Research Note on Captive Atlantic Flows
Estimating Missing Data by Slave-Voyage Routes

Patrick Manning
University of Pittsburgh
pmanning@pitt.edu

Yu Liu
University of Pittsburgh
Yul125@pitt.edu

Abstract

This essay provides new estimates of the number of captives carried in the Atlantic slave trade during each decade from the 1650s to the 1860s. It relies on two categories of known data—on the routes of voyages and the numbers of captives recorded on those voyages—as a basis for estimation of missing data and totals of captive flows. It uses techniques of Bayesian statistics to estimate missing data on routes and flows of captives. As a framework for the Bayesian estimates, it focuses on analysis of 40 distinct routes linking the African coast to the Americas and traces the captive flows—that is, the number of captives embarked on or disembarked from voyages along those routes. The dataset that provides the basis for this research note is available at: https://doi.org/10.7910/DVN/6HLXO3.

Keywords: Slave trade, Voyages, Routes, Captive flows, Missing data, Bayesian estimates, Comparative methods and results

Routes of Atlantic Slave Trade

The underlying data on Atlantic slave trade are documented most fully through individual voyages: for each of some 35,000 known voyages, data have been assembled systematically (Eltis et al., "Slave Voyages"). From this base, various scholars have aggregated and interpreted the voyage-based data in various fashions to provide interpretations of aspects of the historical slave trade. Among the frameworks within which the data have been summarized are frameworks by individual voyage, by time period, by national carrier, by region of African departure, by region of American arrival, and by routes linking regions of African departure and American arrival. (Curtin 1969, Lovejoy 1982, Manning 1990) Our approach emphasizes systematic analysis of routes linking African and American regions.

The underlying voyage-based data, while ample, are incomplete in two different ways. First, there were additional voyages for which records have not survived.[1] Second, data are missing or incomplete for the voyages that are known. As a result, a great deal of effort has gone into estimating the magnitudes of missing data on voyages and on the captives carried on those voyages. The data have generally been reported, with some variation, in terms of conventional regions of African origin and American delivery of African captives.

In this study, we focus on the second type of incompleteness. We do so by linking the African and American regions into routes based on pairs of regions, to yield a significantly different perspective on the process of slave trade. To that degree, this essay adds a new method to those that have been applied earlier. It provides an additional opinion to be compared to past estimates of the volume of the Atlantic slave trade, and it provides additional insights about the character of the Atlantic slave trade through its organization of data in terms of routes. In our previous study within this route-based framework, we conducted a qualitative and descriptive analysis, portraying the Atlantic slave voyages through *routes* by which slave vessels traveled from regions on the African coast to regions in the Americas, by decade from the 1650s to the 1860s. The previous study and the present essay both rely on the WHCDB-2017 database, a modified version of variables selected from the TASTDB-2010 dataset of Eltis et al. (Manning and Liu 2019, Eltis et al. 2010). For ease of comparison with previous work, we based our routes on conventional regions as interpreted by Eltis et al. in the TASTDB-2010. We divide the African coast into eight regions of departure of slave voyages, and divide the Americas into four regions of arrival (plus those

---

[1] The Slave Voyages website (www.slavevoyages.org), online since about 2008, has undergone repeated revisions, including a 2019 revision described in the "About" section. Earlier versions of the site, while not readily retrievable, have presumably been archived. For instance, in 2016 it was proposed that a further set of voyages, documented only indirectly, would be included in the dataset, but this step seems not to have been implemented.

voyages that did not reach the Americas)—yielding a total of 40 possible routes for slave voyages.[2] Table 1 shows our labeling of the routes as pairings of regions: thus, route 1-2 is from Senegambia to the Caribbean; route 5-4 is from the Bight of Benin to Brazil.

Table 1. Regions of embarkation and arrival for Atlantic slave trade.

Code	Embarkation Region	Code	Arrival Region
1	Senegambia	1	Mainland N. America
2	Sierra Leone	2	Caribbean
3	Windward Coast	3	Spanish Mainland
4	Gold Coast	4	Brazil
5	Bight of Benin	5	Africa
6	Bight of Biafra		
7	West Central Africa		
8	Southeast Africa		

Our previous study, in addition to providing descriptive detail on the routes in general and on specific routes, reached three principal conclusions (Manning and Liu 2019, 460). First, of the 40 routes that we documented, ten of them accounted for 85% of the slaving voyages for which the routes were known, in the period from the 1650s to the 1860s. These ten principal routes are displayed in Map 1, which labels each route by its code number, indicating the number of known voyages along each of these routes.

Second, we showed that, within each route, the average numbers of captives embarked and the numbers disembarked were remarkably stable for most of the routes over time, from the 1650s to the 1840s. Third, we showed that a slight simplification of the second observation—the assumption that the average number of captive embarkations and arrivals for each route remained *unchanged* for each route—yielded a remarkably precise estimation of the total embarkations and arrivals of captives carried on slave ships for which we have full documentation (Manning and Liu 2019, 462–64).
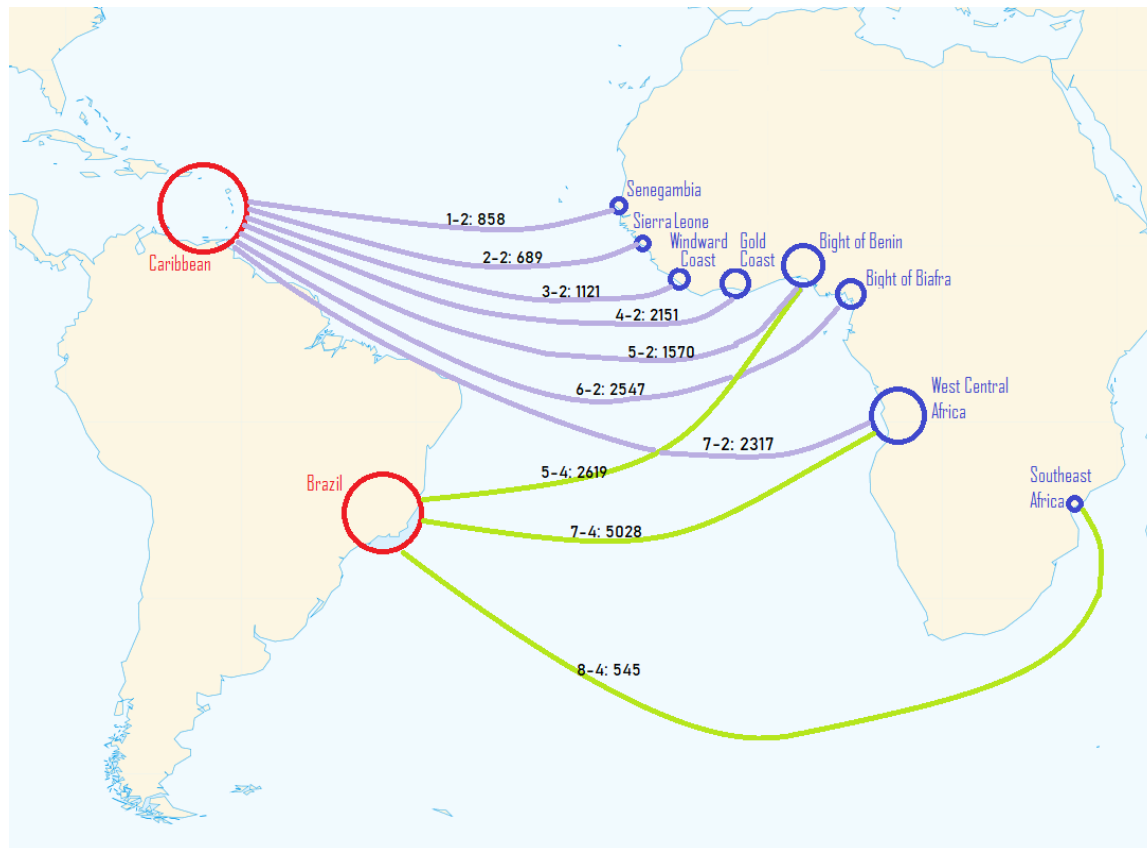
Known vs. Missing Data

Our first step is exploring known data on the Atlantic slave trade, consisting of voyages for which regions of departure and arrival were known, as well as voyages for which we have data on the number of captives embarked, disembarked, or both. Figure 1 summarizes known data on over 33,000 slave voyages, by decade from the 1650s to the 1860s, at five levels of documentation.

In this study, we expand the relationships that we found among known data, extending the analysis to voyages for which data were missing—on the route, the numbers of captives carried

---

[2] The dataset we used to estimate missing data on captive flows is available on the World Historical Dataverse website, at: https://doi.org/10.7910/DVN/6HLXO3.

per voyage, or both. For this analysis, we need to be quite specific on the meaning of "missing data." Known voyages are those listed in the WHCDB-2017 dataset. Missing data consist of specific information on known voyages for which documentation is not available. Missing data on routes are cases of known voyages for which the region of departure, the region of arrival, or both are unknown. Missing data on captives are cases of known voyages for which the number of

Map 1. Top 10 slave routes, with number of known voyages, 1650s – 1860s.

Source: Manning and Liu (2019), 460.

embarking captives, the number of arriving captives, or both are unknown. Our task in this study is to model the missing data on routes and, thereafter, to model the missing data on captives. At the conclusion we will be able to confirm the total of voyages for which routes are known and add the voyages for which routes have been modeled, so that we will have an estimation of the route along which every known voyage traveled. In addition, we will be able to confirm the numbers of captives reported on known voyages, and estimate the numbers of captives aboard voyages for

which they were not reported, so that we will have an estimation of the total number of captives who traveled on all known voyages of each route, by decade.

Figure 1. Total known voyages by decade.[α]
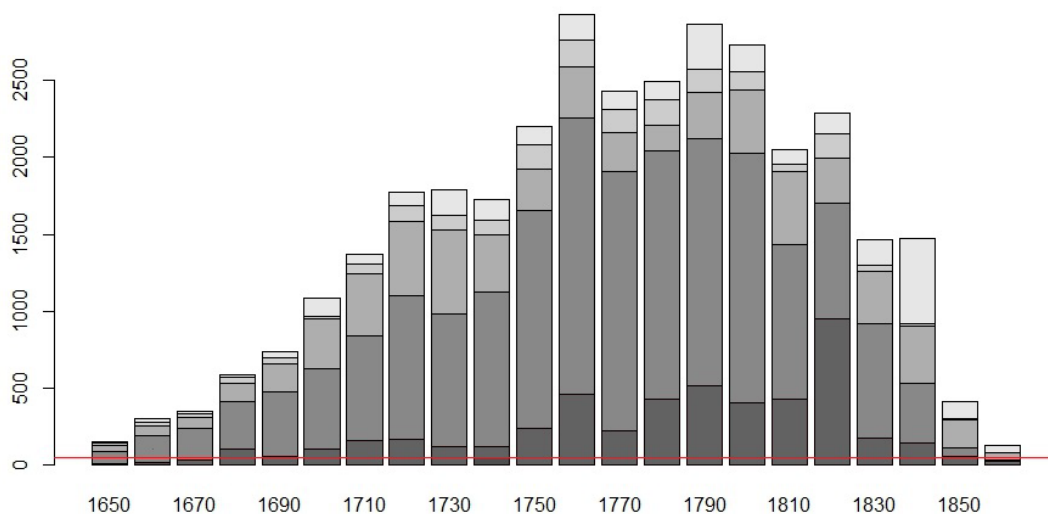
[α] Sections within each bar are defined as follows, from bottom to top: complete region and captive flow data, complete region and incomplete captive flow data, arrival region only, embarkation region only, no information on region. Source: Data0 (See table 2).

Modeling Missing Data in Principle

The steps in our work involve imputing the missing values for routes and captive flows in the WHCDB-2017 dataset, to obtain estimates of embarkations and arrivals of captives in the transatlantic slave trade by route and by decade from the 1650s to the 1860s. We are interested in two key types of statistics: the route distribution (voyage pattern) per decade, and the expected embarkation and disembarkation population per voyage in each route. The proportion of voyages for which we lack information on routes is small (roughly 32% of the voyages); the proportion of voyages for which we lack information on both captive embarkations and arrivals is large (about 85% of voyages). Thus, while both the models of routes and of captive flows are essential to our imputations, the imputation of regions should be more reliable than the imputation of population. And we find that the region/route information can be a significant factor in estimating the captive flow of a voyage. To estimate the above parameters, we propose a **two-stage estimation strategy.** First, we conduct a **regional imputation**, to complete the missing regional information for all voyages in the database to get a complete pattern of routes within voyages for each decade. Second, we conduct a **population imputation,** building a Markov Chain Monte Carlo (MCMC) model to

estimate the expected embarkation and disembarkation population per voyage in each route.[3] The two stages of our imputation of data address missing documentation of routes and missing data on flows of captive population. Combining the route and captive flow information, we will have a picture of the volume and distribution of Atlantic slave trade from the 1650s to the 1860s.

In the first stage, we model the known voyages so as to impute missing routes. We observe that the proportion of voyages with complete information on regions is relatively high, i.e. 68% for the whole database level and above 60% in most decades. At the same time, we realize that we cannot assume that the missing data are completely randomly distributed among the routes, as the distribution is demonstrably non-random in several regards.[4] Even though we fail to find effective factors to explain most patterns of missing data, we find that—given the relatively high proportion of complete regional data—we are comfortable in comparing the distribution of known captive embarkations and arrivals for ships with unknown routes to the pattern for known routes, thereby gaining at least a partial basis for imputing unknown routes. The basic idea is to expand the current distribution of documented records to include estimations for unknown records in each decade.[5] In this step, we rely only on documented routes and not on other explantory factors.[6] Further, thanks to the ample regional data for each decade, the imputation can be performed decade by decade, thus accounting for the fact that the distribution of routes changed significantly over time. In a nutshell, the first step in our analysis is thus the imputation of all missing routes, so that all voyages in the analysis are attributed either known or imputed routes.[7] Success in this imputation gives us clearly identified routes for each voyage, whether the route was documented or imputed.

In the second stage, we turn to modeling captive flows to impute the missing population information. We focus on three analytical points. First, on the distribution of captive flows for each of the various routes, we found that the route-specific averages of embarkation and arrival varied among routes. These variations in average captive flow per route can be seen by comparing the 40

---

[3] Markov Chain Monte Carlo (MCMC) analysis provides a systematic and comprehensive method for estimating the parameters for the dataset containing missing data; it imputes the missing data at the same time. It also provides the inference interval along with the estimation.

[4] For instance, absence of data on routes and on captive flows clearly depended on time and on the national carrier of slave voyages.

[5] Another advantage of this algorithm that it will not result in revision of known data. That is, it will not move any known voyages from one route to another.

[6] In fact, there are voyages with population information but incomplete information on region. We did some adjustment for those cases. Details can be found in later sections.

[7] For further details, see the section below entitled "Imputation, Stage 1."

routes shown in Figure 4.[8] Second, averaging the numbers of embarkations, arrivals, and losses per voyage shows that variations within a given route can usually be neglected. Therefore, we aggregate the data in terms of route and focus on the population characteristics of each route instead of a single voyage or a single decade. Meanwhile, considering that the proportion of voyages in the dataset with documented captive flows is less than 20%, we leverage the assumption of constant captive flows in each route among decades to maximize the power of completed population records. Third, distribution of routes is quite concentrated. The top ten routes, for which we have confidence in the validity of our estimates, include more than 80% of voyages, which therefore provides us with confidence in our overall estimation. Even the routes that we label as "weakly conforming" to our assumption of stable captive flows yield results that fit with our framework.[9]

To summarize this discussion of our second-stage analysis, we define our model of captive flows by assuming that the expected numbers of embarkations, arrivals, and loss rates remain constant across decades for each of the 40 routes, and thus may be taken as estimates of parameters identifying the expected number of captives for each route. These parameters, when multiplied by the numbers of voyages, for each route and each decade, will yield our estimates of the numbers of embarkations, arrivals, and loss rates for voyages where data are missing. Relying on this model, we reaffirm our conclusion that the variation in slave trade over time was mainly in the number of voyages per decade and, especially, in the distribution of voyages among competing routes of slave trade, while the number of captives per voyage along a given route varied only slightly. By finishing this step, we can replace all of the missing information in the database, WHCDB-2017, with imputed data for routes and populations, giving us a coherent estimate of the totality of the Atlantic slave trade from 1650 through the 1860s.

Applying these principles for modeling our existing and missing data with relevant statistical techniques, we conduct our imputation in two stages. Stage 1 is to estimate unknown routes using a multinomial model for the distribution of routes, to give us a full set of routes.[10] Stage 2 is to estimate unknown flows of captives, using a Poisson model for the distribution of embarkation population and a binomial model for the distribution of arriving populations. This gives us a nearly full set of estimates for embarkations, arrivals, and rates of loss at sea.

---

[8] For instance, embarkations on the African coast were smaller for Senegambia than for Angola; arrivals on the American coasts were larger for the Caribbean than for North America. In an important further point, the variations in average captive flows *within a given route* are smaller and more random than the variations in average captive flows *among routes*.

[9] For further details on our handling of these weakly conforming routes, see the section below, "Imputation, stage 2."

[10] At the last step in Stage 1, all the known flows of captives are linked to now-known routes.

Table 2: Sub-datasets for known and imputed data in WHCCDB-2017 dataset.

Before Imputation			After Regional Imputation			After Population Imputation		
Category	Definition	Voyages	Category	Definition	Voyages	Category	Definition	Voyages
Data0 (Fig. 1)	"Full original dataset." All documented slave voyages in WHCCDB-2017	33,345						
Data1 (Fig. 2)	"Routes known" – **Documented**: departure and arrival regions	22,803	**Data4** (Fig. 3)	"Regions known & imputed." ***Documented and imputed:*** departure and arrival regions	33,345			
Data2a	"Embarkations and Routes known." **Documented**: departure and arrival regions, departure populations	7,229	**Data5a** (Fig. 4)	"Embarkations known." ***Documented and imputed:*** departure and arrival regions. **Documented**: departure populations	7,318			
Data2b	"Arrivals and Routes known." **Documented**: departure and arrival regions, arrival populations	14,008	**Data5b** (Fig. 4)	"Arrivals known." ***Documented and imputed:*** departure and arrival regions. **Documented**: arrival populations	17,514			
Data3	"Captive flows and Routes known." **Documented**: departure and arrival regions, paired departure and arrival populations	4,899	**Data6**	"Regions & captive flows known." ***Documented and imputed:*** departure and arrival regions. **Documented**: paired departure and arrival populations	5,125	**Data7** (Table 3)	"Regions & captive flows known & imputed." ***Documented and imputed:*** departure and arrival regions, paired departure and arrival populations	33,345

Table 2 lists the full set of datasets and sub-datasets in our analysis. Data0 through Data3 are sets of original information on documented voyages. Data4, 5a, 5b, and 6 are generated through stage 1 of the imputation. To indicate the expansion in available information that is achieved with imputation of missing data on routes and captive populations, Table 2 shows the number of voyages that result for each of the post-imputation data sets (from Data4 to Data6), as compared with the original number of documented voyages (Data0 to Data3).

Imputation, stage 1: estimation of all routes by multinomial model
We use a multinomial model to impute the allocation of the remaining 10,500 documented voyages among the 40 routes (the difference between Data0 and Data1). The multinomial distribution is a classic model for categorical data, which is suitable for modeling the distribution of routes. The resulting voyage pattern by route and by decade, pre- and post-imputation, is shown in Figures 2 and 3. We emphasize that, in this step, we conduct the imputation decade by decade, to account for the changing distribution of routes across decades. Since more than 60% of routes are known for each decade, we are confident in modeling the distribution of routes independently by decade. We assume that the voyage pattern follows the multinomial distribution $multinomial(N, p_{ij})$, $\sum_{ij} p_{ij} = 1$, where N is the total number of voyages and p_{ij} is the probability of a voyage from region i to region j. Let X be the embarkation region code and Y be the embarkation region code. We have three types of observations:

$$\begin{cases} Z_{ij}: \text{ the number of voyages from region } i \text{ to region } j. \\ X_i: \text{the number of voyages from region } i \text{ to unknown region.} \\ Y_j: \text{the number of voyages from unknown region to region } j. \end{cases}$$

The estimator of p_{ij} is

$$p_{ij} = \frac{Z_{ij}}{n} + \frac{X_i}{n}\frac{Z_{ij}}{Z_{i\cdot}} + \frac{Y_j}{n}\frac{Z_{ij}}{Z_{\cdot j}},$$

where $Z_{i\cdot} = \sum_j Z_{ij}$, $Z_{\cdot j} = \sum_i Z_{ij} \sum_{i\,j}(Z_{ij} + X_i + Y_j) = n.$

The estimator of the voyage pattern is: For the voyages with only embarkation information, namely, the X_i voyages from region i, the distribution of the disembarkation region is $Y_j \sim multinomial(X_i, \frac{p_{ij}}{p_{i\cdot}})$. For the voyages with only disembarkation information, namely, the Y_j voyages from region j, the distribution of the disembarkation region is $X_i \sim multinomial(Y_j, \frac{p_{ij}}{p_{\cdot j}})$.

For the voyages with neither information, namely, the N voyages without regions, the distribution of the route is $Z_{ij} \sim multinomial(N, \ p_{ij})$.

In the full dataset of 33,345 voyages, we have 22,803 voyages with complete regional information. In addition, the dataset has 1708 voyages with only embarkation regions, 6048 voyages with only disembarkation regions, and 104 voyages with neither of them. The missing regions can be assigned by the above model. Further, we find 3506 voyages containing captive disembarkation but incomplete route information, and there is a strong correlation between the embarkation population and the region—that is, a voyage with high disembarkation population has a better chance of having departed from certain specific regions. To capture this feature, we divide the dataset into two subsets: the ones with captive disembarkations less than or equal to 300, and the ones with captive disembarkations above 300.[11] We assume the multinomial

Figure 2. Voyages by route and decade (for 33,345 voyages, after regional imputation).
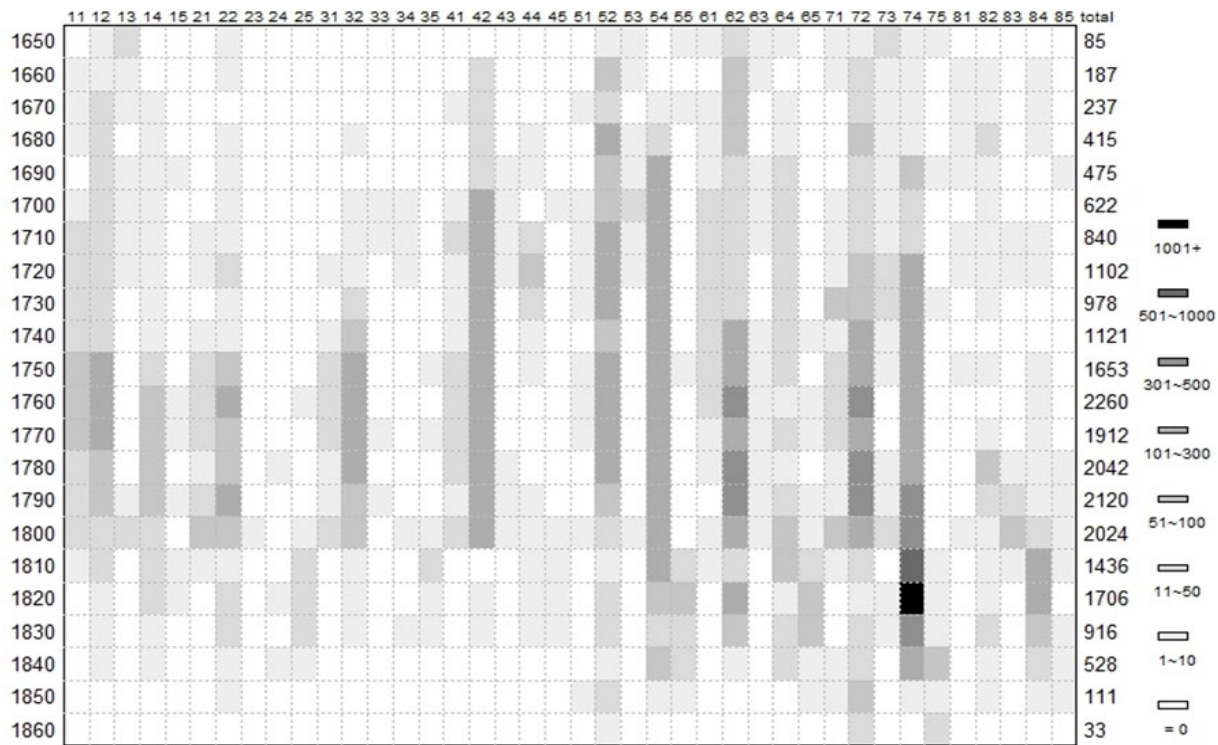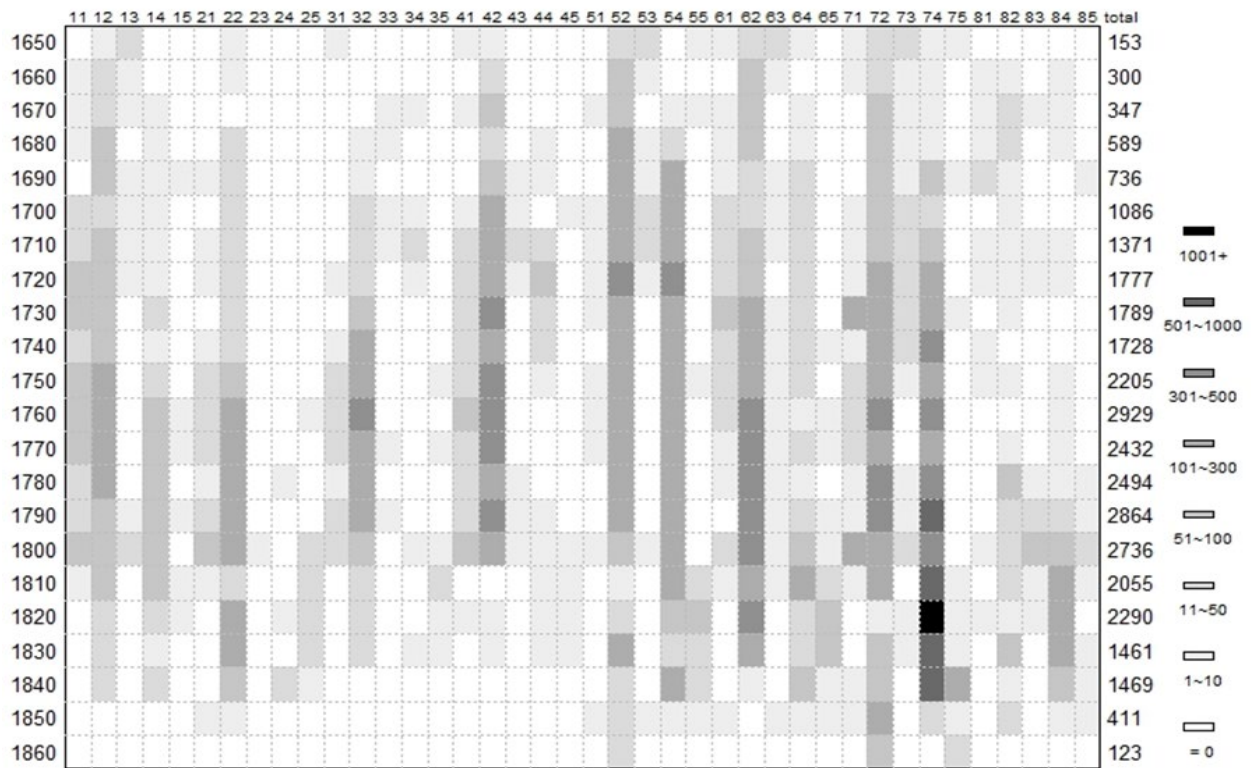

Source: Data1

---

[11] We focus only on the impact of disembarkation population and ignore that of embarkation population since the number of voyages with only embarkation population, but incomplete route info is fewer than 500 in total (the precise number is 428).

parameters are different for those groups. With 22 decades in total and three groups (missing arrival population, arrival population <= 300, arrival population > 300) within each decade, we divide the dataset into 66 subsets, although some of them are empty. We apply our model (implemented in the R programming language) to the 66 subsets separately in order to get the final imputation of regions and routes for all the voyages (See R code in Appendix). Figure 3 shows the imputed voyage pattern for all voyages, after the above procedure: it summarizes the numbers and directions of voyages by decade for all 33,345 voyages, in contrast to Figure 2, which gives parallel data for the 22,803 voyages. Left-hand column shows decade; right hand two columns show total voyages per decade and the key (range of voyage numbers for each shading). Note differences in numbers of voyages shown in keys to Figs.2 and 3. Visual inspection confirms that the overall pattern of voyages is preserved, even though one third of the routes in Figure 3 are imputed.

Figure 3. Voyages by route and decade (for 33,345 voyages, after regional imputation).
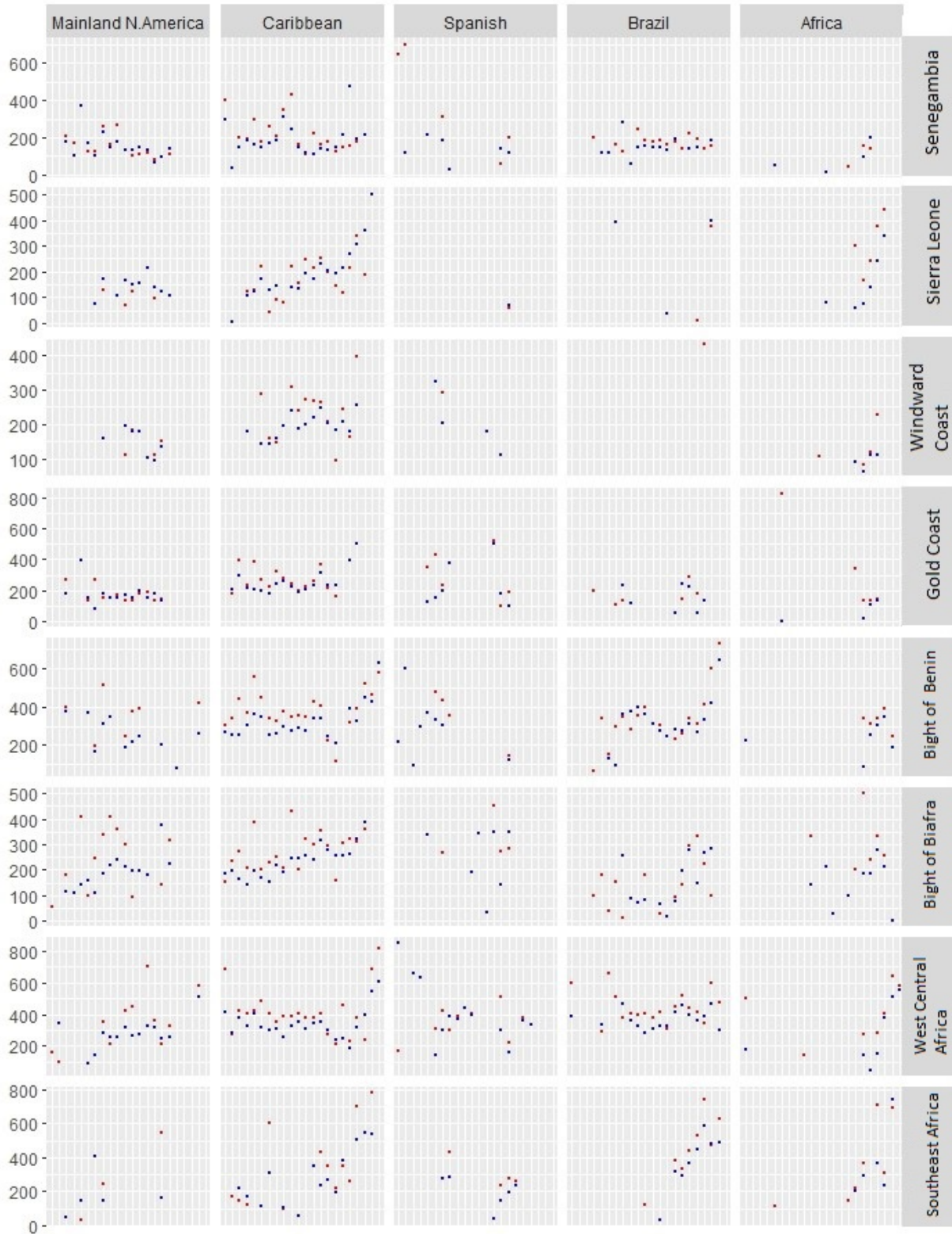

Source: Data4

Imputation, stage 2: MCMC estimate of missing and total captive populations

Based on the results of stage 1, we know the route information for all known voyages. We now allocate all known captive flows among the complete set of voyages after imputation: Figure 4 shows the known captive flows after imputation of routes: this is the database after stage 1 of the imputation. Figure 4, showing all known captive flows for the routes of 33,345 known and imputed voyages, is an expanded version of an earlier figure showing a smaller number of captive flows for 22,803 known voyages (Manning and Liu, 2010, 461). Comparison of the two figures confirms the pattern of stability in the captive flows in each case.

From Figure 4, we observe that the departure population and the arrival population remain at the same level over time, especially from the 1650s to the 1810s. Based on this observation, we propose an assumption for our model: the population numbers for each route (i.e. departure population, arrival population, and loss rate) stay the same across time. This assumption, while quite novel in slave-trade studies, is nonetheless plausible. It is the simplest assumption: every voyage along a given route is the same. This is a natural choice for simplifying the model and it matches our observation of the dataset as well. While more complicated assumptions can be considered, they would increase the model's complexity or require analyzing the data in smaller batches, thus reducing available sample sizes and most likely narrowing the statistical validity of results. To back up this default statistical logic, we argue that practical and empirical considerations back up this assumption, in that each route has a unique collection of population characteristics and also of other characteristics. For population characteristics we argue that, in practice, the slave trade established a firm pattern for each route, both in embarkations and arrivals. The other characteristics of each route included the regional system for supply of captives, the size of ships, the participation of European and African merchants in trade, the level of demand for captives in varying regions of the Americas, and the sailing conditions by route, including the timing of voyages by route. All of these other factors differed for each route and may have combined to give a stable character to the slave trade for each route (Miller 1981).

For two groups of voyages, discrepancies among decennial flows appear relatively serious. That is, within these groups, the means for the key variables have a greater variance than that for the dataset overall. We label these as voyages that are "weakly conforming" rather than "strongly conforming" to assumptions of constant flow along each route. In total, the number of voyages that we describe as weakly conforming to our assumption of constant flow is less than 5% of all known voyages. There are several categories of "weakly conforming" routes. First, the voyages during the decades of the 1850s and 1860s (especially for departures from the Bight of Benin and West Central Africa). In those cases we have made an adjustment of our imputation of slave-trade parameters in route 5-2 and 7-2, using a step function to capture the change of the captive flows in 1850s to 1860s.

Figure 4. Known captive flows for <u>all</u> voyages by route. Average embarkation (red) and arrival (blue) populations per voyage, by route by decade.

Source: Data 5a (7318 voyages) and 5b (17,514 voyages).

This was the era of fully illegal slave trade, during which the previous conditions of routes were disrupted by anti-slavery squadrons; in addition, the number of fully documented routes was well under 100 for the 1850s and 1860s, so that the small samples are not dependable for estimation. The second problem area is that of the routes from Southeast Africa to all regions of the Americas (routes 8-1 to 8-4 in Figure 4), from the 1780s to the 1860s: the combination of weak documentation and long voyages yields relatively erratic patterns each of for these routes. Third, an additional variation is found with routes terminating in "Africa" from the 1810s to the 1860s, consisting mainly of voyages captured by anti-slavery squadrons, for which the captives were embarked mainly in the Bight of Benin, Bight of Biafra, and West-Central Africa and disembarked principally in Sierra Leone and St. Helena. The patterns of these voyages, represented in the right-hand column of Figure 4, appear on inspection to have a relatively high variance. Fourth, besides the above cases, certain routes include less than 10 documented records for which we have complete data on captive flows. This is more of a limitation of the data in the database than a challenge to our constant-captive-flows assumption—the number of such routes (fifteen) may seem large, but in fact all those routes account for less than 2% of the voyags in the dataset.[12] In practice, of these four types of weakly conforming routes, we found that only the first case required us to adjust our procedure, which we did by calculating separate parameters for routes 52 and 72 for the decades of the 1850s and 1860s.

Based on inspection of the data, we make three assumptions as given above in our model: for all voyages along each route, across time, we assume a constant expected 1) number of embarkations, 2) number of arrivals, and 3) rate of loss of captives. Since the decades of all voyages are known (because of stage 1 imputation), that leaves only the distribution of captives among routes to be estimated at this stage of the imputation. For voyages on which the embarkation population is unknown, we assume that the data follow the *Poisson distribution*. In probability theory and statistics, the Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event. The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area or volume.[13] We assume that the expected level of embarkation varies for different routes and decades: it may be affected by the vessels, the current economic situation, and even the weather. This expected level of embarkation determines the parameter in the Possion distribution—and we use the average embarkation population per voyage to estimate.

---

[12] The 15 routes meeting this criterion include 1-3, 1-5, 2-3, 2-4, 3-3, 3-4, 3-5, 4-3, 4-4, 4-5, 5-1, 5-3, 6-3, 8-1, 8-5.

[13] The distribution describes the properties of each route rather than each voyage: that is, each voyage within a given route is assumed to have the same number of embarkations, arrivals, and losses.

For voyages in which the arrival population is unknown, but the embarkation population is known, we assume that the arrival population follows the *binomial distribution*. The binomial distribution with parameters **n** and **p** is the discrete probability distribution of the number of successes in a sequence of **n** independent yes/no experiments, each of which yields success with probability **p**; the Binomial-Poisson hierarchy model is the most classic way to model survival. The parameter in the binomial distribution is a measure of the expected survival, in other words, 1 –loss rate.

We divide all the records into four groups, so that the estimation of parameters can utilize the information on incomplete voyages.

D_f , A_f embarkation and arrival population when both are known (5125 voyages).

D_d , A_d embarkation and arrival population when only the embarkation population is known (2193 voyages).

D_a , A_a embarkation and arrival population when only the arrival population is known (12,389 voyages).

D_m , A_m embarkation and arrival population when both are unknown (13,638 voyages).

Thus, D_{full} , A_{full} , D_{dep} , A_{arr} are values we already know, and D_{mis} , D_{mis} , D_{arr} , A_{dep} are values we need to estimate. Our strategy is to utilize all voyages with known data to estimate the parameter in the model, then predict the unknown number by the model. We run the model route by route and the results give each parameter its interpretation. The model for the **known data** is:

$$\begin{cases} D_f \sim Poisson\ (d) \\ A_f \sim Binomial\ (1-l, D_f) \\ D_d \sim Poisson(d) \\ A_a \sim Poisson(d(1-l)) \end{cases}$$

The parameter for Poisson distribution **d** is the expected embarkation population per voyage, for a specific route; and the parameter for binomial distribution **l** is the expected corresponding loss rate. Then we impute the **missing populations** with the following model

$$\begin{cases} A_d \sim Binomial\ (1-l, D_d) \\ D_a \sim A_a + negtive\ binomial\ (l, A_a) \\ D_m \sim Poisson\ (d) \\ A_m \sim Binomial\ (1-l, D_m) \end{cases}$$

The imputation is done by the r-stan procedure. We employ a non-informative prior to fit the model. We run it for 4 chains with random initials and 20,000 iterations for each route, of which the first 5000 iterations are for burn-in each time. As a Bayesian method, MCMC can also be interpreted as a multi-level model. The multi-level structure also explains the over-dispersion for Poisson assumption.[14] Furthermore, the model itself ensures that all the imputed voyages have a positive loss rate.[15]

In the results of the imputation as displayed in Table 3, the parameters are calculated through common procedures for all data, with two exceptions that are discussed below. The table shows the estimated parameter for each route, the 95% confidence range for each parameter, and the number of voyages included in the calculation of each parameter. Overall, parameters are calculated on the basis of 5125 voyages out of the total of 33,345 voyages. We believe this is a statistically adequate basis for projecting embarkation, survival, and disembarkation for the full 33,345 voyages.

The error levels we display in Table 3 are underestimates of error margins. The first source of additional inaccuracy is that our estimation is based on a fixed imputation of regions, so that we do not take into account the "variance" of the distribution of voyages. Secondly, our estimation of totals is based on the captive flow character of each route, yet each single voyage will also be affected by some randomness. The specific interpretation of the margin of error as reported here is that it is the margin of error in the expected total captive flow.

There are two exceptions in our calculation: for "weakly conforming" data within routes 52 and 72, from the Bight of Benin and West-Central Africa to the Caribbean. In these exceptions, parameters are calculated separately (with the same algorithm, but only for the data of the 1850s and 1860s). The highly illegal slave trade to Cuba dominated these routes in the 1850s and 1860s. In Table 3, estimates for routes 52 and 72 are for the period 1650s–1840s, while estimates for routes 52(late) and 72(late) are for the 1850s–1860s. As can be seen, cargoes were larger and survival rates lower for these "late" routes. For other routes and voyages that we identified earlier as potentially "weakly confirming," we found that, in practice, the data conformed surprisingly strongly to our assumption of constant flow. Of the voyages disembarking in "Africa," the WHCCDB-2017 dataset includes 354 such voyages, all in the decades 1810s to 1860s; they came in roughly equal numbers from the Bight of Benin, Bight of Biafra, and West Central Africa and

---

[14] The "non-informative prior" means that we have only the data and no belief on the prior condition. The Poisson distribution assumes that the mean and variance should be equal, but in the data we always find the sample variance to be greater than the sample mean: this phenomenon is called over-dispersion.

[15] The model is run route by route, but there are different types of voyages on each route.

**Table 3. Estimated captive flows for all known voyages:
Parameters for Embarkations, Arrivals, and Loss Rates.**[16]

route	estimated embarkation parameter			estimated survival rate parameter			estimated disembarkation parameter			documented case	with imputed cases	all cases
	mean	2.50%	97.50%	mean	2.50%	97.50%	mean	2.50%	97.50%			
11	157.27	155.78	158.81	0.87	0.86	0.87	136.33	135.11	137.53	67	74	590
12	193.03	191.97	194.09	0.86	0.85	0.86	165.12	164.26	165.99	178	189	1559
13	257.63	249.66	265.67	0.59	0.57	0.61	152.84	148.53	157.19	7	7	70
14	167.46	165.82	169.09	0.91	0.91	0.92	152.70	151.25	154.14	77	78	557
15	99.26	93.45	105.27	0.93	0.90	0.95	91.90	86.47	97.50	4	4	20
21	153.63	151.44	155.87	0.89	0.89	0.90	137.15	135.44	138.85	23	26	265
22	222.45	221.21	223.68	0.91	0.91	0.91	202.35	201.32	203.40	120	136	1329
23	87.60	76.95	99.12	0.82	0.76	0.87	71.73	62.46	81.68	2	2	7
24	354.78	343.58	366.25	0.96	0.95	0.98	341.60	331.46	351.82	1	1	17
25	175.67	171.91	179.50	0.82	0.80	0.83	143.65	140.87	146.43	18	18	75
31	182.18	178.83	185.64	0.87	0.86	0.89	159.37	156.77	162.00	12	14	143
32	241.63	240.36	242.89	0.87	0.86	0.87	209.05	208.17	209.93	120	134	1633
33	301.27	286.26	317.48	0.88	0.84	0.91	264.38	254.56	274.44	1	1	12
34	433.27	392.44	475.70	0.50	0.02	0.97	216.56	9.58	427.97	0	0	22
35	84.41	80.51	88.45	0.97	0.95	0.98	81.45	77.81	85.18	5	5	28
41	163.34	161.37	165.30	0.88	0.88	0.89	144.21	142.47	145.90	58	59	273
42	257.33	256.59	258.08	0.87	0.87	0.87	223.06	222.45	223.67	494	518	3561
43	251.63	244.99	258.30	0.72	0.71	0.74	181.62	176.89	186.40	9	9	45
44	199.21	193.31	205.35	0.81	0.79	0.83	160.63	155.46	165.88	4	4	206
45	247.90	238.00	257.95	0.34	0.31	0.36	83.26	77.46	89.24	5	5	16
51	304.13	296.93	311.35	0.82	0.81	0.83	249.40	244.20	254.70	8	8	43
52	359.09	358.06	360.10	0.82	0.82	0.82	294.03	293.21	294.85	534	555	2470
53	473.73	465.18	482.37	0.66	0.65	0.67	311.72	306.66	316.80	9	9	75
54	338.31	337.10	339.52	0.94	0.94	0.94	317.34	316.27	318.41	140	147	3153
55	284.38	281.70	287.05	0.90	0.90	0.91	256.60	254.10	259.09	112	117	166
61	266.34	262.07	270.60	0.70	0.69	0.71	187.34	185.53	189.13	13	16	307
62	311.90	310.96	312.84	0.82	0.82	0.82	255.52	254.88	256.15	346	380	4043
63	371.05	362.57	379.73	0.77	0.75	0.78	284.56	277.61	291.56	9	9	41
64	219.42	216.80	222.01	0.82	0.81	0.82	179.14	177.09	181.19	27	28	530
65	273.28	270.56	276.04	0.80	0.79	0.80	217.33	214.95	219.74	116	116	156
71	319.08	315.20	322.98	0.82	0.81	0.83	261.16	259.12	263.23	13	18	321
72	352.09	351.21	352.98	0.89	0.88	0.89	312.03	311.31	312.75	483	517	3397
73	364.79	360.63	369.04	0.83	0.83	0.84	303.86	300.51	307.30	26	27	161
74	423.78	423.06	424.50	0.91	0.91	0.91	385.03	384.35	385.71	1416	1434	6339
75	416.44	412.63	420.27	0.89	0.88	0.89	368.97	365.40	372.56	85	85	169
81	307.67	293.23	322.80	0.53	0.50	0.55	162.01	156.37	167.79	2	2	46
82	465.22	460.98	469.46	0.68	0.68	0.69	317.79	315.41	320.15	28	35	332
83	269.95	265.89	273.98	0.71	0.70	0.72	191.30	188.13	194.49	40	43	87
84	528.81	526.58	531.03	0.84	0.84	0.84	445.39	443.42	447.37	249	252	661
85	341.97	334.77	349.24	0.76	0.75	0.77	259.40	254.02	264.88	9	10	45
52(late)	564.40	555.61	573.27	0.86	0.85	0.87	482.94	477.43	488.60	5	7	61
72(late)	686.27	680.77	691.78	0.81	0.80	0.81	553.73	550.30	557.19	24	26	314
									total:	4899	5125	33345

disembarked at Sierra Leone and St. Helena. Remarkably, the parameters calculated for routes 5-5, 6-5, and 7-5 show high precision, consistent with the assumption of constant flow. Similarly, the data for embarkations in Southeast Africa show precise parameters for all routes except to North America. We still need to emphasize that for certain routes, highlighted in Table 3, the number of voyages was very small, so that the estimates for those voyages lack precision.

---

[16] Estimates show 95% confidence range. The 52 and 72 are the parameters of routes 52 and 72 in 1650s–1840s respectively. The 52 (late) and 72 (late) are the parameters of route 52 and 72 in 1850s–1860s respectively. Highlighted rows have less than ten documented cases. Note that this is a summary of Data7.

In the concluding step of our imputations, we multiply each of the parameters from Table 3 by the number of voyages in each decade for each route, as given in Data4.[17] Table 4 summarizes the results of those calculations: it gives estimates for the transatlantic slave trade by decade, giving numbers of known voyages and imputed embarkation and arrival populations and loss rate.

In a further summary of the results of our imputations, Table 5 displays data on the busiest ten routes of the 40 routes in the Atlantic slave trade. The table shows the number of voyages along the ten routes with the largest amount of traffic for six successive periods. Total departures and arrivals for each route in each period are estimated as the number of voyages along that route multiplied by the appropriate embarkation parameter. These ten routes accounted for 83.5% of the voyages and over 85% of the captives in the WHCCDB-2017 dataset, and the remaining 30 routes accounted for the other 16.5% of the voyages.

Table 4. Estimated total flows of captives.

decade	Number of Voyages	Departure Population	Arrival Population	Lost at Sea	Loss rate (%)
1650s	153	51,643	40,246	11,397	22.1
1660s	300	96,373	79,627	16,746	17.4
1670s	347	108,562	90,429	18,133	16.7
1680s	589	188,103	157,360	30,743	16.3
1690s	736	231,716	200,752	30,964	13.4
1700s	1,086	330,266	285,226	45,040	13.6
1710s	1,371	414,688	353,334	61,354	14.8
1720s	1,777	546,370	474,033	72,337	13.2
1730s	1,789	551,579	476,060	75,519	13.7
1740s	1,728	555,432	484,992	70,440	12.7
1750s	2,205	644,396	560,912	83,484	13.0
1760s	2,929	844,433	736,128	108,305	12.8
1770s	2,432	705,598	613,526	92,072	13.0
1780s	2,494	775,874	670,162	105,712	13.6
1790s	2,864	904,733	791,061	113,672	12.6
1800s	2,736	843,222	729,161	114,061	13.5
1810s	2,055	744,918	658,577	86,341	11.6
1820s	2,290	857,664	755,366	102,298	11.9
1830s	1,461	530,665	458,382	72,283	13.6
1840s	1,469	563,706	506,095	57,611	10.2
1850s	411	241,556	196,355	45,201	18.7
1860s	123	72,356	60,069	12,287	17.0
	33,345	10,803,853	9,377,853	1,426,000	

Source: Table 3 and Data7

---

[17] See the Appendix, which shows the location of this and other datasets discussed in this essay.

Table 5. Number of voyages for major routes, 1650s-1860s.

route name	route code	number of voyages per route, by decades							total voyages	departures per voyage	total departures	arrivals per voyage	total arrivals
		1650s-1690s	1700s-1720s	1730s-1750s	1760s-1780s	1790s-1810s	1820s-1840s	1850s-1860s					
W Central Af-Brazil	74	117	281	792	875	1819	2421	34	6,339	424	2,686,350	385	2,395,655
W Central Af-Caribbean	72	273	257	715	1030	925	197	314	3,711	352\686	1,411,148	312\554	1,233,820
B Biafra-Caribbean	62	299	176	616	1252	1078	619	3	4,043	312	1,261,030	256	956,664
B Benin-Brazil	54	223	857	555	553	724	239	2	3,153	338	1,066,698	317	929,798
B Benin-Caribbean	52	404	757	507	455	169	178	61	2,531	359\564	921,134	294\483	755,643
Gold Coast-Caribbean	42	267	829	913	940	602	10	0	3,561	257	916,352	223	734,749
Windward Coast-Caribbean	32	5	53	393	878	257	47	0	1,633	242	394,579	209	340,331
Southeast Africa	84	5	4	7	10	233	398	4	661	529	349,669	445	294,145
Senegambia-Caribbean	12	183	227	321	519	237	72	0	1,599	193	300,931	165	227,211
Sierra Leone-Caribbean	22	31	56	116	459	304	356	7	1,329	222	295,630	202	262,651
sum of the above 10 routes		1807	3497	4935	6971	6348	4537	425	28,560	--	9,603,521	--	8,130,667
sum of all the routes		2125	4234	5722	7855	7655	5220	534	33,345	--	10,803,853	--	9,377,853
top 10 routes as % of all routes		85.0%	82.6%	86.2%	88.7%	82.9%	86.9%	79.6%	85.7%	--	88.9%	--	86.7%

Source, Table 3 and Data7

Conclusion

Figure 5 presents the data from Table 4 in graphical form. A comparison with Figure 1 shows that the overall pattern of known rise and decline in transatlantic slave trade, as shown in the documented records of slave trade, is confirmed by these imputations of missing data in Figure 5. In addition, Figure 5 compares our new imputations (solid lines) with the previous estimates of Eltis and Lachance as given in the Slave Voyages website (dotted lines), showing that our estimates average roughly 10% larger in volumes of embarkations and arrivals than these previous estimates.

With the completion of these estimates, we now have three types of estimates of the volume of the Atlantic slave trade. All three rely on the TAST-2010 database, but they are based on different ways of estimating missing data. The estimates by Eltis and Lachance, based especially on linear equations estimating 155 types of specific situations, and give a total of 9.7 million African embarkations, 1650-1870 (Eltis and Lachance 2010). The estimates by Manning, Zhang, and Bowen, based on MCMC analysis of 80 separate embarkation and arrival categories by region and by decade, give a total of 10.3 million African embarkations, 1650-1870 (Manning, Zhang, and Yi 2015). The estimates here, based on MCMC analysis of combined embarkations and arrivals for 40 routes by region and by decade, give a total of 10.8 million African embarkations, 1650-1870.[18] In each succeeding case, the imputations are based on progressively larger numbers of cases in each imputation, thus increasing the precision of the analysis. In addition, the two MCMC analyses allow estimation of error estimates. In sum, the combination of the separate

---

[18] Table 4. Also note that the Slave Voyages site has updates after TASTDB-2010 with revisions and expansions, which could be explored with the methods used here.

analyses shows that the various approaches are relatively consistent with each other, but that different statistical techniques provide different perspectives on the size of missing data. In

Figure 5. Overall estimation of Atlantic slave trade volume.
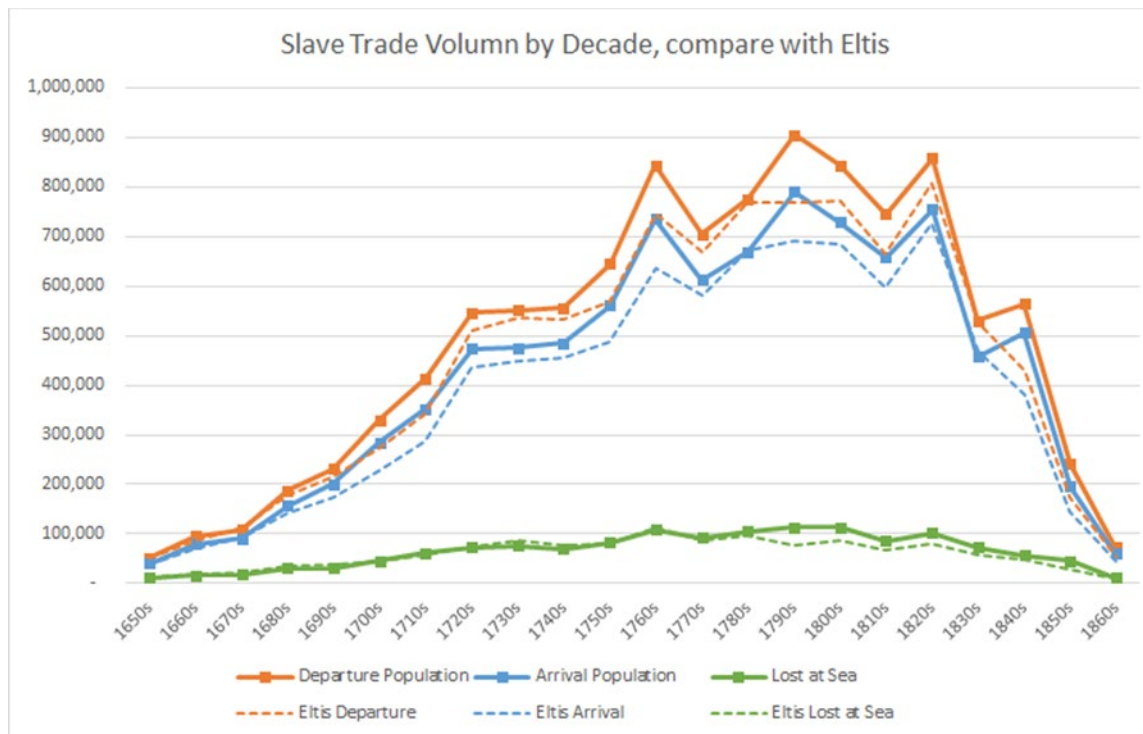
particular, we argue that our approach on analyzing missing data through the framework of routes shows the significance of the routes of slave trade, not only as a technique for estimation, but also the routes as a framework through which to consider the complex history of the Atlantic slave trade. The analysis by routes included the largest number of cases in each imputation, and arguably gives the most statistically valid result. As a final observation, we emphasize a result that may seem surprising but that confirms the logic of our analysis. That is, the aggregated results show a declining *average loss rate* of captives over time, even though we have assumed that the number and proportion of *losses* per individual voyage remained unchanged over time for each route. This means that the captive loss rates declined over time not because of improved health conditions on the vessels but because, with time, larger proportions of voyages took low-mortality routes.

About the Authors: Patrick Manning is Andrew Mellon Professor of World History, Emeritus, at the University of Pittsburgh. He has published widely on the demography of slavery and the slave trade. Yu Liu completed her PhD in Statistics at the University of Pittsburgh in 2018 and is employed in the private sector. She has worked since 2016 on modeling the demography of the slave trade.

Disclosure Statement: Any conflicts of interest are reported in the acknowledgments section of the article's text. Otherwise, authors have indicated that they have no conflict of interests upon submission of the article to the journal.

Appendix

Datasets Available in the World-Historical Dataverse

The main datasets associated with this analysis are available for download at the World-Historical Dataverse, sponsored by the Collaborative for Historical Information and Analysis (CHIA), at https://doi.org/10.7910/DVN/6HLXO3. Sections of this Excel file include:

   A. "Data0: Full original dataset, WHCDB-2017." (33,345 voyages)
   B. "Data1: Routes known." (22,803 voyages)

Note that Data2a (7229 voyages), Data2b (14,008 voyages), and Data3 (4899 voyages) may easily be created from Data1.

   C. "Data4: Regions known and imputed." (33,345 voyages)"
   D. "Data5a: Documented departure populations after regional imputation." (7318 voyages)
   E. "Data5b: Documented arrival populations after regional imputation." (17,514 voyages)
   F. "Data6: Paired departure and arrival populations, documented and imputed." (5125 voyages)

These are the data used in the MCMC estimations of the parameters shown in Table 3.

   G. "Table 3. Parameters for Embarkations, Arrivals, and Loss Rates."

Each of these parameters, when multiplied by the appropriate number of voyages in Data4, yields the number of persons embarked, arriving, or lost at sea for a given route and decade.

   H. "Table 5. Embarkations, Arrivals, and Lost at Sea, by Route and by Decade." This table displays the results of the calculations indicated just above. Summaries are shown in Table 4 and Figure 5.
   I. Data7. "Capture departure and arrival flows by route and by decade, 1650s-1860s."
   J. "Code steps 1-4." This is the R-language code for replicating the simulation of missing data routes and captive flows.

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.

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

Eltis, David, et al. "Slave Voyages." https://www.slavevoyages.org.

Eltis et al. 2010. SPSS Dataset. http://www.slavevoyages.org/voyage/download; identified in the text as "TASTDB-2010."

Eltis, David, and Paul Lachance. 2014–15. "Comments on Manning, Zhang, and Yi's 'Volume and Direction of the Atlantic Slave Trade, 1650-1870'." *Journal of World-Historical Information*, 2–3: 150–53.

Lovejoy, Paul E. 1982. "The Volume of the Atlantic Slave Trade: A Synthesis." *Journal of African History*, 23: 473–501.

Manning, Patrick. 1990. *Slavery and African Life: Occidental, Oriental, and African Slave Trades*. Cambridge: Cambridge University Press.

Manning, Patrick, and Yu Liu. 2019. "Routes of the Atlantic Slave Trade: Revised Framework and New Insights," *Journal of World-Systems Research* 25: 449–466.

Manning, Patrick, Yun Zhang, and Bowen Yi. 2014–15. "Volume and Direction of the Atlantic Slave Trade, 1650-1870: Estimates by Markov Chain Carlo Analysis." *Journal of World-Historical Information* 2–3: 127–49.

Miller, Joseph C. 1981l "Mortality in the Atlantic Slave Trade: Statistical Evidence on Causality." *Journal of Interdisciplinary History* 11: 385–423.