Supplementary Materials: When Can Multiple Imputation Improve Regression Estimates?

Vincent Arel-Bundock[*]   Krzysztof J. Pelc[†]

September 7, 2017

_____

[*]Department of Political Science, Université de Montréal, vincent.arel-bundock@umontreal.ca

[†]Department of Political Science, McGill University, kj.pelc@mcgill.ca

1 Supplementary materials: Replication

To illustrate the benefits of the guidelines proposed in text, we replicate one of the political economy studies criticized in Lall (2016).[1] One of the key takeaways from our note and from the literature on MI is that proper use of this technique requires serious consideration and a deep understanding of the data at hand. Unfortunately, this means that carefully reviewing all forty-two of the studies replicated by Lall is not feasible. Instead, we focus on one study, Pelc (2011), which we are familiar with, and which appears as a good *prima facie* candidate for imputation.

1.1 Argument

Pelc (2011) asks what explains variation in the level of flexibility that countries inject into their trade tariffs. Such flexibility, called "binding overhang," is equivalent to the difference between the trade duties actually levied at the border and the "bound," or maximum rate, that countries commit to, and which they cannot legally exceed. The greater the difference, the greater countries' flexibility to legally raise trade protection. Pelc argues that binding overhang generates costly uncertainty, and shows that governments that enjoy alternative sources of flexibility—floating exchange rates or the ability to use antidumping measures—retain lower levels of binding overhang.[2] Lall (2016) applies multiple imputation to Pelc's replication dataset, and produces new coefficient estimates that fail to cross standard thresholds of statistical significance.

1.2 Descriptive statistics

Table 1 shows descriptive statistics from Pelc's replication dataset. Most missing observations are found in the dependent variable *Binding Overhang*, and three of the independent variables: *Products Imports*, *Floating Currency*, and *Regime Type*. Interestingly, the sample of completely observed units is descriptively similar to the sample of incompletely observed units: most of the variables' means are similar in the complete and incomplete data (Table 1). But if there are few differences in descriptive statistics across complete and incomplete rows of the dataset, there are important

---

[1] For full replication materials for these estimations and the simulations below, see Dataverse doi:10.7910/DVN/S9G9XS (Arel-Bundock and Pelc, 2017).

[2] Since countries' bound duties rarely move over time, data are cross-sectional. Observations are made at the country-product level during a country's first year following WTO accession.

differences in the variables' means between units for which the dependent variable is observed or missing. In particular, the proportion of *Floating Exchange Rates* – the key independent variable – is 10% for units where the dependent variable is observed, but much smaller where the dependent variable is missing (2%). This suggests that the patterns of missingness on the left and right-hand sides of Pelc's regression equation may be conceptually different.

Missingness in the dependent variable. As we noted above, a large fraction of observations (23%) in Pelc's replication data do not carry information about the dependent variable, *Binding Overhang*. Indeed, about 40% of the "new" observations that Lall's imputation introduces originally showed missing values on that key variable. But just because we *can* impute those missing values does not mean that we *should*. Sometimes, knowledge of the data dictates that an observation remain "missing."

Recall the meaning of *Binding Overhang*: it reflects the flexibility of a country's tariff rate commitments at the WTO. A country's international commitment on a given tariff line only becomes binding once the WTO records it; the organization holds a complete record of all binding tariff commitments.[3] By construction, this variable is complete.

"Missingness," in this case, reflects the fact that some countries choose not to make commitments on every product. For instance, Bangladesh has long been unwilling to bind itself within the international trade regime, refusing to commit to a cap on a great number of its tariff lines. For each of the 696 products where Bangladesh remains unbound to this day, the replication dataset from Pelc (2011) records a missing value for the dependent variable. With imputation, these blank spaces get filled in, creating out of whole cloth a total of $89,030$ tariff commitments that never took place.

These are not "unobserved" data points; they are "non-existent."[4] Re-

---

[3]When the WTO itself calculates a country's official average tariff rate, it thus excludes unbound tariff lines. WTO Statistics. `https://www.wto.org/english/res_e/statis_e/popup_indicator_help_e.htm`.

[4]One interesting possibility is that the original results in Pelc (2011) could suffer from a form of selection problem not highlighted by Lall. We could think of tariff negotiations as a two-step process. First, countries decide whether they want to be bound on a given tariff line. Second, they choose the specific value of the tariff they want to commit to. This problem lies outside the scope of the replication, but it is worth noting that the share of country-product observations with floating exchange rates is much higher in the sample where tariff commitments are made than where countries refuse to be bound (10% vs. 2%, respectively - Table 1). If we think of unbound tariff lines as having very high flexibility, introducing the "dogs that didn't bark" into the estimation amounts to including many

---

taining such missing observations in a dataset is inconsequential when applying listwise deletion. But when using multiple imputation, care must be taken to remove theoretically irrelevant observations before regression analysis.

Missingness in the independent variables. Turning to right-hand side variables, it is interesting to note that most of the missing data in Pelc are found in *Products Imports*, *Floating Currency*, and *Regime Type*. In a series of tests not reported here, we found that the observations dropped by LWD due to missingness in the first two variables generally conform to Pelc's expectations; including those units in the sample via MI or other means does not affect the results.[5] We thus focus our attention on the *Regime Type* variable, whose effect on sample composition turns out to be consequential.

In the original estimation, over 36,000 observations were excluded by listwise deletion due to lack of information about democracy. Does ignoring those observations introduce bias in the estimates? We can offer a preliminary answer to this question and develop some intuition about the threat to inference by explicitly theorizing the missingness generation mechanism.

1.3 Why are data missing? Should we expect listwise deletion to bias the regression estimates?

Recall that the complete case OLS estimator is biased when the dependent variable remains associated with missingness after we condition on the regressors. This can happen if (a) values of the dependent variable directly determine if a given case is fully observed, or (b) some unobserved variable drives both values of the dependent variable *and* the pattern of missingness.

A priori, it seems unlikely that the level of tariff commitment flexibility directly affects whether we can measure a country's level of democracy. The question thus becomes: Can we think of (unobserved) variables which drive both values of the dependent variable *and* the pattern of missingness? The answer is "yes."

---

new observations with fixed exchange rates *and* very high binding overhang. In this view, Pelc's original estimates appear more conservative.

[5]We manually included information on *Floating Currency* for ten countries/regions which were not coded in the currency regime dataset of Reinhart and Rogoff. We use two alternative approaches to include the observations with missing *Products Imports*: We drop that variable from the regression model and use listwise deletion, or we apply multiple imputation to the subset of observations which show missing *Products Imports* but are otherwise complete.

Case selection for the *Regime Type* variable is a deterministic function of population size: the Polity Project only includes countries with a population greater than $500,000$.[6] Similar sample selection strategies are often adopted by international organizations and scholars who choose not to collect data on extremely small countries, island nations, and semi-autonomous territories.

There are good reasons why political economy theories may apply differently to such anomalous units. Consider five of the small countries excluded from Pelc's original analysis because they do not feature in Polity IV: Antigua and Barbuda, Dominica, Grenada, St-Kitts and Nevis, and St-Lucia. Polity excludes such countries in part because in spite of their formal sovereign status, states with populations of less than 50,000 (as is the case of St-Kitts in the sample period) give rise to different expectations over the presence of political institutions. One can argue that similar concerns arise in the case of trade: these island nations are linked by their use of the East Caribbean dollar, which has been pegged to the US dollar since 1976. They are also highly dependent on imports, so they protect only a small number of domestic industries, and their tariff line commitments thus feature relatively low binding overhang. In this case, the pattern of missingness is clearly related to values of the dependent variable; taking those five countries into account pulls the estimates in a direction counter to the theory's expectations.[7]

This does not pose a problem as such: Pelc's argument is stated in probabilistic terms, and researchers rarely expect every unit to behave as theory predicts. However, as is well known, the leverage of an observation in regression analysis is a function of its proximity to the centroid of the data (Greene, 2011, 99-100). Since island nations tend to be exceptional along most dimensions, they can exert an influence on the overall results which is disproportionate to their importance in the world economy. In other words, by including countries that are purposefully left uncoded by Polity or the World Bank, multiple imputation risks introducing outliers, with attendant consequences on the overall results.

---

[6]Polity IV Dataset User's Manual. 2015. `www.systemicpeace.org/inscr/p4manualv2015.pdf`

[7]Note that the European Union, for example, counts as a single unit of observation in the analysis, on the same order as each island nation that Lall introduces through imputation. Given that EU member states are part of a customs union, this is the correct treatment, but this comparison highlights how MI may give disproportionate weight to units that occupy a marginal position in the world economy.

1.4 Why do results diverge under MI and LWD?

Table 2 shows how each of the choices described above affects Pelc's results. The first two columns show, respectively, the original estimation and the replication using multiple imputation to fill in missing values in all rows of the dataset. In Model 3, we exclude the theoretically irrelevant observations with missing dependent variable.[8] In Model 4, we also exclude the five small East Caribbean states mentioned above.[9]

These two corrections suffice to restore the original results that Lall claims as "disappeared."[10] Lall's divergent results thus appear driven by the introduction of nearly 90,000 non-existent country commitments and, less problematically, by the addition of observations for five outliers: Antigua and Barbuda, Dominica, Grenada, St-Kitts and Nevis, and St-Lucia.

What are we to make of this? Model 3 should be uncontroversial, since it corrects a substantive mistake. However, since Pelc (2011) does not explicitly consider whether extremely small countries are subject to different incentives from others, discussion of Model 4 necessarily amounts to *post hoc* theorizing about scope conditions. The question then becomes whether the weight of these five island nations, with a combined population of less than 450,000, should serve as falsifying evidence against a relationship which holds for the rest of the sample.

---

[8] We also fill in true values of the main independent variables in a few cases where data were originally missing, but for which reliable information are now available. This yields new observations for 12 countries on the exchange rate variable, and 2 countries on trade remedies usage. These countries are missing in the Reinhart and Rogoff dataset because of uncertainty about the precise exchange rate regime (e.g. whether a country's de facto regime corresponds to a moving band of $+/- 2\%$ or $+/- 5\%$), but filling these in for the binary variable "Fully Floating Currency" proves straightforward: all but the EU are not fully floating for 1995, or the relevant year of WTO entry. The other countries are Taiwan, Angola, Namibia, Fiji, Oman, Cuba, Macao, Macedonia, St-Kitts, Rwanda, and Djibouti. As for the trade remedy variable, the Bown (2011) data lack observations for Trinidad and Tobago and St-Kitts and Nevis, neither of which was a trade remedy user at the moment of its WTO accession.

[9] The results are very similar if we drop all nine of the very small countries that were originally unaccounted for by the Polity Project.

[10] Lall replicates three other models from Pelc (2011). The corrections we propose here also bring those models in line with the original published estimates, with Models 1, 3, and 4 from Pelc's Table 2 yielding "conclusive" results, and Model 2 producing "mixed" results.

1.5   Robustness checks

In our paper, we recommended that applied researchers probe the robustness of their results by comparing results using different imputation procedures and tuning parameters. In that respect, Lall (2016) should be commended, since he replicated a random sample of C/IPE studies using both the `Amelia` and `mice` imputation routines. In the case at hand, the choice of imputation routine turns out to be much less consequential than the problems we discussed above.

Yet it is worth noting that imputing Pelc's replication data using `Amelia`'s default settings (as Lall does) produces highly implausible values for many variables. Consider the imputed values of *Binding Overhang* for the aforementioned case of Bangladesh. There, Lall's "complete" datasets show (imputed) commitments far lower than the country's (actual) average tariff, moving the average Bangladeshi bound rate from 167% in the original data to 80% in the imputed data. But, if anything, unbound tariff lines should be thought of as having maximum rates far *higher* than the average, since they can be raised at will. `Amelia` also produces a range of impossible values, such as negative bound tariffs of -107%—effectively a commitment by countries to pay exporters the full value of their exports at the border.

Referring back to our proposed best practices, empiricists should let their knowledge of the data guide their assessment of imputed data. When these appear implausible, it may be that MI is being used in ways that the data-generating process may not support.

Table 1: Descriptive statistics

	Missing %	Range	SD	Mean				
				Full	Complete	Incomplete	Obs.DV	Miss.DV
Binding Overhang	23.10	864.63	26.90	18.63	18.67	18.59	18.63	
Logged Products Imports	24.50	13.82	2.32	3.29	2.98	3.54	3.37	3.07
Fully Floating Currency	13.60	1.00	0.27	0.08	0.10	0.06	0.10	0.02
Regime	9.40	108.00	10.24	3.29	3.19	3.41	4.19	0.27
Logged GDP	3.80	10.39	2.52	23.97	24.12	23.78	24.16	23.27
MFN	2.90	871.40	13.67	12.08	11.46	12.89	11.44	14.53
Logged GDP per capita	2.50	5.66	1.55	7.83	8.00	7.62	8.02	7.20
Remedies User	1.60	1.00	0.47	0.34	0.36	0.30	0.38	0.21
Agricultural Product	0.00	1.00	0.32	0.11	0.09	0.15	0.14	0.03
LDC	0.00	1.00	0.36	0.15	0.19	0.10	0.10	0.32
Recent Entrant	0.00	9.00	1.84	0.69	0.57	0.86	0.76	0.49

Table 2: The Effect of Policy Substitutes on Tariff Flexibility: Replicated vs. Imputed Data

	(1) Original Estimation	(2) Lall Replication	(3) Tariffs Correction	(4) Carribean Correction
Applied Rate	-0.522	-0.348	-0.403	-0.441
	(0.081)	(0.064)	(0.081)	(0.083)
Logged GDP per capita	-1.708	-1.243	-0.682	-1.719
	(3.300)	(1.704)	(2.278)	(2.329)
Logged GDP	0.894	-2.850	-2.159	-0.664
	(1.315)	(0.900)	(1.054)	(1.040)
Regime	0.125	0.177	0.188	0.173
	(0.319)	(0.131)	(0.217)	(0.234)
Logged Products Imports	0.091	0.032	0.006	-0.057
	(0.165)	(0.112)	(0.128)	(0.129)
LDC dummy	4.667	3.407	5.148	6.989
	(7.453)	(4.754)	(6.702)	(6.409)
Agricultural Product	23.929	19.042	18.816	17.446
	(3.773)	(2.800)	(3.046)	(3.061)
Recent Entrant	-5.023	-4.612	-4.577	-4.602
	(0.631)	(0.448)	(0.617)	(0.680)
Fully Floating Currency	-9.909	-3.965	-7.758	-10.356
	(3.828)	(3.856)	(4.657)	(4.581)
Remedies User	-16.168	-8.696	-11.084	-12.866
	(7.718)	(5.360)	(6.678)	(6.587)
Constant	22.052	104.739	85.488	58.503
	(38.362)	(25.137)	(31.422)	(30.859)
N	163097	385798	296768	284286

Dependent variable is binding overhang. OLS estimates with robust standard errors in parenthesis clustered on common country. Column (1) is the original regression from Pelc 2011. Column (2) is Lall's replication after imputation of all missing observations. Column (3) is model (2), excluding non-existent tariffs, and adding known currency regime data and trade remedies data. Column (4) is model (3), excluding imputed data for five Caribbean countries that Polity IV avoids coding due to their small size.

2 Supplementary materials: Simulation

In the text, we developed rules of thumb to help researchers identify the conditions under which MI is most likely to be beneficial. Here, we use a set of very simple Monte Carlo experiments to illustrate.

2.1 New assumptions, new problems?

`Amelia` combines an expectation-maximization algorithm with a bootstrap approach to impute missing values in partially-observed datasets. It makes two main assumptions: data must be MAR and be distributed following a multivariate normal law (Honaker, King and Blackwell, 2011, 3-4).

Proponents of MI often claim that the approach performs well under NMAR. Unfortunately, as we mentioned in Footnote 8 of the text, evidence in support of that contention is scant. One important problem is that the NMAR concept covers a vast array of potential dependence patterns, and that any performance assessment will be highly dependent on the specific data generating process under investigation. This means that any attempt to compare the performance of LWD and MI in NMAR data will at best be partial in scope. That said, a recent unpublished manuscript by Pepinsky (2016) raises some concerns. Based on extensive Monte Carlo simulations, the author concludes that "multiple imputation yields results that are frequently more biased than listwise deletion when data are [NMAR] [...] even with very strong correlations between fully observed variables and variables with missing values, such that the data are very nearly MAR." In short, while we do not have access to strong evidence either way, there are good reasons to remain cautious in the (quite typical) case where researchers are unable to claim that the MAR assumption holds.

The multivariate normality assumption also raises potential issues since, as `Amelia`'s authors concede, it is "often a crude approximation to the true distribution of the data" (Honaker, King and Blackwell, 2011, 4).[11] And even if analysts can use truncation or transformations to make individual variables look more "normal", the multivariate normal assumption imposes requirements beyond marginal distributions: it also constrains the *structure of relationships* between variables. Below, we show that even if every variable, on its own, is standard normal, `Amelia`'s performance can be severely

---

[11]Other imputation procedures relax the multivariate normality assumption, but open several "researcher degrees of freedom." For example, the `mice` routine (van Buuren and Groothuis-Oudshoorn, 2011) requires that the analyst makes seven choices in the specification of the imputation model.

degraded when the *dependence* between variables is not normal.

2.2 Simulation design

We wish to use a linear regression model to estimate the association between regressand y and regressor x_1, controlling for x_2. To study the effect of deviations from multivariate normality, we draw values for y, x_1, x_2 using four different random numbers generators ($N = 1000$).[12] The first produces multivariate normal data with mean zero, variance one, and covariances equal to 1/3. The other three use *Clayton*, *Gumbel*, and *Frank* copulas to produce data whose marginal distributions are standard normal, but whose dependence structure is non-normal.[13] Figure 1 illustrates the difference between variables drawn from a Gumbel copula and others drawn from a multivariate normal. In both cases, the marginal distributions (density plots on the diagonals) are standard normal. However, the scatterplots show that the structure of dependence is slightly different in the Gumbel and normal data.

Figure 1: Variables with normal marginal distributions drawn from three Archimedean copulas and a multivariate normal (N=1,000).

(a) Gumbel   (b) Multivariate Normal

---
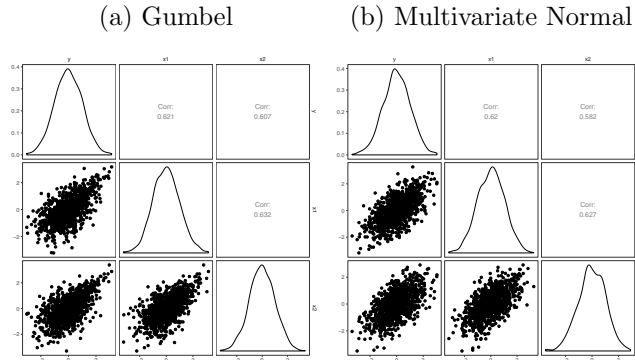
[12]In a series of tests not reported here, we found that adding correlated auxiliary variables to the imputation stage does not materially affect our overall conclusions for the multivariate normal case. Unfortunately, since we cannot manipulate individual covariance parameters with the `copula` package for `R`, it is impossible to use imputation-only variables in the other experiments.

[13]Copulas are multivariate probability distributions for which all marginals are uniform over [0,1]. This property allows us to use the probability integral transformation to model the dependence structure between variables separately from their marginal distributions (Yan et al., 2007). The (arbitrary) tuning parameters that control the strength of association between x_1, x_2, and y are 1.8 (*Gumbel*), 5.4 (*Frank*), and 1.5 (*Clayton*).

To see how the missingness mechanism affects the performance of MI, we eliminate observations in each dataset by drawing binary indicator Q_i from a binomial distribution, where the probability that unit i is completely observed is given by:

$$Pr(Q_i = 1) = logit(\theta_1 + \theta_{x_1}x_{1i} + \theta_{x_2}x_{2i} + \theta_y y_i). \tag{1}$$

θ_1 controls the share of partially observed cases. When $\theta_{x_1} = \theta_{x_2} = \theta_y = 0$, data are MCAR and we expect no bias. When $\theta_{x_1} = \theta_{x_2} = 1$ and $\theta_y = 0$, missingness is a function of the observed regressors (MAR), but it is conditionally independent of the regressand; again, we expect no bias. When $\theta_y = 1$ and $\theta_{x_1} = \theta_{x_2} = 0$, the outcome variable remains associated with the pattern of missingness, even after we control for x_1 and x_2. In practice, this situation could arise when the outcome variable itself drives missingness, or when an unobserved variable determines both the outcome and missingness. In such cases, we expect LWD to introduce bias in regression estimates.

To assess the performance of MI when different variables need to be imputed, we use the Q_i indicator to create three versions of each partially-observed dataset. To begin, we use Equation 1 to erase values of the dependent variable, but leave the x_1 and x_2 regressors intact. Then, we repeat the exercise with the other two independent variables. All datasets are imputed ten times using the `Amelia` software.

Figure 2 shows the mean absolute deviation from full-data estimates of the x_1 coefficient under different data generation mechanisms. Four main conclusions emerge.

First, columns 3 and 4 show that MI does not materially improve upon LWD when data are MCAR or where we can control for the determinants of missingness.[14] This is consistent with the analytical results we presented in text, which show that LWD estimates are unbiased when missingness is conditionally independent of the dependent variable.

Second, when the missingness mechanism is related to the dependent variable and data are multivariate normal, imputing data with `Amelia` can yield important benefits.

Third, improvements with MI seem particularly large when the control variable (x_2) is affected, rather than the main independent or dependent

---

[14]`Amelia` does seem to produce slight improvements at very high levels of missingness (e.g., 85%). However, this is only the case when the control variable is missing, and not when either the outcome or the main independent variables are missing. Note that these simulations probably understate the benefits of MI where analysts can leverage auxiliary variables with high predictive power.

variables of interest. This makes sense because, as Little (1992, 1227) points out, if "the X's are complete and the missing values of Y are missing at random, then the incomplete cases contribute no information to the regression of Y on $X_1, ..., X_p$."[15] Conversely, the imputation of auxiliary variables yields a benefit because it allows for estimation based on units that are otherwise fully observed with respect to the Y and the X's of interest.

Fourth, the rest of Figure 2 shows that even if all the variables in our experiments are standard normal, deviations from *multivariate* normality can severely degrade the imputation algorithm's performance. Indeed, as we manipulate the structure of dependence between variables, we see that LWD often *outperforms* `Amelia`.

This is not to say that all departures from multivariate normality will hinder the imputation procedure. The copulas we used above were chosen for convenience[16], and because they are widely used in the statistical literature. We do not argue that these distributions represent better approximations of social phenomena, nor do we claim that all non-normal multivariate data will degrade the performance of `Amelia`. Nevertheless, the results in Figure 2 are interesting because, to our eyes at least, the *Clayton*, *Frank*, and *Gumbel* data do not look more "atypical" than the observational data we regularly work with. If seemingly innocuous departures from multivariate normality can have such deleterious effects on `Amelia`'s performance, one wonders how well it can be expected to work in real-life settings, where data are messier and the multivariate central limit theorem does not rule all.

In sum, Monte Carlo experiments support our earlier contention that missingness does not substantially impair our ability to obtain accurate regression estimates using LWD, as long as we can control for the determinants of missingness. MI using `Amelia` can sometimes improve our estimates, but the procedure relies on two very strong assumptions: MAR and multivariate normality. In practical applications, NMAR data are ubiquitous, the MAR condition is untestable, and existing evidence does not allow us to conclude that MI dominates LWD when MAR is violated. Multivariate normality is often a poor descriptor for real-world data, and departures from that dependence structure can severely degrade the performance of the imputation model.

This discussion suggests that analysts would do well to probe the sensitivity of their results by trying different MI routines and tuning parameters.

---

[15]Leaving units with missing outcome in the dataset for imputation can still help estimation if it improves the imputation model for missing values of the regressors.

[16]Random number generators are readily available for `R` (Hofert et al., 2016).

Moreover, when results under MI and LWD diverge, analysts will generally be unable to make an *a priori* claim that one set of results is more credible than the other. Case-specific judgment and knowledge of the data remain important.

Figure 2: Performance of different estimation procedures in 5,000 Monte Carlo simulations.
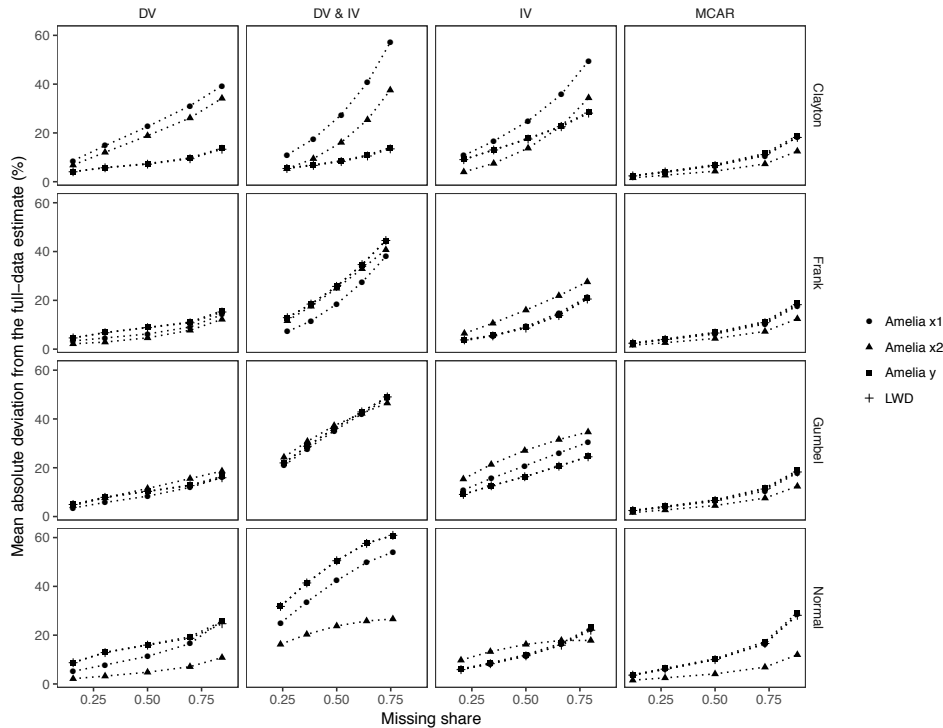
References

Arel-Bundock, Vincent and Krzysztof J. Pelc. 2017. "When Can Multiple Imputation Improve Regression Estimates?".
URL: *http://dx.doi.org/10.7910/DVN/S9G9XS*

Greene, William H. 2011. "Econometric Analysis, Fifth Edition.".

Hofert, Marius, Ivan Kojadinovic, Martin Maechler and Jun Yan. 2016. *copula: Multivariate Dependence with Copulas.* R package version 0.999-15.
URL: *http://CRAN.R-project.org/package=copula*

Honaker, James, Gary King and Matthew Blackwell. 2011. "Amelia II: A Program for Missing Data." *Journal of Statistical Software* 45(7):1–47.
URL: *http://www.jstatsoft.org/v45/i07/*

Lall, Ranjit. 2016. "How Multiple Imputation Makes a Difference." *Political Analysis* p. mpw020.

Little, Roderick J. A. 1992. "Regression With Missing X's: A Review." *Journal of the American Statistical Association* 87(420):1227–1237.
URL: *http://www.jstor.org/stable/2290664*

Pelc, Krzysztof J. 2011. "How States Ration Flexibility: Tariffs, Remedies, and Exchange Rates as Policy Substitutes." *World Politics* 63(4):618–646.

Pepinsky, Thomas. 2016. "A Note on Listwise Deletion Versus Multiple Imputation.".

van Buuren, Stef and Karin Groothuis-Oudshoorn. 2011. "mice: Multivariate imputation by chained equations in R." *Journal of statistical software* 45(3).

Yan, Jun et al. 2007. "Enjoy the joy of copulas: with a package copula." *Journal of Statistical Software* 21(4):1–21.