2}{|r| - 1} \frac{1}{|r|} \quad (7)$$

$$\widehat{Cov}(\bar{E}_r, \bar{E}) = \widehat{V}(\bar{E}_r) \frac{|r|}{\sum_j |r_j|} \quad (8)$$

This also undergoes the small area estimate modification in the sparser cases, with both region and decade records.

Small Estimate: We use small area estimation when performing either an estimation with a region mean, or via a ratio estimate. In both cases we use a technique called composition or more generally shrinkage, by which the estimator is replaced by a linear combination ($\hat{\theta}_r^C$) of the regional estimator from the pooled regions.

$$\hat{\theta}_r^C = (1 - b)\hat{\theta}_r + b\hat{\theta} \quad (9)$$

We will construct an optimal estimate for b , b_r , based on previously computed estimates for the region estimate variance, $\mathbf{V}(\hat{\theta}_r) = v_r$, the pooled regions estimate variance, $\mathbf{V}(\hat{\theta}) = v$, and the covariance between the region and the pooled regional estimates, $cov(\hat{\theta}, \hat{\theta}_r) = c_r$.

$$b_r = \frac{v_r - c_r}{v_r + v - 2c_r + \sigma_B^2} \quad (10)$$

$$\hat{\sigma}_B^2 = \frac{1}{n} \left(S - \sum_{r=1}^R m_r v_r \right) \quad (11)$$

$$S = \sum_{r=1}^R m_r (\hat{\theta}_r - \hat{\theta})^2 \quad (12)$$

where m_r and m are the number of voyages in each region and the total number of voyages respectively. On the rare occasion that this estimator is negative then we conclude that the actual value is very small, in which case, $\hat{\theta}$ should be used to estimate θ_r .

Region/Decade Assignment Model

We will call the idea of assigning embarkation counts without region or decade or both values according to a multinomial distribution the “region/decade assignment model.” Some mention should be made as to what actual imputations are being performed. The missing regions are not being imputed for given voyages, rather we would like an estimate of each region’s contributions to the pool of voyages with missing region records. Consider that the region of departure may take one of several discrete values, which we index 1, 2... I . Denote the embarkation counts from region i in decade j as N_{ij} , and the embarkation counts with missing region in decade j as n_j . Assuming conditional on n_j , each region’s contributions n_{1j}, \dots, n_{Ij} follow multinomial distribution with probabilities q_1, q_2, \dots, q_I , then the maximal likelihood estimate of q_i and variances are given by

$$\hat{q}_i = \frac{N_{ij}}{\sum_{i=1}^I N_{ij}} \quad (13)$$

$$\hat{n}_{ij} = n_j \hat{q}_i \quad (14)$$

$$\widehat{\mathbf{V}}(\hat{q}_i) \approx \frac{\hat{q}_i(1 - \hat{q}_i)}{\sum_{i=1}^I N_{ij}} \quad (15)$$

$$\widehat{\mathbf{V}}(\hat{n}_{ij}) \approx \widehat{\mathbf{V}}(\hat{n}_j) \widehat{\mathbf{V}}(\hat{q}_i) + \hat{n}_j^2 \widehat{\mathbf{V}}(\hat{q}_i) + \hat{q}_i^2 \widehat{\mathbf{V}}(\hat{n}_j) \quad (16)$$

One additional remark is that N_{ij} and n_j are computed using the most recent calculation (imputation) of the embarkation slave count (from those samples with no missing in region and decade, and samples with missing region respectively)

Identical procedures are performed to handle voyages with missing decades. Consider that the decade of departure may take one of several discrete values, which we index 1, 2... J . Denote the embarkation counts from region i in decade j as N'_{ij} , and the embarkation counts with missing decade for region i as n_i . Assuming that conditional on n_i , each decade’s contributions n'_{i1}, \dots, n'_{iJ} follow the multinomial distribution with probabilities p_1, p_2, \dots, p_J , then the maximal likelihood estimates of q_j and variances are given by

$$\hat{q}_j = \frac{N'_{ij}}{\sum_{j=1}^J N'_{ij}} \quad (17)$$

$$\hat{n}'_{ij} = n_i \hat{q}_j \quad (18)$$

$$\widehat{\mathbf{V}}(\hat{q}_j) \approx \frac{\hat{q}_j(1 - \hat{q}_j)}{\sum_{j=1}^J N'_{ij}} \quad (19)$$

$$\widehat{\mathbf{V}}(\hat{n}'_{ij}) \approx \widehat{\mathbf{V}}(\hat{n}_i) \widehat{\mathbf{V}}(\hat{q}_j) + \hat{n}_i^2 \widehat{\mathbf{V}}(\hat{q}_j) + \hat{q}_j^2 \widehat{\mathbf{V}}(\hat{n}_i) \quad (20)$$

It should be noted that N'_{ij} is not equal to N_{ij} for the reason that $N'_{ij} = N_{ij} + \hat{n}_{ij}$. Equivalently, N'_{ij} can be thought of as the updated version of N_{ij} . Likewise, if both region and decade are missing, then we estimate the percentage of these embarkation counts that contribute to each region and decade in the analogous way using the most updated imputations.

This estimate of decennial slave embarkations by region applied a complex method with many independent parts. Many modeling assumptions were made, corresponding to each type of estimate. The most important assumption was that of random missingness. While this assumption is reasonable in most situations, the assumption that the missingness of a region is not confounded with the embarkation count of the region is somewhat questionable. The same considerations arise when information on the decade is missing. These are different assumptions to test, as the missing values may not be observed at the time they are missing. There are also instances where the procedure uses samples with a certain type of missingness repeatedly to train parameters that will allow the imputation of embarkation counts from samples with other types of missingness. For convenience, we ignore the variance brought by repeated measurements when we estimate the standard error of the imputations.

New Method 2: Markov Chain Monte Carlo Analysis

Our second new method relies on Bayesian statistical procedures. For a clear introduction to Bayesian principles in statistics, see Casella and Berger (2001: 324). The strategy of the analysis is to estimate the number of embarkations and the number of arrivals for each decade and each region, where the data are organized by voyage and where the length of the voyage has been previously estimated. That is, in the MCMC analysis, everything is known except the number of embarkations and arrivals for each decade and region: these two variables are estimated in the analysis. In fact, there were missing data for region and voyage length as well as for embarkations and arrivals and these needed to be simulated in order to carry out the MCMC analysis for missing embarkations and arrival data; in contrast, there were no missing data for decade of departure from Africa. A typical MCMC estimation includes numerous iterations of the process of randomly drawing results from the algorithm. After an early set of burn-in iterations that are discarded, a selection of later iterations is retained. For instance on voyage length, we conducted 300 runs, using runs 50-99, 150-199, and 255-299.

Our primary goal is to impute embarkation count by decade and embarkation region. Note that we have some missingness in almost every field which includes embarkation count (E), arrival count (A), voyage length (L) and the regions of embarkation and arrival (R_E, R_A). Since imputed decade is

used, there is no missingness in decade. All the models we introduce below are built for each decade separately.⁴

(1) Imputation for missing Embarkation Region and Arrival Region

We first impute embarkation region and arrival region by MCMC method, particularly, the Gibbs-Sampler, assuming

$$\vec{\alpha} \sim \text{Dirichlet}(\vec{1}) \quad (21)$$

$$\vec{\beta} \sim \text{Dirichlet}(\vec{1}) \quad (22)$$

$$R_E \sim \text{Multinomial}(\vec{\beta}) \quad (23)$$

$$R_A \sim \text{Multinomial}(\vec{\alpha}) \quad (24)$$

This results in the conditional distribution:

$$\vec{\alpha} | R_A \sim \text{Dir}(\vec{1} + c_{R_A}) \quad (25)$$

$$\vec{\beta} | R_E \sim \text{Dir}(\vec{1} + c_{R_E}) \quad (26)$$

where c_R is the count of voyage in region R . Given a value of $\vec{\alpha}^{(t)}$ *alpha* drawn at iteration t :

Imputation Step: Draw $R_{E,mis}^{(t+1)}$ with density $p(R_{E,mis} | R_{E,obs}, \vec{\alpha}^{(t)})$

Posterior Step: Draw $\vec{\alpha}^{(t)}$ with density $p(\vec{\alpha}^{(t)} | R_{E,obs}, R_{E,mis}^{(t+1)})$

Following a sufficient burn-in period, the iterative procedure can be shown eventually to yield a draw, from the joint posterior distribution of $R_{E,mis}$, given $R_{E,obs}$. Convergence diagnostics can also be conducted.

The first step of the analysis was imputing missing values for regions—embarkation regions and arrival regions—using the Gibbs Sampler. Assuming embarkation and arrival regions following multinomial distributions separately with parameters $\vec{\beta}$ and $\vec{\alpha}$, which have Dirichlet distributions, using Maximal likelihood estimates as the initial input parameters, we performed the Gibbs Sampler with 500 iterations and obtained our estimated value for $\vec{\beta}$ and $\vec{\alpha}$. And we imputed the missing values of embarkation regions and arrival regions based on multinomial distributions with parameter $\vec{\beta}$ and $\vec{\alpha}$ respectively. Note that the estimated value is that of the final iteration, using rules for the Gibbs Sampler. Since there are no missing values with respect to decades and, among different decades, behaviors of embarkation and arrival can vary significantly, the MCMC method is applied to each decade separately.

Imputation for Missing Embarkation and Arrival Counts, accounting for Voyage Length

Once all the voyages have regions of embarkation and arrival, the main MCMC can take place (though it must also overcome any missing values of voyage length). The results of this work assign specific though randomly-generated regions to all of the missing embarkation regions and arrival regions. Calculations for assigning values to

missing data are carried out separately for embarkation and arrival, but are carried out in the same program for convenience.

Following a sufficient burn-in period, the iterative procedure can be shown eventually to yield a draw from the joint posterior distribution of $R_{E,obs}, \vec{\alpha}$ given $R_{E,obs}$. Convergence diagnostics can be then be conducted.

Imputation for missing Embarkation Count and Arrival Count. The distributions that we assume for these variables are as follows:

$$\begin{aligned} E|A, L, R_E, R_A, \mu, \lambda &\sim A + Poisson(\lambda_{R_E}(1 - e^{-L\mu R^*(R_E, R_A)})) \\ A|E, L, \mu, R_E, R_A &\sim Binomial(E, e^{-L\mu R^*(R_E, R_A)}) \\ f_L|A, E, R_A, R_E, \mu &\propto L^{k-1} e^{-L(1/\delta R^*(R_E, R_A) + \mu R^*(R_E, R_A)A)} (1 - e^{-L\mu R^*(R_E, R_A)}) \end{aligned}$$

Exploratory analyses and graphics indicate that the assumptions of distributions are appropriate.

The fully conditional distributions are as follows:

$$\begin{aligned} E|A, L, R_E, R_A, \mu, \lambda &\sim A + Poisson(\lambda_{R_E}(1 - e^{-L\mu R^*(R_E, R_A)})) \\ A|E, L, \mu, R_E, R_A &\sim Binomial(E, e^{-L\mu R^*(R_E, R_A)}) \\ f_L|A, E, R_A, R_E, \mu &\propto L^{k-1} e^{-L(1/\delta R^*(R_E, R_A) + \mu R^*(R_E, R_A)A)} (1 - e^{-L\mu R^*(R_E, R_A)}) \end{aligned}$$

Parameters μ, λ, δ are estimated from the voyages we have full records of. Imputation procedures proceed in much the same manner as imputation of region except that Multiple Imputations (MI) are implemented. We conduct posterior mean of the Monte Carlo samples imputation $m = 3$ times for each voyage, and then combine the results across the multiply-imputed data.

Suppose that \hat{Q}_j is an estimate obtained from data set j ($j = 1, \dots, m$) and U_j is the standard error associated with \hat{Q}_j . The overall estimate is the average of the individual estimates,

$$\bar{Q} = \frac{1}{m} \sum_{j=1}^m \hat{Q}_j \quad (27)$$

The between-imputation variance is

$$B = \frac{1}{m-1} \sum_{j=1}^m (\hat{Q}_j - \bar{Q})^2 \quad (28)$$

The total variance is

$$T = \frac{1}{m} \sum_{j=1}^m U_j^2 + \left(1 + \frac{1}{m}\right) B \quad (29)$$

Missing data for voyage length. Missing data for voyage length must be estimated before embarkations and arrivals can be estimated. The Metropolis-Hastings algorithm is used to draw L_{mis} from the distribution where the

normalizing constant is very difficult to compute. The term $(1 - e^{-\mu R(P_A, P_E)L})^{E-A}$ can be expanded into positive and negative terms and if we use the positive parts as the proposed distribution we obtain reasonable rejection thresholds. We are able to simulate from the distribution by noting that this proposal distribution may be thought of as a mixture. Namely,

$$f_{L|rest} \propto L^{k-1} e^{-L(1/\delta_{R^*} + \mu A)} (1 - e^{-\mu L})^{E-A} \quad (30)$$

$$= L^{k-1} e^{-L(1/\delta_{R^*} + \mu A)} \sum_i^{[E-A]_0} (-1)^i e^{-i\mu L} \quad (31)$$

$$= L^{k-1} e^{-L(1/\delta_{R^*} + \mu A)} (f_0(L) - f_1(L)) \quad (32)$$

where $f_0(L) = \sum_{i, \text{even}}^{[E-A]_0} e^{-i\mu L}$ and $f_1(L) = \sum_{i, \text{odd}}^{[E-A]_0} e^{-i\mu L}$ and note that,

$$L^{k-1} e^{-L(1/\delta_{R^*} + \mu A)} f_0(L) \quad (33)$$

$$\propto \sum_{i, \text{even}}^{[E-A]_0} (1/\delta_{R^*} + \mu(A+i))^{-k} \text{Gamma}(L|k, 1/\delta_{R^*} + \mu(A+i)) \quad (34)$$

This is clearly a mixture of Gamma distributions. The acceptance threshold is

$$\rho = \min\left\{1 - \frac{f_1(V)}{f_0(V)}, \left(1 - \frac{f_1(L_{t-1})}{f_0(L_{t-1})}\right), 1\right\} \quad (35)$$

where V is a draw from the proposal distributions.

Further, voyage length (Gamma) is estimated with the Gibbs Sampler and Metropolis-Hastings algorithm for unknown distributions. This requires making preliminary estimates of embarkations and arrivals in order to permit estimation of voyage length. The voyage length estimated in this process is then treated as known data in the next stage, the Monte Carlo estimation of embarkation and arrival totals. We performed imputation for missing values of embarkation numbers E , arrival numbers A and voyage length L . With the assumption that embarkation number E follows a Poisson distribution with λ_{R_E} related to the embarkation region, arrival number A follows a binomial distribution with parameters associated to the number of embarkation population E , voyage length L , the region of embarkation R_E and the region of arrival R_A , and voyage length L follows a Gamma distribution with parameters related to embarkation region R_E and arrival region R_A . Again we applied the Gibbs Sampler method to the data. Among those parameters, we paid special attention to the length of voyages. The reason is that compared with other unknown information, it is relatively complicated to derive the fully conditional distribution of voyage length. So we decided to use the Metropolis-Hastings algorithm, which enables us to deal with unknown distributions. And we did 40 iterations for the voyage length within each single run of the Gibbs Sampler.

Standard errors are reported for embarkations and arrivals, but not for voyage length because its distribution is unknown and the algorithm for calculating variance and standard error is too complex.

For each of 300 iterations in Gibbs Sampler, our procedure was to do 40 iterations for Voyage Length and take the final one as the voyage length.

Estimation of embarkations and arrivals. Overall, we performed 300 runs for the Gibbs Sampler method, using the 50th to 99th, 150th to 199th and 250th to 299th to make imputation for the missing embarkation numbers E, arrival numbers A, voyage length L, simultaneously we computed standard errors for our estimates.

The Estimates: Comparison and Evaluation

Five Appendices, each presented in the form of a set of tables, present the results of our analysis. In sum, they focus on estimating the number of persons embarking and arriving on slaving voyages from African regions by decade, in comparison with the previous estimates of TASTDB. Appendix 1 shows estimates of embarkations, comparing two estimates by multiple-methods and two by MCMC. Appendix 2 shows estimates of arrivals, displaying the one MCMC estimate completed; Appendix 3 compares our preferred MCMC estimates of embarkations with those of TASTDB. Appendix 4 compares our MCMC estimates of arrivals with those of TASTDB. Appendix 5 summarizes the raw data, the number of captives recorded as embarking and arriving in TASTDB—these are the data from which all estimates of missing data and total flows have been constructed.

The estimates of total flows are summarized in four graphs. Figure 2 shows a comparison of embarkations (E) by decade according to our MCMC estimate and TASTDB 2010. Figure 3 compares the arrivals (A) as estimated by MCMC and TASTDB 2010. These two graphs illustrate difference between the Bayesian MCMC estimates and the TASTDB estimates: the two estimates are broadly parallel, but the MCMC estimates are higher by an average 20 percent for embarkations and by an average 15 percent for arrivals, with variations from decade to decade. Figure 4 displays embarkations and arrivals as estimated by MCMC, and shows the difference between the two, which is an implicit estimate of the number or proportion of persons lost or dying at sea. The MCMC estimate is that five percent of those embarked were lost at sea. Figure 5 similarly displays embarkations, arrivals, and the implicit losses at sea for the TASTDB 2010 estimates. But in this case, the estimated arrivals averaged more than five percent higher than the embarkations.

The implicit loss rates give a strong indication of the difference between the two sets of estimates. At present, the MCMC results appear intuitively to be more consistent with demographic realities, although each set of estimates may include errors. For instance, mortality rates for transatlantic voyages suggest that an average ten to fifteen percent of captives embarked failed to survive the voyage. Further, it seems certain that the counts of arriving captives were not undercounts, which suggests that the estimated totals of arrivals (for either algorithm) were not undercounts. This reasoning suggests that the estimations of embarkations in both Slave Voyages and MCMC methods are too low by proportions of at least five percent and perhaps as much as fifteen percent.

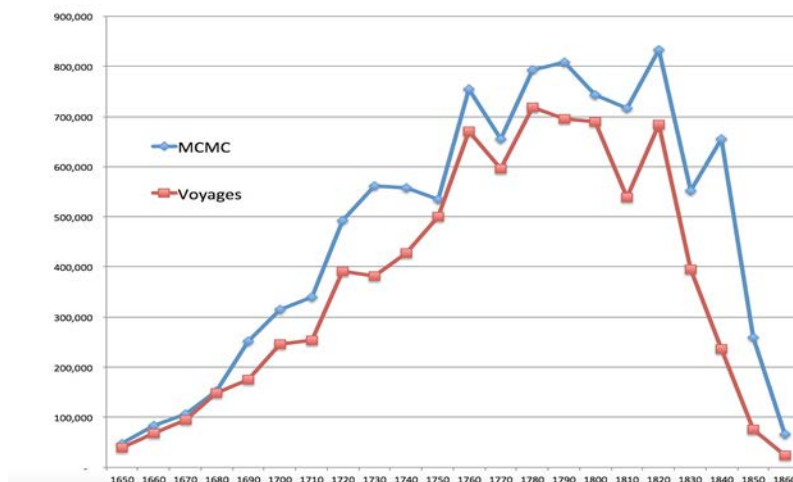


Figure 2. Embarkations: Estimates by MCMC and TASTDB 2010.

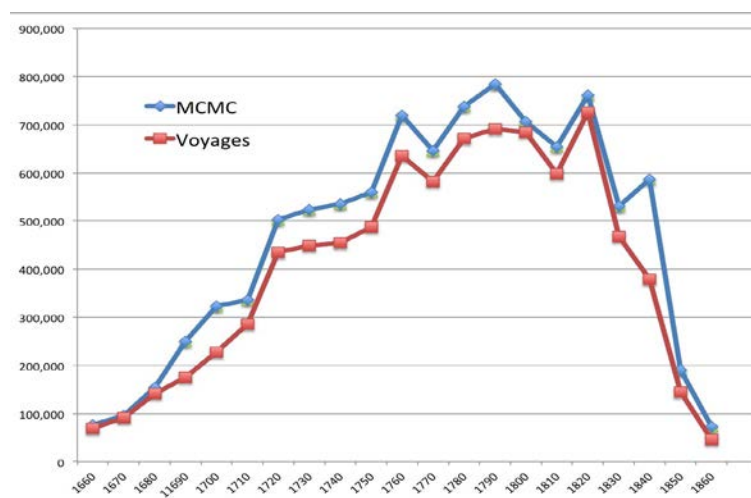


Figure 3. Arrivals: Estimates by MCMC and TASTDB 2010.

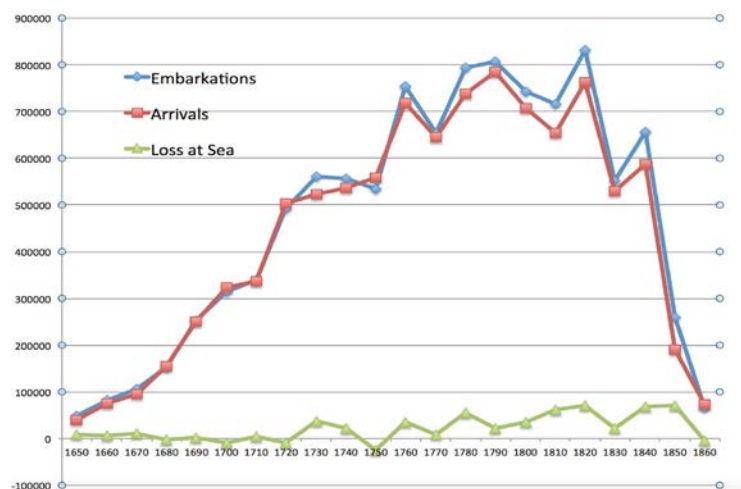


Figure 4. MCMC estimates: Embarkation, Arrival, and Imputed Loss at Sea.

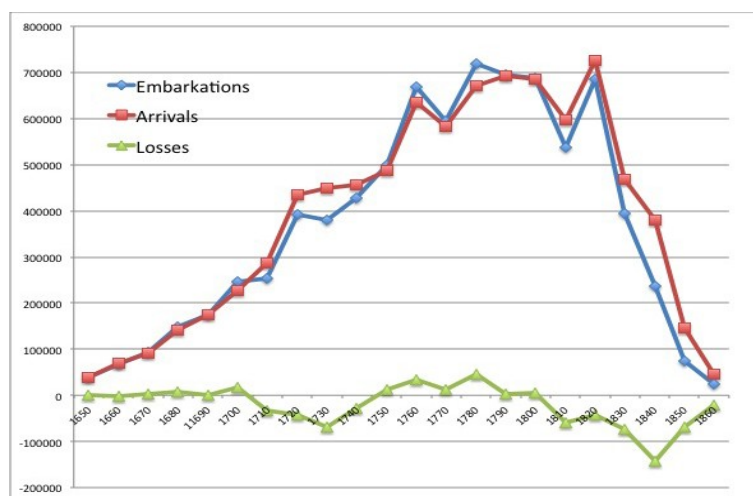


Figure 5. TASTDB 2010 Estimates: Embarkation, Arrival, and imputed Loss at Sea.

Table 2 summarizes the results of three analyses in tabular form. Within it, Table 2a shows the results of the multiple-method analysis for the decades of the 1680s and 1690s plus the three succeeding 50-year periods. (Data were insufficient for estimates by this set of techniques before 1680 and in the latest decades of slave trade.) Table 2b shows the results of MCMC analysis for four 50-year periods plus the 1850s and 1860s. As can be seen, the African totals for the multiple-method and the MCMC approaches were very close, from the 1700s to 1840s, although estimates by region varied somewhat more. Table 2c shows the TASTDB 2010 imputed embarkations of captive migrants.

Totaling the three fifty-year periods shown in Table 2 for which the three methods have comparable figures (from the 1700s to the 1840s), the multiple-methods approach suggests a total of 9.6 million embarkations, the MCMC approach gives 9.3 million embarkations, and TASTDB 2010 suggests 7.4 million embarkations. Comparing the totals for 1700–1850, the multiple-methods estimate exceeds TASTDB 2010 by 29 percent, while the MCMC estimate exceeds TASTDB by 25 percent.

Appendices 1 and 2 give full details on our estimates. That is, for two versions of the multiple-method analysis, these tables report the standard error for each decennial estimate of captive migrants and also report the 95 percent confidence levels for migration in each decade. For two versions of the MCMC analysis, the tables report the standard error for each decennial estimate of captive migrants and also report the estimated decennial mean length in days of the transatlantic voyage. As can be seen, while the MCMC analysis could be completed for all decades from the 1650s to the 1860s, the multiple-method analysis could not be completed for the 1650s, 1660s, 1670s, 1850s, or 1860s, because of the relatively small and dispersed quantity of data.

Appendices 3 and 4 report only the estimated number of captive migrants, rather than error margins, for each region and decade: this enables a straightforward comparison of the TASTDB 2010 estimates with the estimates prepared here. Overall, this comparison confirms that the multiple-method and MCMC estimates were quite close to one another where both could be calculated. In contrast, the MCMC results were just over 25 percent higher than the TASTDB 2010 results in total and for each decade. Results for individual regions by decade varied somewhat more greatly, so that a detailed comparison of them might give some hints as to possible reasons for the discrepancy.

Table 2a. Multiple-method estimation of captive migration						
Region	1680s-90s	1700s-40s	1750s-90s	1800s-40s	1850s-60s	1680s-1840s
SENEGAMBIA AND OFFSHORE ATL	36,331	157,598	218,384	52,272		464,585
SIERRA LEONE	5,509	19,369	125,723	135,037		285,638
WINDWARD COAST	-	37,005	115,533	34,406		186,944
GOLD COAST	32,133	372,316	594,538	34,044		1,033,031
BIGHT OF BENIN	172,858	588,527	508,665	420,505		1,690,555
BIGHT OF BIAFRA	38,142	95,652	477,332	430,169		1,041,295
Gabon	6,602	16,190	80,000	68,624		171,416
Loango	55,336	300,828	675,835	228,304		1,260,303
Angola	32,161	520,696	914,789	1,591,397		3,059,043
Southeast Africa and Indian ocean	9,256	6,618	73,609	420,626		510,109
Other Africa	-	33,654	198,379	26,985		259,018
SUM	388,328	2,148,453	3,982,787	3,442,369		9,961,937
Table 2b. MCMC estimation of captive migration						
Region	1650s-90s	1700s-40s	1750s-90s	1800s-40s	1850s-60s	1650s-1860s
SENEGAMBIA AND OFFSHORE ATL	67,070	148,725	211,840	65,911	800	494,346
SIERRA LEONE	5,370	14,677	163,330	128,130	2,881	314,388
WINDWARD COAST	625	24,973	163,140	30,016	-	218,754
GOLD COAST	94,474	434,742	461,963	67,643	700	1,059,522
BIGHT OF BENIN	205,728	641,212	483,092	409,466	48,967	1,788,465
BIGHT OF BIAFRA	70,825	161,036	559,376	239,591	1,972	1,032,800
Gabon	23,480	38,287	89,118	87,894	1,200	239,979
Loango	48,255	210,886	321,178	639,697	179,622	1,399,638
Angola	115,522	534,167	904,887	1,331,170	36,211	2,921,957
Southeast Africa and Indian ocean	10,965	15,395	59,322	478,400	52,010	616,092
Other Africa	616	42,025	126,452	20,245	2,788	192,126
SUM	642,930	2,266,125	3,543,698	3,498,163	327,151	10,278,067
Table 2c. Slave Voyages estimation of captive migration						
Region	1650s-90s	1700s-40s	1750s-90s	1800s-40s	1850s-60s	1650s-1860s
SENEGAMBIA AND OFFSHORE ATL	43,040	79,624	184,550	74,500	-	381,714
SIERRA LEONE	2,713	9,377	142,009	82,892	1,540	238,531
WINDWARD COAST	697	20,014	155,954	23,518	-	200,183
GOLD COAST	56,297	256,397	371,833	57,816	-	742,343
BIGHT OF BENIN	175,106	572,312	492,319	277,502	14,144	1,531,383
BIGHT OF BIAFRA	72,500	134,100	571,155	288,217	575	1,066,547
WEST CENTRAL AFRICA	122,462	491,768	1,085,001	1,384,126	70,622	3,153,979
Southeast Africa and Indian ocean	13,982	8,666	48,030	319,706	12,126	402,510
Other Africa	37,328	127,900	127,495	35,663	-	328,386
SUM	524,125	1,700,158	3,178,346	2,543,940	99,007	8,045,576
Table 2. Captive Embarkations by decade and region for three methods.						

On the face of it, it appears that the Slave Voyages method of estimation may have a consistent downward bias. We hope that this discrepancy can be modeled and explained. According to the MCMC analysis, a total of 10.2 million persons were embarked on slave ships crossing the Atlantic from 1650 to 1870. Since the TASTDB 2010 database proposes that, in earlier times from roughly 1450 to 1650, an additional 2 million persons were carried across the Atlantic, this would give a total of some 12.2 million captives transported over the whole Atlantic slave trade.

Further fine-tuning may modify the details of the MCMC results. In particular, the "Other Africa" classification of voyages in the Slave Voyages database was treated in our analysis as if it were indeed a separate region. Another possibility would be to treat these voyages as if they had missing data on region, and reclassify them among the other regions. The result would give little change to the total number of captives transported but would give different results for embarkations by region.

A further issue to explore is that the MCMC estimates for voyage length seem inconsistent with the data for known voyages, with the estimated voyage length varying surprisingly over time. It maybe that careful study of the existing data on voyage length and on the process of estimating unknown voyage lengths may clarify this issue.

Conclusion

Initial results in this research make clear the advantage of including these statistical approaches in the estimation of missing historical data for the transatlantic slave trade and for other such topics. The use of established statistical techniques confirms the basic approach of previous work, but suggests specific historical estimates that differ from previous estimates in important particulars. The introduction of formally calculated error margins is especially important. On face, the MCMC method gives standard error estimates that are remarkably small, but they make very clear that no such estimates can be taken at face value.

One striking contrast in the MCMC estimates is that between estimated total embarkations and arrivals. The two totals, calculated according to the same principles, differ from each other by only about 5 percent. This figure is surprisingly small in that, for voyages on which we have data, the rates of death per voyage generally ranged between 10 and 15 percent. Reasons for this discrepancy might be sought in various directions. One possibility is that the number of persons embarked was systematically underestimated. It is conceivable that captains and merchants exaggerated the number of persons arriving live in the Americas, but that seems very unlikely. Investigation of this issue will be important both because it addresses the logic of the Atlantic leg of slave trade and because it affects the estimate of the total number of persons embarked from African shores.

Finally, the comprehensiveness of the MCMC analysis raises the question of whether it could be extended in detail, in two directions. First, can one create a more detailed regional breakdown on both the African and American sides of the Atlantic? Second, could one estimate the movement of captives from specific African regions to specific regions in the Americas—thus, from the Bight of Biafra to Virginia or from Senegambia to the French Antilles? The first of these two possibilities is more likely than the second. By selecting smaller regions and running the MCMC program for larger numbers of iterations, one could get decennial estimates for these regions on both sides of the Atlantic, though with larger error margins. As for linking African to American regions, the additional complexity of the analysis would reduce its precision seriously. Still, such region-to-region analysis might be worth attempting for the times and places in which the number of persons undergoing enslavement was the largest.

Appendix

The Appendix provides topics, titles, and links to Excel sheets, showing numbers of embarkations (and, in some cases, arrivals) by decade from as early as the 1650s to as late as the 1860s. It presents results from the multiple-method analysis, the MCMC analysis, and the TASTDB 2010 database. Appendix 1 and Appendix 2 provide both estimates of migrant flows and error margins, where available. Appendix 3 and Appendix 4 provide estimates of migrant flows, simplified to facilitate comparability among methods. Appendix 5 summarizes the raw data from TASTDB 2010 on which all analysis in this paper is based: numbers of voyages, documented embarkations, and documented arrivals by decade. Appendix 5 also compares these raw data with MCMC estimates of embarkations and arrivals.

Appendix 1. New Estimates of Embarkations, 1650s - 1870s

Table A. Multiple method, version 1, with standard error results.

Table B. Multiple method, version 2, with standard error results.

Table C. MCMC, version 1, with standard error results.

Table D. MCMC, version 2, with voyage length and standard error results.

Excel Sheet: <http://dx.doi.org/10.7910/DVN/BJIMZA>

persistentId=doi:10.7910/DVN/BJIMZA

Appendix 2. New Estimates of Arrivals, 1650s - 1870s

Table A. MCMC, version 2, with voyage length and standard error results.

Excel Sheet: <http://dx.doi.org/10.7910/DVN/D3JUYL>

persistentId=doi:10.7910/DVN/D3JUYL

Appendix 3. Comparative Embarkation Estimates, 1650s-1870s

Table A. Multiple method, version 2 Table B. MCMC, version 2

Table C. Slave Voyages 2010

Excel Sheet: <http://dx.doi.org/10.7910/DVN/GOIFI4>

persistentId=doi:10.7910/DVN/GOIFI4

Appendix 4. Comparative Arrival Estimates, 1650s-1870s

Table A. MCMC, version 2 Table B. Slave Voyages 2010

Excel Sheet: <http://dx.doi.org/10.7910/DVN/VYJGQU>

persistentId=doi:10.7910/DVN/VYJGQU

Appendix 5. Raw Data for Captive Migrants from TASTDB 2010

Table A. African embarkations, by region and decade.

Table B. American arrivals, by region and decade.

Excel Sheet: <http://dx.doi.org/10.7910/DVN/9KOWMI>

persistentId=doi:10.7910/DVN/9KOWMI

NOTES

¹ The authors express thanks to Qiyao Wang for important assistance in editing and commentary on this study. James Sharpnack provided useful initial formulation of the multiple-method analysis. In addition, the authors thank Paul Lachance and David Eltis for a detailed and insightful critique of a draft of this article, in which they identified several important errors and clarified their alternative perspective.

² In the Slave Voyages dataset (<http://www.slavevoyages.org>), see the “Downloads” section within the “Voyages Database” tab, and go to the section on “Estimates Spreadsheet.” The “Methodology essay” may be downloaded.

³ For additional detail, see Yun Zhang, “TAST: Technical Report 1,” unpublished paper (Aug. 13, 2013); and Yun Zhang, “Imputation Procedure for Embarkation Count,” unpublished paper (June 10, 2013); both available at <http://www.worldhistory.pitt.edu/Afropop.php>.

⁴ For additional detail on our application of this procedure, see Yun Zhang, “TAST: Technical Report 2,” unpublished paper (August 13, 2013), available at <http://www.worldhistory.pitt.edu/Afropop.php>.

REFERENCES

- Casella, George, and Roger L. Berger. 2001. *Statistical Inference*, 2nd ed. Pacific Grove, CA: Duxbury Press.
- Curtin, Philip D. 1969. *The Atlantic Slave Trade: A Census*. Madison: University of Wisconsin Press
- Curtin, Philip D. 1976. “Measuring the Atlantic Slave Trade Once Again.” *Journal of African History* 17:595-605.
- Eltis, David. 1979. “The Direction and Fluctuation of the Transatlantic Slave Trade, 1821-1845: A Revision of the 1845 Parliamentary Paper,” in Henry A. Gemery and Jan S. Hogendorn, eds., *The Uncommon Market: Essays in the Economic History of the Atlantic Slave Trade* (New York: Academic Press), 273-301.
- Eltis, David. 1983. “The Volume of the Transatlantic Slave Trade, 1761-1867,” unpublished paper cited in Eltis 1987a, pp. 244-248.
- Eltis, David. 1987a. *Economic Growth and the Ending of the Atlantic Slave Trade*. New York: Oxford University Press.

Eltis, David. 1987b. "The Nineteenth-Century Transatlantic Slave Trade: An Annual Time Series of Imports into the Americas Broken down by Region." *Hispanic American Historical Review* 67: 109-138.

Eltis et al. 1999. SPSS dataset available online at <http://www.slavevoyages.org/voyage/download>; identified in the text as "TASTDB-1999."

Eltis, David E., Stephen D. Behrendt, David Richardson, and Herbert S. Klein, ed. 2000. *The Transatlantic Slave Trade: A Database on CD-ROM*. New York: Cambridge University Press.

Eltis et al. 2008. SPSS dataset available online at <http://www.slavevoyages.org/voyage/download>; identified in the text as "TASTDB-2008."

Eltis et al. 2010. SPSS dataset available online at <http://www.slavevoyages.org/voyage/download>; identified in the text as "TASTDB-2010."

Eltis, David, and Paul F. Lachance. 2010. "Estimates of the Size and Direction of Transatlantic Slave Trade." <http://www.slavevoyages.org/documents/download/2010estimates-method.pdf>.

Inikori, J. E. 1976. "Measuring the Atlantic Slave Trade: An assessment of Curtin and Anstey." *Journal of African History* 17:197-233.



Articles in this journal are licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by the University Library System of the University of Pittsburgh as part of its D-Scribe Digital Publishing Program and is cosponsored by the University of Pittsburgh Press.