A Jackknife Model Averaging Analysis of RMB Misalignment Estimates

Yin-Wong Cheung, Wenhao Wang

CESifo

A Jackknife Model Averaging Analysis of RMB Misalignment Estimates

Abstract

We adopt the Jackknife Model Averaging (JMA) technique to conduct a meta-regression analysis of 925 renminbi (RMB) misalignment estimates generated by 69 studies. The JMA method accounts for model selection and sampling uncertainties, and allows for non-nested model specifications and heteroskedasticity in assessing effects of study characteristics. The RMB misalignment estimates are found to be systematically affected by the choices of data, the theoretical setup and the empirical strategy, in addition to publication attributes of these studies. These study characteristic effects are quite robust to the choice of benchmark study characteristics, to alternative model averaging methods including the heteroskedasticity-robust Mallows approach, the information criterion approach, and the Bayesian model averaging. In evaluating the probabilistic property of RMB misalignment estimates implied by hypothetical composites of study characteristics, we find the evidence of a misaligned RMB, in general, is weak.

Yin-Wong Cheung
Hung Hing Ying Chair Professor of
International Economics
City University of Hong Kong
Hong Kong
yicheung@cityu.edu.hk

Wenhao Wang
Department of Economics and Finance
City University of Hong Kong
Hong Kong
wrwang5-c@my.cityu.edu.hk

1.   Introduction

Exchange rates affect a country's economic interactions with the rest of the world via their implications for trade competitiveness and capital flows. A country by manipulating its exchange rate can enjoy an advantageous competitiveness position in the global market, and promote its exports sector and economic growth. The benefits of a misaligned exchange rate, however, come with costs imposed on the global economy. Misaligned exchange rates are perceived to be a main cause of current account imbalances and distorted flows of capital, which generate pressure on the stability of the global financial system and the international economic order (Benassy-Quere, *et al.,* 2008; Cline and Williamson. 2010; Morrison and Labonte, 2013).

Does a country pursue policies to maintain an artificially undervalued currency to boost its exports sector and growth? While exchange rate misalignment is by now a rather standard concept in international economics, it is practically difficult to determine if a currency is misaligned and, even less likely to agree on the degree of misalignment.

A critical issue in assessing misalignment is the choice of an appropriate equilibrium exchange rate. In the literature, there are alternative concepts of equilibrium exchange rate and empirical specifications. For a given actual exchange rate, different notions of equilibrium exchange rate would generate different assessments of misalignment. Montiel and Hinkle (1999), for instance, offer an early discussion of issues on defining and measuring exchange rate misalignment. The poor performance of empirical exchange rate models presents another challenge to the evaluation of exchange rate misalignment. Meese and Rogoff (1983) cast doubt on the ability of structural or time series models to describe exchange rate movements – specifically, it is quite difficult for these models to beat a random walk specification. The so-called Meese and Rogoff puzzle highlights the difficulty of finding a commonly agreed framework to model exchange rates and, hence, assess equilibrium exchange rates. The general conclusion of this seminal study is mostly affirmed by subsequent analyses (Cheung, *et al.*, 2005; Cheung, *et al.*, 2019; Engel, *et al.*, 2018; Rossi, 2013).

The complex nature of exchange rate misalignment is aptly illustrated by the debate in the last two decades on the valuation of the Chinese currency, the renminbi (RMB). The debate was triggered by China's ballooning current account surplus and the accompanied rapid economic expansion, which have put China's foreign exchange policy in the limelight. A typical accusation is that China artificially depresses the value of the RMB to gain unfair competitive edge in the

international market and, as a result, leads to global imbalances that threaten the stability of the global financial market. China's efforts in revamping its foreign exchange policy including the modification of the RMB central parity formation mechanism in August 2015, however, have not silenced all her critics. For instance, during the recent China-US trade dispute, China was asked to avoid any kind of competitive devaluation.

The debate has triggered numerous studies on RMB misalignment. Despite the widespread suspicion of an undervalued RMB, both overvalued and undervalued RMB misalignment estimates are reported. These diverse RMB misalignment estimates are generated from studies with different specifications of equilibrium exchange rate, different empirical specifications, and different estimation methods. Cheung, *et al.* (2007), Dunaway, *et al.* (2009), and Schnatz (2011), for example, point out that the lack of a consensual exchange rate model and the sensitivity of estimates to assumptions underlying empirical specifications make it very difficult to obtain a precise estimate to assess the degree of RMB misalignment.

The current exercise is not meant to generate another RMB misalignment estimate. Instead, we would like to take stock of the plethora of empirical studies on RMB misalignment, and to identify heterogeneous patterns of reported misalignment estimates and infer what we can learn about RMB misalignment. We adopt a meta-regression approach to investigate the linkages between the heterogeneity of RMB misalignment estimates and the characteristics of the studies including data choices, model specifications, publication venues, and author attributes. We extend the meta-regression analyses of RMB misalignment estimates in Bineau (2010) and Korhonen and Ritola (2011) by examining a longer sample period and a larger sample of empirical studies.

Further, we explicitly address the uncertainty associated with model selection. Since we do not have good theories on effects of different study characteristics on reported RMB misalignment estimate, the inclusion of relevant and exclusion of irrelevant study characteristics become an empirical model selection issue. We also recognize that different combinations of potential study characteristics involve non-nested model specifications, and misalignment estimates can display heteroskedasticity due to, say, clustering on study effects.

Our exercise adopts the Jackknife model averaging (JMA) technique to handle model selection uncertainty encountered in selecting study characteristics. Comparing with the commonly used Bayesian model averaging (BMA) procedure, the JMA technique is a frequentist model averaging method that allows for non-nested model specifications and residual

heteroskedasticity.[1] In addition to addressing some data issues encountered in the current exercise, the use of the JMA technique illustrates an alternative approach of conducting model averaging.

To anticipate results, the JMA based meta-regression shows that the RMB misalignment estimates are systematically affected by the choices of data and theoretical and empirical strategies, and by publication attributes of these studies. Specifically, studies that use PPP-based data, a Penn effect setup, a cointegration specification, or are a non-academic journal publication are likely to give strong evidence of RMB undervaluation. Studies that adopt a BEER setup, have an author affiliated with a mainland China institution, or generate estimates for the period 2011 to 2014 tend to yield weak evidence of RMB undervaluation. These study characteristic effects are quite robust to the choices of benchmark study characteristics, alternative model averaging methods including the heteroskedasticity-robust Mallows approach, the AIC and BIC approaches and the BMA.

Using the Jackknife model average results, we evaluate the probabilistic property of RMB misalignment estimates implied by hypothetical composites of study characteristics. We consider the composites defined by the average values of study characteristics, the sets of most frequently and least frequently used study characteristics, and the sets of study characteristics adopted by the most cited English study and the most cited Chinese study. For all these hypothetical cases, the evidence of a misaligned RMB, in general, is weak.

2.   RMB Valuation

China's foreign exchange policy has been closely scrutinized since the turn of the 21st century. The heavily managed RMB coupled with China's fast buildups of current account balance and international reserves have invited complaints that the RMB valuation is the culprit of global imbalances. Studies in the 2000s usually favored the view that the RMB is undervalued, if not seriously undervalued (Bergsten, 2007; Coudert and Couharde, 2007; Funke and Rahn, 2005; Goldstein and Lardy, 2009; Wang, *et al.*, 2007). Indeed, based on the undervaluation estimates available at that time, the 2005 Schumer–Graham bipartisan bill proposed to impose a tariff rate of 27.5% on all imports from China to force China to stop currency manipulation.[2]

---

[1]   Relative to the frequentist approach, the BMA is quite commonly used in economics and finance. Indeed, Cheung and He (forthcoming) adopts the BMA approach to handle model selection uncertainty in studying RMB misalignment.

[2]   See http://schumer.senate.gov/SchumerWebsite/pressroom/press_releases/2005/PR4111.China020305.html.

As part of its on-going efforts to reform the economy, China has adjusted and revamped its foreign exchange policy several times since 1994, the year it replaced a dual-exchange rate arrangement with a policy targeting the US dollar. The modification of the RMB central parity formation mechanism in August 2015 is the most recent major effort to enhance the role of market forces in determining the value of the RMB. While the reform effort is welcome by IMF,[3] it has met with cautious skepticism by critics – for instance, the US Department of the Treasury (2006) asserts that "the increased (RMB) flexibility is considerably less than is needed." Indeed, the RMB valuation is directly and indirectly referred to in the 2018-19 China-US trade dispute.

With its growing importance in the global market, if China manipulates its currency, there will be serious economic and political backlashes. The contentious nature of assessing whether China is manipulating its currency is attested by the intense policy debate in the last two decades. A recent US report, for instance, asserts that "China has a long history of pursuing a variety of economic and regulatory policies that lead to a competitive advantage in international trade, including through facilitating the undervaluation of the RMB" (United States Department of the Treasury, 2018, p. 3).[4]

An overarching question is: in view of the difficulty of defining equilibrium exchange rate and modeling exchange rate; especially for currencies of transitional economies,[5] how confident we are in determining whether the RMB is misaligned, and its precise degree of over- or undervaluation? Admittedly, it is not straightforward to resolve these longstanding issues in exchange rate economics. On the other hand, the literature abounds with different empirical estimates of RMB misalignment. What can we learn from these empirical estimates?

3.   RMB Misalignment Estimates, Study Characteristics, and Control Variables

The current study conducts a meta-regression analysis of RMB misalignment estimates. Meta-regression analysis is a commonly used technique to aggregate information from different

---

[3]   See the IMF press release http://www.imf.org/external/country/CHN/rr/2015/0811.pdf.
[4]   The US designated China as a currency manipulator (https://home.treasury.gov/news/press-releases/sm751) on August 5, 2019. The last time the US labelled China for manipulating its currency regularly was the period between May 1992 and July 1994 (United States Department of the Treasury, 1992, 1994).
[5]   Frankel (2006), for instance, questions what is the "equilibrium exchange rate" when there is more than one sector to consider. Engel (2009) argues that "external balance" needs to be defined in terms of efficiency in global resource allocation, rather than trade balances in the usual sense.

studies with a common object of interest and gain further insight on the common object.[6] Bineau (2010) and Korhonen and Ritola (2011) adopted this framework to investigate the links between characteristics of the studies and heterogeneity of RMB misalignment estimates.[7] The current meta-regression analysis is based on

$$Y_i = \alpha + \Sigma_{j=1}^{J}\beta_j X_{ij} + \Sigma_{k=1}^{K}\gamma_k Z_{ik} + \varepsilon_i \; ; i = 1, \ldots, \text{n},\tag{1}$$

where Y_i is the i-th RMB misalignment estimate, X_{ij} is the j-th characteristic type of the study that reported Y_i, Z_{ik} is the control variable that captures the k-th exchange rate policy regime in which Y_i is reported, J and K are the total numbers of, respectively, study characteristic types and policy regimes under consideration, ε_i is an error term, and n is the number of observations. The exchange rate policy regime control variables are included to isolate the study characteristic effects from influences of China's exchange rate policy.

Meta-regression analysis allows the pooling of results from different studies on RMB misalignment. If random errors are the source of study-to-study variability of estimates, information aggregation can improve the precision of the inference of RMB misalignment. The pooling of studies also makes it possible to investigate if the reported estimate depends on, say, who is the author, which is the model employed, where is the study published, …, etc.

3.1   RMB misalignment estimates

Our sample of RMB misalignment estimates is largely the same as the one used by Cheung and He (forthcoming). They are obtained from 69 studies that report quantitative estimates of RMB misalignment – of which 62 are English publications and 7 are Chinese publications.[8] These studies were published between 1998 and 2018, and provided 925 RMB misalignment estimates given by the difference of the actual and the (estimated) equilibrium exchange rate in percentage between 1994 and 2014.[9] We focus on estimates after 1994 because China instituted a major

---

[6]   Applications in economics and finance includes meta-regression analyses on determinants of FDI spillovers (Meyer and Sinani, 2009; Irsova and Havranek, 2013), determinants of financial performance (Capon, *et al.*, 1990), economic growth (Valickova, *et al.*, 2015), labor market (Groot and Van Den Brink, 2000), and asset pricing (Coggin and Hunter, 1993).

[7]   Cheung (2012) and Government Accountability Office (2005), for example, offer narrative reviews of selected studies on RMB misalignment but not quantitative analysis of reported estimates.

[8]   These studies are listed in Appendix A.1, and selected from an initial collection of 283 studies, which represent to the extent possible studies that report inferences about RMB misalignment (Cheung and He, forthcoming).

[9]   We dropped 12 extreme estimates from Chang (2008) which the author considered not reliable.

change of its exchange rate policy in January 1994, replacing the dual-exchange rate arrangement with a managed exchange rate targeting the US dollar. After 2014, there is only a paucity of estimates available. [10] A positive (negative) RMB misalignment estimate indicates RMB undervaluation (overvaluation). The terms "RMB misalignment estimates" or "misalignment estimates" here refer to these quantitative estimates of RMB misalignment.

Figure 1. The Number of Publications and Observations across Years
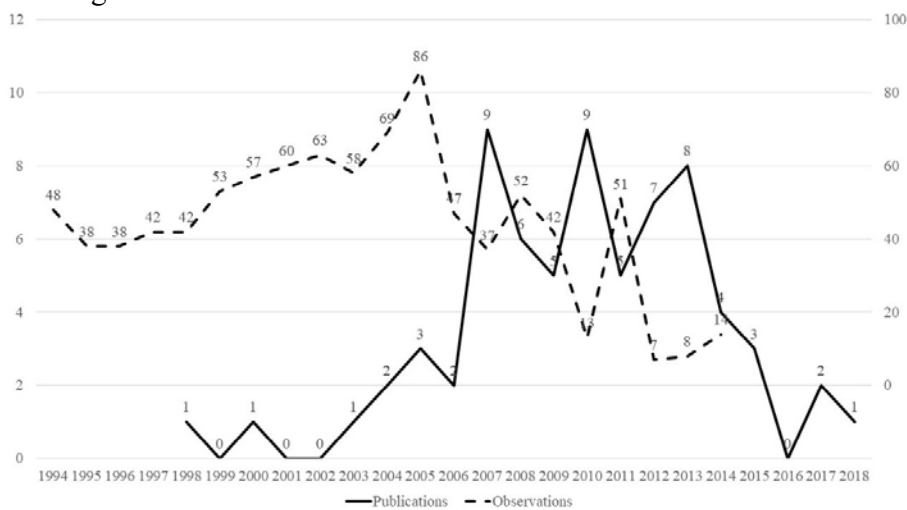
Notes: The numbers of publications and the numbers of misalignment estimates in a given year are given, respectively, along the solid line (left scale) and the dash line (right scale).

Figure 1 depicts, for each year, the numbers of studies and RMB misalignment estimates in the sample. There are relatively few studies on RMB misalignment before the early 2010s when China's trade balance and holding of international reserves were relatively small. The number of studies has subsided since 2015, the year in which the International Monetary Fund stated that the RMB is no longer undervalued (International Monetary Fund Communications Department, 2015). Probably due to data not available at the time these empirical studies were conducted, there is only a relatively small number of estimates reported for 2012 to 2014.

Table 1 presents, for each year, some descriptive statistics of the RMB misalignment estimates. Both the mean and median statistics indicate that the RMB is undervalued for most years. (Recall a positive value implies undervaluation). Relatively speaking, the degree of

---

[10]   Bineau (2010) and Korhonen and Ritola (2011) cover, respectively, 17 and 30 studies, and 130 and 99 misalignment estimates in their exercises.

undervaluation is higher in the first decade of the 21st century. Nevertheless, for any given year, the estimates can be quite variable; the coefficient of variation is larger than one for 15 out of 21 years. The observation is in line with the wide ranges given by the yearly minimum and maximum estimates. Further, these estimates exhibit considerable degree of skewness. Both the standard error and skewness indicate of a high level of heterogeneity of within-year RMB misalignment estimates. Further, the dispersion of these statistics across years collaborates with time-varying data heterogeneity.

Table 1.　　　RMB Misalignment Estimates: Descriptive Statistics

year	Obs	Mean	SE	Median	Skewness	Min	Max
1994	48	22.26%	13.80%	20.45%	53.11%	-4.00%	53.00%
1995	38	7.81%	11.73%	7.70%	27.99%	-17.35%	39.00%
1996	38	0.86%	16.94%	2.50%	-53.58%	-50.00%	36.00%
1997	42	5.16%	17.13%	3.62%	94.03%	-30.00%	58.00%
1998	42	1.42%	18.27%	1.46%	78.38%	-35.00%	57.00%
1999	53	-0.20%	19.64%	-0.27%	28.44%	-51.00%	58.00%
2000	57	8.32%	22.34%	3.02%	59.68%	-36.00%	63.50%
2001	60	3.68%	26.84%	0.16%	-82.70%	-108.00%	63.50%
2002	63	11.44%	17.60%	6.00%	123.37%	-11.26%	63.50%
2003	58	18.00%	17.48%	12.50%	71.60%	-6.86%	63.50%
2004	69	22.70%	22.43%	18.00%	58.60%	-13.26%	76.00%
2005	86	20.22%	22.76%	16.99%	-39.53%	-48.70%	78.30%
2006	47	20.94%	18.69%	12.60%	115.49%	-5.10%	75.80%
2007	37	17.45%	17.95%	13.06%	135.07%	-10.00%	79.00%
2008	52	16.53%	19.95%	15.60%	-4.44%	-42.00%	78.10%
2009	42	9.94%	14.42%	9.40%	-19.57%	-15.10%	40.20%
2010	13	19.65%	14.91%	22.46%	-13.86%	-3.85%	40.70%
2011	51	-3.29%	24.14%	0.00%	-76.37%	-87.00%	46.00%
2012	7	6.74%	15.14%	3.10%	127.34%	-11.40%	38.10%
2013	8	10.49%	10.34%	5.95%	108.09%	2.00%	31.00%
2014	14	-1.24%	21.52%	0.00%	43.79%	-27.20%	33.30%

Notes: For each year, the Table presents the number of RMB misalignment estimates (Obs), and the corresponding sample means (Mean), standard errors (SE), medians (Median), skewness coefficients (Skewness), minima (Min), and maxima (Max).

There is substantial within-year and across-year heterogeneity in RMB misalignment estimates. The variability of misalignment estimates can be the outcome of contrasting ideas of what constitutes an equilibrium exchange rate and competing econometric methodologies (Hinkle and Montiel, 1999). Other study designs including data choice can also systematically affect empirical findings.

3.2   Study Characteristics

To shed insight on study-to-study variability of RMB misalignment estimates, we collected information on ten characteristics of these studies. These ten study characteristics are grouped into three categories; namely, a) data specifications, b) theoretical and estimation specifications and (c) publication attributes.

We consider three alternative ways to classify data specifications. First, we consider the data frequency, and differentiate studies using either annual or non-annual data. The two types are selected such that there are sufficient observations for each characteristic type. Second, we consider whether a study uses PPP-based data typically derived from ICP surveys or non-PPP-based data.[11] Third, we consider whether the dollar-based real exchange rate is used or not.

The category of theoretical and estimation specifications is deployed to capture the study designs related to the theoretical underpinning and empirical methodology. In assessing misalignment, a critical issue is the choice of equilibrium exchange rate model. It is not surprising that different measures of equilibrium exchange rate yield different inferences about the level of misalignment.[12] We group the studies into four alternative theoretical settings; namely, the behavioral equilibrium exchange rate (BEER) approach, fundamental equilibrium exchange rate (FEER) approach, Penn effect approach, and other approaches. The BEER and FEER approaches are quite commonly used to assess exchange rate misalignment of developing countries including China. The Penn effect approach is motivated by the robust empirical relationship between real exchange rates and *per-capita* income, and perceived to be a useful setup for evaluating the RMB misalignment (Frankel, 2006).[13] The type "Other Approaches" includes the absolute or relative PPP approach and the shadow price of foreign exchange approach.[14]

---

[11]   PPP-based data refer to data derived from the internationally comparable price indices (labelled purchasing power parities, PPPs) that are generated from surveys conducted by the International Comparison Program (ICP) (World Bank, 2013, 2015). These data offer "consistent" information that facilitates cross-country comparison of real exchange rates and, hence, misalignment.

[12]   Ideally, a useful model shall inform us how the RMB is affected by both canonical economic factors and institutional arrangements including capital controls.

[13]   The Penn effect is documented in, for example, Kravis and Lipsey (1983, 1987), Kravis, *et al.* (1978) and Samuelson (1994). Theoretical explanations are offered by, for example, Balassa (1964), Samuelson (1964), Bhagwati (1984), Bergstrand (1991), and Bergin, *et al.* (2006).

[14]   Cheung, *et al.* (2010), for example, offers an alternation topology of theoretical approaches adopted by studies on RMB misalignment. Cenedese and Stolper (2012) presents a recent recount of equilibrium exchange rate models commonly used to assess exchange rate misalignment.

Table 2. RMB Misalignment Estimates and Study Characteristics

	Obs	Mean	SE	Rank-Sum
Entire Sample	925	11.4%	21.3%	
Data Characteristics				
Annual	516	16.5%	22.0%	0.00
Non-Annual	409	4.9%	18.4%	0.00
PPP-based	359	17.4%	24.2%	0.00
Non-PPP-based	566	7.6%	18.2%	0.00
Dollar-based RER	612	13.0%	23.0%	0.00
Non-Dollar-based ER	313	8.2%	17.1%	0.00
Theoretical and Estimation Specifications				
BEER	354	4.2%	19.6%	0.00
FEER	180	19.1%	19.3%	0.00
Penn Effect	258	18.7%	22.2%	0.00
Other Approaches	133	6.0%	18.2%	0.11
Time series	504	4.4%	19.6%	0.00
Non-Time-series	421	19.7%	20.3%	0.00
Cointegration	427	6.4%	21.1%	0.00
Non-cointegration	498	15.7%	20.5%	0.00
Reduced-form	679	9.5%	22.3%	0.00
Structural	246	16.7%	17.1%	0.00
Publication Attributes				
Journal	761	10.2%	21.8%	0.00
Non-Journal	164	17.0%	17.7%	0.00
Mainland	345	5.5%	19.4%	0.00
Non-mainland	580	14.9%	21.6%	0.00
Academics	504	11.3%	21.5%	0.35
Non-Academics	421	11.4%	21.1%	0.35

Notes: The columns labelled "Obs", "Mean" and "SE" present the number of observations, the mean and the standard error of misalignment estimates. The numbers shown in the first row are statistics based on the entire sample. Numbers in other rows are for misalignment estimates of study characteristics identified in the first column. The rightmost column presents the p-values of Wilcoxon Rank-Sum test under the null hypothesis of $E(Y_i) = E(Y_j)$ and the alternative hypothesis of $E(Y_i) \neq E(Y_j)$, where Y_i and Y_j belong to different types of a study characteristic.

Besides theoretical underpinning, we consider three ways to characterize empirical methodology. First, we differentiate the uses of time series and non-time series (*a la*, cross-sectional and panel) estimation techniques. Second, we consider whether a cointegration setup is used or not. Third, we consider whether a reduced-form or structural setup is used.

The third category of study characteristics includes three publication attributes. First, the studies are sorted according to whether they are published in academic journals or not. The peer-review process commonly adopted by academic journals is regarded as a quality signal of journal publications. The second and third characteristics pertain to an author's affiliation – whether an author is affiliated with an institution in mainland China, and whether an author is affiliated with an academic institution. Given the economic and political sensitivities of RMB misalignment, an author's institutional affiliation may influence empirical findings.

Appendix A.2 lists the definitions of these study characteristics, which are coded as zero-one dummy variables. Table 2 presents the number of observations (Obs), the sample average (Mean), and the standard error (SE) of the RMB misalignment estimates in the entire sample, and under alternative study characteristic classifications. The sample averages indicate that the misalignment estimates are on average indicative of an undervalued RMB. With the exception of the case of academic affiliation, the degree of undervaluation appears to differ across study characteristic types.

We employ the nonparametric Wilcoxon Rank-Sum test to investigate, for a given study characteristic, if its component types yield RMB misalignment estimates with the same mean. Does the observed difference in sample averages imply different component types of a study characteristic yield estimates of different degrees of misalignment? Specifically, for a given study characteristic, we consider the null hypothesis of $E(Y_j) = E(Y_i)$, where Y_i and Y_j belong to different component types of the study characteristic, and the alternative hypothesis of the two expected values are different. The p-values of the two-tailed Wilcoxon Rank-Sum test are presented in the column labelled "Rank-Sum."

The Wilcoxon Rank-Sum test indicates that the RMB misalignment estimate is not significantly affected by whether an author is affiliated with an academic institution or not. Further, the test provides only a marginal evidence (a p-value of 0.11) of the "Other Approaches" type is different from other types of the theory underpinning characteristic. For other cases, the test shows that the RMB misalignment estimates from different types of a study characteristic are not the

same. That is, the nonparametric Wilcoxon Rank-Sum test suggests that, for most of the study characteristics, the reported degree of misalignment varies across their component types; some types can be associated with a larger estimate of undervaluation than others.

The standard errors of RMB misalignment estimates are quite large. Indeed, the coefficient of variation is larger than one for all the cases reported in Table 2. That is, the RMB misalignment estimates included in any one of these classification schedules display considerable variability. A possible interpretation is that the degree of misalignment is not precisely estimated.

Two remarks are in order. First, with the exception of the theoretical underpinning characteristic, we consider only two component types for each study characteristic. In doing so, we attempt to strike a balance between how fine we would like to group the estimates and the number of observations available under each type. Further, the estimation burden of the JMA meta-regression exercise is exponentially related to the number of study characteristic types.

Second, we do not conduct the usual publication bias analysis which requires information about the standard error of the RMB misalignment estimate. It is because studies in our sample typically do not provide the information. Our choices of study characteristics include the venue of publication, and the author's attributes. These study characteristics can shed light on other forms of publication biases beyond that implied by standard errors. Fidrmuc and Lind (2018) and Campos, *et al.* (2019), for example, show that institution-specific publication bias may be significant especially for politically sensitive issues.

3.3   Control Variables

After 1994, China has revamped its foreign exchange policy a few times. It moved to a *de facto* dollar peg after the 1997 Asian Financial Crisis. In July 2005, the *de facto* dollar peg was replaced with a "managed floating exchange rate regime." China replaced the managed regime with a stable RMB/dollar rate policy in the midst of the 2008 Global Financial Crisis, and reestablished it in 2011. That is, there are roughly five policy regime periods 1994–1997, 1998–2004, 2005–2008, 2009–2010, and 2011–2014.

These exchange rate regimes, together with accompanying capital control policies, reflect China's policy responses to both external and internal pressures on its exchange rate and economy. To the extent that the currency is managed, it interferes with the market adjustment process and has implications for currency misalignment. Indeed, the Wilcoxon Rank-Sum test indicates that

these policy regimes have RMB misalignment estimates of different average values – the period of 2005 to 2008 has the largest level of undervaluation and the period of 2011 to 2014 displays evidence of overvaluation (Appendix A.3). That is, China's exchange rate policy regime affects the misalignment level. Given this consideration, we include exchange rate policy regime dummy variables in (1) to control for policy influences in evaluating study characteristic effects.[15]

4. Jackknife Model Averaging Meta-Regression Analysis

The Wilcoxon Rank-Sum test results in Table 2 illustrate the bilateral interactions between individual study characteristics and RMB misalignment estimates. In this section, we adopt the JMA technique to conduct a meta-regression analysis of the study-to-study variability of RMB misalignment estimates.

4.1. The Procedure

The ability of (1) to reveal the effects of study characteristics depends on a proper specification of X_{ij} s – which are the study characteristics should be included and which should be excluded? One practical issue is that, for a given study characteristic, because of perfect linearity we cannot include all its component types, which are coded as a zero-one dummy variable. To circumvent perfect linearity, we study the effect of a type relative to the benchmark type. And, for each study characteristic, we label the characteristic type that is used by most studies as the benchmark type. The bold-font in Appendix A.2 identifies the selected benchmark study characteristic types. Not counting the ten benchmark types, the X_{ij} s in (1) can have up to $J = 12$ study characteristic types.

The lack of a strong theory on the appropriate set of study characteristics for explaining the variability of RMB misalignment estimates across empirical studies creates a technical issue. Indeed, disagreement over the choice of important study characteristics is always an issue (Stanley, 2001). A typical empirical solution is to select the "best" model from the model space that comprises all models constructed from the potential study characteristics. In the current case, the total number of contending models in the model space is $2^J = 2^{12} = 4,096$. Similar to uncertainty

---

[15] Some studies (Capon, *et al*., 1990; Doucouliagos and Ulubaşoğlu, 2008; Havranek, *et al*., 2015a, Valickova, *et al*., 2015) use time period dummy variables to capture improvements in methodology or data quality. Given China's special circumstances, we viewed the policy effect is an important exogenous element.

associated with parameter estimation, there is uncertainty associated with comparing and selecting models that can be nested or non-nested from the model space. If model selection uncertainty is simply ignored, the inference based on a chosen model may be overly optimistic and misleading (Buckland, *et al*., 1997; Fragoso, *et al*., 2018; Steel, 2019; Hoeting, *et al*., 1999).

We adopt the JMA technique which is a frequentist model averaging technique to evaluate effects of study characteristics. In essence, model averaging exploits the information from all (contending) models in the model space and accounts for model uncertainty in the process. Instead of identifying a single model, it derives the selected model combination and the combination of relevant study characteristics. Hoeting, *et al.* (1999), Fragoso, *et al.* (2018) and Steel (2019), for example, argue that model averaging offers better performance than selecting a single model.

The frequentist model averaging technique derives the weights for averaging models from empirical data – while the BMA approach also makes use of *priors*. The forecast combination exercise in Bates and Granger (1969) presents an early application of frequentist model averaging.[16] The JMA, compared with other frequentist model averaging methods that are based on, say, information criteria and the Mallows criterion (Buckland, *et al*., 1997; Hansen, 2007; Wan, *et al*., 2010), have two properties that are relevant for the current exercise. Specifically, the JMA allows the error term in (1) to be heteroskedastic and have a non-diagonal variance-covariance structure, and the contending models to be non-nested (Hansen and Racine, 2012; Zhang, *et al.*, 2013). In our sample, multiple RMB misalignment estimates are from individual studies. Even after accounting for study characteristics, the variability of misalignment estimates can cluster around individual studies, and display heteroskedasticity and a non-diagonal residual variance matrix. Further, there are both nested and non-nested contending models in the model space.

In this exercise, we derive the JMA weights as follows.[17] First, for model τ ($\tau = 1, \dots, 2^{12}$), we obtain its deleted-1 Jackknife residual vector $\hat{\varepsilon}_{\tau}$. The deleted-1 cross-validation criterion is given by $w'\hat{\varepsilon}'^{Jack}\hat{\varepsilon}^{Jack}w$, where $\hat{\varepsilon}^{Jack} = (\hat{\varepsilon}_1 \ \dots \ \hat{\varepsilon}_{2^{12}})$, and w is a 2^{12}x1 vector of model weights. The weights used to construct the Jackknife version of model averaging estimator are given by

$$\hat{w} = \underset{w \in \{w \leq 1, w \geq 0, w'1=1\}}{\arg\min} w'\hat{\varepsilon}'^{Jack}\hat{\varepsilon}^{Jack}w . \qquad (2)$$

---

[16] Recent applications include Wan and Zhang (2009) on tourism, and Magnus, *et al.* (2011) on housing market.
[17] Appendix A.4 gives a brief description of the JMA and other selected model average procedures.

The Jackknife averaging criterion incorporates covariances of residuals between models in the model space.

The JMA estimator of β_j ($j = 1, \ldots, 12$) and its standard error are given by

$$\hat{\beta}_j = \sum_{\tau=1}^{2^{12}} \hat{w}_\tau \hat{\beta}_{j,\tau} I(\beta_{j,\tau}),$$ (3)

and

$$\mathrm{SE}(\hat{\beta}_j) = \sum_{\tau=1}^{2^{12}} \hat{w}_\tau [Var(\hat{\beta}_{j,\tau}) + (\hat{\beta}_{j,\tau} - \hat{\beta}_j)^2]^{1/2} I(\beta_{j,\tau}),$$ (4)

where τ is the model index, $\hat{w}_\tau$ is the τ-th element of $\hat{w}$, $\beta_{j,\tau}$ is the coefficient of the study characteristic j ($X_{.j}$) included in model τ, $I(\beta_{j,\tau})$ is the indicator function that assumes the value of one if the study characteristic j is included in model τ and is zero otherwise, and $Var(.)$ is the variance operator.

By the same token, we define $\hat{\gamma}_k$ the JMA estimator of γ_k, the coefficient of the exchange rate policy regime control variable and its standard error. The Jackknife version of the model combination that is the best approximate of the true model is[18]

$$Y_i = \hat{\alpha} + \Sigma_{j=1}^J \hat{\beta}_j X_{ij} + \Sigma_{k=1}^K \hat{\gamma}_k Z_{ik} + \hat{\varepsilon}_i.$$ (5)

The contribution of model τ to the combination is given by its model weight $\hat{w}_\tau$. The model with the largest model weight is the best model relative to others in the model space identified by the JMA procedure. By combining the weights of models that include β_j, we get

$$\mathrm{IW}(\beta_j) = \sum_{\tau=1}^{2^{12}} \hat{w}_\tau I(\beta_{j,\tau}),$$ (6)

the total "weight" of β_j in the best approximate model combination (5). We label $\mathrm{IW}(\beta_j)$ the inclusion weight of β_j that indicates the relevance of the j-th study characteristic type.

It is noted that, under a frequentist model averaging framework, it is possible that different choices and criteria of selecting weights of the model averaging estimator give different results. Thus, besides presenting results from the JMA procedure in next subsection, we in the next section report results from a few alternative model averaging procedures, including BMA.

---

[18]   Hansen and Racine (2012) prove the optimality properties of JMA.

Table 3. Jacknife Model Averaging Results

	IW	MA	BEST
Weights			0.4126
Intercept		-0.0102	-0.0308
		(0.0491)[0.0649]	(0.0327) [0.0511]
Data Characteristics			
Non-Annual	0.7392	-0.0724	**_-0.0938_**
		(0.0232)[0.0351]	(0.0173)[0.0387]
PPP-based	0.8744	0.1089	**_0.1168_**
		(0.0274)[0.0390]	(0.0248)[0.0382]
Non-Dollar-based RER	0.6734	-0.0325	_-0.0357_
		(0.0172)[0.0258]	(0.0165)[0.0314]
Theoretical and Estimation Specifications			
FEER	0.1830	0.0260	
		(0.0215)[0.0227]	
Penn Effect	0.7899	0.1655	**_0.1970_**
		(0.0484)[0.0601]	(0.0391)[0.0586]
Other Approaches	0.0589	-0.0047	
		(0.0046)[0.0051]	
Non-Time-series	0.1008	0.0120	
		(0.0110)[0.0116]	
Cointegration	0.9182	0.1432	**_0.1806_**
		(0.0461)[0.0590]	(0.0254)[0.0474]
Structural	0.8170	0.1453	**_0.1872_**
		(0.0364)[0.0442]	(0.0203)[0.0348]
Publication Attributes			
Non-Journal	0.9014	0.0687	**_0.0810_**
		(0.0179)[0.0300]	(0.0172)[0.0313]
Mainland	0.9771	-0.1166	**_-0.1251_**
		(0.0174)[0.0392]	(0.0158)[0.0388]
Non-Academics	0.8487	0.0627	**_0.0728_**
		(0.0174)[0.0311]	(0.0154)[0.0352]
Exchange Rate Policy Regime			
1998-2004		-0.0147	-0.0160
		(0.0178)[0.0516]	(0.0167)[0.0515]
2005-2008		0.0054	-0.0013
		(0.0235)[0.0493]	(0.0206)[0.0482]
2009-2010		-0.0925	_-0.0974_
		(0.0305)[0.0715]	(0.0287)[0.0712]
2011-2014		-0.2599	**_-0.2656_**
		(0.0324)[0.0574]	(0.0306)[0.0559]
Adj.R2			0.3150

Notes: The intercept and exchange rate policy regime dummy variables are included in all regressions under the Jackknife model averaging exercise. The column labelled "IW" gives the inclusion weights of the corresponding characteristic types to the left. The column labelled "MA" gives the Jackknife model averaging estimates and their weighted standard errors. Two kinds of weighted standard errors are calculated: one based on White standard errors in single regressions is in parentheses and one based on study clustered standard errors in single regressions is in squared brackets. The column labelled "Best" gives the OLS results of the model that has the largest model weight, which is given in the row labelled "Weights". The coefficient estimates significant at the 10% level under the White robust standard errors are underlined, and those significant at the 10% level under clustered robust standard errors are in bold.

4.2   Results

Table 3 presents results from the JMA analysis of RMB misalignment estimates. The inclusion weights and the JMA estimates of β_j s are reported under the columns labelled, respectively, "IW" and "MA."

By construction, inclusion weights lie between zero and one. However, we do not have a theory to determine the significance of an inclusion weight value. If we relate the notion of inclusion weight to the concept of posterior inclusion probability (PIP) of the BMA approach, we can consider the *j*-th study characteristic type to have an acceptable, substantial, strong, or decisive effect if IW(β_j) is between, respectively, 0.5 and 0.75, 0.75 and 0.95, 0.95 and 0.99, and 0.99 and 1. If IW(β_j) is less than 0.5, the *j*-th study characteristic type is considered ignorable.[19] According to the rule of thumb of 0.5, we identify nine non-ignorable study characteristic types; three for each of the three categories of study characteristics.

Under the column labelled "MA," two model averaging standard errors are reported underneath each weighted estimate. The weighted White robust standard errors are in round parentheses, and the weighted standard errors robust to clustering on study effects are in squared brackets. The t-statistics constructed from either type of standard errors indicate statistical significance is associated with inclusion weights. For study characteristic type that has an inclusion weight larger than 0.7, its t-statistic is larger than, say, the value 1.645. The case of non-dollar based RER that has an inclusion weight of 0.6734 (larger than 0.5) is a marginal case – the t-statistic based on the White robust standard error is -1.890 and the one robust to clustering on study effects is -1.260. However, Zhang and Liu (2018) show that these weighted model averaging coefficients do not have standard asymptotic distributions. Therefore, we have to interpret these t-statistics with cautions.

To shed additional light on the relevance of these study characteristics, we report under the heading "BEST" the results of estimating the model that garnered the largest model weight $\hat{w}_\tau$. The best model has a model weight of 41.3%, and includes nine study characteristic types.[20] Indeed,

---

[19]   The next section presents a BMA analysis. If PIP falls between 0.5–0.75, 0.75–0.95, 0.95–0.99 and 0.99–1, respectively, the corresponding variable type has an acceptable, substantial, strong, or decisive effect (Kass and Raftery, 1995; Havranek, *et al*., 2015). If it is less than 0.5, the variable is considered ignorable.
[20]   The second best model has a model weight of 22.4% and does not include the "non-dollar based RER" variable. A plot of the model weights is in Appendix A.5.

these included variables coincide with those with an inclusion weight larger than 0.5. Again, the White robust and clustering-on-study robust standard errors are presented in round parentheses and squared brackets, respectively. The included study characteristic types are statistically significant under clustering robust standard errors, with the exception of the "non-dollar based RER" variable that is significant under the White robust standard error. The coefficient estimates are quite close to the corresponding weighted estimates from model averaging.

The different pieces of evidence in Table 3 together yield strong evidence for eight study characteristic effects, and marginal evidence for one. The multivariate framework that controls for inter-dependence between study characteristics identifies a smaller number of significant study characteristic types than the one indicated by the Wilcoxon Rank-Sum test results in Table 2.

It is of interest to look at the significant factors even though we do not have a sharp theory on effects of study characteristics. Thus, we take a "pragmatic" approach and interpret the significant effects in Table 3 without connotations of rejecting or validating any theory or hypothesis.

Of the three data characteristic types, the use of non-annual tends to yield relatively weak evidence of RMB undervaluation; suggesting that high frequency (quarterly and monthly) data tend to give weaker evidence than low (annual) frequency data. One possible interpretation is that, compared with annual data, quarterly and monthly data can better capture short-term effects of currency market shocks – a point that is in accordance with the discussion of long-run and short-run equilibrium real exchange rates (Montiel and Hinkle, 1999).[21]

The use of PPP-based data tends to yield relatively strong evidence of RMB undervaluation. PPP-based data account for differences in relative prices in individual countries to facilitate cross-country comparison of real exchange rates and, hence, misalignment. The Chinese economy has experienced different growth patterns in different sectors in the past few decades. To the extent that the evolution of market prices and the exchange rate does not reflect relative productivity growth, the uses of PPP-based and market-based data can generate different misalignment results.

---

[21]   Bineau (2010), Cheung and He (forthcoming) and Korhonen and Ritola (2011) reported similar negative, but statistically insignificant, non-annual data effects. Appendix A.6 presents a comparison of results from these studies. In general, the results are usually consistent in terms of the sign, but may differ in terms of significance. The difference in significance is likely due to differences in the samples. In the case of Cheung and He (forthcoming), the omission of 12 extreme estimates from Chang (2008) – see footnote 9 – is the main source of differences.

Our results indicate that, after controlling for these cross-country data differences, it is more likely to report evidence of RMB undervaluation.

The significance of the non-dollar-based RER variable is not robust and, thus, we consider the evidence is weak. The result suggests the possibility that studies using dollar-based RER tend to yield stronger evidence of RMB undervaluation. If currencies move in a similar pattern against the US dollar, then the level of RMB misalignment against non-dollar currencies can be less than against the US dollar. Our finding offers weak support for this interpretation.

It is found that, under the category of theoretical and estimation specifications, the Penn Effect, Cointegration, and Structural study characteristic types are significant with a positive coefficient estimate. That is, these characteristics tend to be associated with an undervalued RMB misalignment estimate.

The Penn Effect setup, which is based on the well-documented empirical link between real exchange rates and per-capita income is commonly used to evaluate RMB valuation. Among a variety of theoretical explanations, most discussions refer to the Balassa-Samuelson hypothesis that draws on differential trade-sector and non-trade-sector productivity growth rates. Compared with the non-trade-sector, China's trade-sector, evidently, has experienced a very strong growth in the past few decades. The differential growth patterns across sectors affect the evolution of China's equilibrium exchange rate.[22] The Penn Effect setup that allows for differential productivities is likely to suggest the RMB is undervalued.

The cointegration specification incorporates empirical long-run relationship in the estimation process. The positive Cointegration study characteristic effect suggests that allowing for empirical long-run linkages can lead to strong evidence of RMB undervaluation. Our significant "Structural" characteristic type result is comparable to Wang and Yao (2008) who find that structural setup usually gives larger misalignment estimates relative to reduced-form setup.

Beside these three characteristic types, the other theoretical and estimation specification characteristic types (FEER, Other Approaches, and Non-Time-series) have low inclusion weights and are insignificant under the "BEST" specification.

---

[22]   Theoretically, the Balassa-Samuelson mechanism assumes intersectoral factor mobility so that the equilibrium exchange rate is consistent with internal balance and external equilibrium. China, however, is unlikely to satisfy this assumption.

All the three types of publication attributes in Table 3 have an inclusion weight larger than the 0.5 threshold and are statistically significant under the "BEST" specification. That is, the publication venue and the author's affiliation have implications for the reported RMB misalignment estimates. Specifically, a weak empirical evidence of RMB undervaluation is likely to be reported by studies published in academic journals, or authored by researchers affiliated with a mainland China institution or a non-academic institution.

As noted in Section 3, we do not have the information on standard errors to conduct the publication bias about RMB misalignment estimates. Our results pertaining to publication attributes, however, offer an alternative perspective to assess the behavior of reporting politically sensitive RMB misalignment estimates (Fidrmuc and Lind, 2018; Campos, *et al.*, 2019). While we do not pretend to test a sharp hypothesis, the negative association between a mainland institution affiliation and RMB undervaluation seems to be non-controversial.[23]

In the presence of these study characteristics, the four exchange rate policy regime dummy variables have a negative coefficient estimate. However, only the time dummy variable of 2011-2014 is statistically significant; indicating that weak evidence of RMB undervaluation is likely to be found for this period. The result is in accordance with, say, the Wilcoxon Rank-Sum test result mentioned in subsection 3.3 (Appendix A.3), the finding of Cheung, *et al.* (2017), and the International Monetary Fund 2015 assessment.

4.3   Alternative Study Characteristic Benchmarks

In the previous subsection, study characteristic effects are assessed with reference to benchmarks defined by the most adopted study characteristic types. The choice of benchmarks can be viewed as arbitrary. Here we explore the sensitivity of alternative benchmark choices. Specifically, we swap the roles of regressors and benchmarks in the previous subsection. That is, based on equation (1), we use the 10 benchmark study characteristic types identified in Appendix A.2 as regressors X_{ij}'s to study the effects of these study characteristics.

---

[23]   Similar results of author's affiliation are reported by Korhonen and Ritola (2011).

Table 4. Jackknife Model Averaging - Benchmark Study Characteristic Types as Regressors

	IW	MA	BEST
Weights			0.4090
Intercept		0.3444	**0.4112**
		(0.0757)[0.1070]	(0.0399) [0.0897]
Data Characteristics			
Annual	0.4842	0.0425	
		(0.0239)[0.0307]	
Non-PPP-based	0.8109	-0.1017	**-0.0903**
		(0.0352)[0.0480]	(0.0188)[0.0402]
Dollar-based RER	0.9466	0.0655	**0.0658**
		(0.0161)[0.0316]	(0.0163)[0.0331]
Theoretical and Estimation Specifications			
BEER	0.6196	-0.1036	**-0.1724**
		(0.0429)[0.0560]	(0.0274)[0.0661]
Time-series	0.7368	-0.0899	**-0.1208**
		(0.0281)[0.0375]	(0.0195)[0.0372]
Non-Cointegration	1.0000	-0.1507	**-0.1859**
		(0.0507)[0.0778]	(0.0326)[0.0757]
Reduced-form	0.3804	-0.0500	
		(0.0318)[0.0345]	
Publication Attributes			
Journal	0.7790	-0.0552	**-0.0794**
		(0.0180)[0.0277]	(0.0148)[0.0289]
Non-mainland	0.9681	0.1010	**0.1085**
		(0.0160)[0.0376]	(0.0154)[0.0382]
Academics	0.9147	-0.0385	_-0.0447_
		(0.0131)[0.0327]	(0.0132)[0.0337]
Exchange Rate Policy Regime			
1998-2004		-0.0058	-0.0113
		(0.0181)[0.0535]	(0.0169)[0.0508]
2005-2008		0.0127	0.0036
		(0.0227)[0.0496]	(0.0206)[0.0473]
2009-2010		-0.0838	_-0.0974_
		(0.0316)[0.0720]	(0.0284)[0.0703]
2011-2014		-0.2468	**-0.2566**
		(0.0322)[0.0598]	(0.0305)[0.0573]
Adj.R2			0.2965

Notes: See notes to Table 3.

Table 4 presents the corresponding JMA meta-regression analysis. The inclusion weights and the JMA estimates are given under the "IW" and "MA" columns. While the signs of weighted average estimates are consistent with those in Table 3,[24] the 0.5 inclusion weight rule of thumb indicates that there are eight non-ignorable study characteristics. Comparing with results in Table 3, the study characteristics under the category of data characteristics have the corresponding (opposite) signs but the annual data type has an inclusion weight less than 0.5; that is there is one less non-ignorable study characteristic. The category of theoretical and estimation specifications yields the same number – three – of non-ignorable study characteristics. However, only two overlap with those in Table 3. The "time-series" is found to be non-ignorable and tends to yield weak evidence of RMB undervaluation. The "reduced-form" type is an ignorable type – a result that is at odds with the finding of a non-ignorable "structural" type in Table 3. The results of the three publication attribute study characteristics are qualitative similar to those in Table 3.

The model with the largest model weight $\hat{w}_\tau$ ($\hat{w}_\tau = 40.9\%$) is presented under the "BEST" column.[25] It includes the eight study characteristic types that have an inclusion weight larger than 0.5. In terms of statistical significance, the use of standard errors robust to clustering-on-study, instead of the White robust standard errors, reduces the number of significant coefficient estimates from eight to seven – the "academics" type becomes insignificant after controlling for clustering effects.

The evidence presented in Tables 3 and 4 indicates that the reported RMB misalignment estimates vary systematically across studies; depending on whether a) PPP-based data, b) dollar-based RER, c) BEER model, and d) cointergration method are used. Also it depends on whether it is published in academic journals, and whether authors are affiliated with a China institution, or not. The evidence is weak when a study uses annual data, a time series setup, or a reduced-form model. Further, compared with other historical periods, the RMB misalignment estimates reported for 2011-2014 tend to yield a smaller degree of undervaluation.

*4.4   Misaligned or Not*

---

[24]   Under the theoretical setting, the non-ignorable BEER type has a sign opposite to that of the non-ignorable Penn type.
[25]   The second best model has a model weight of 26.3% and includes five study characteristics, which also are in the best model. A plot of the model weights is in Appendix A.5.

As discussed earlier, differing theoretical views and empirical strategies can lead to diverse and even conflicting RMB misalignment estimates in our sample. Our quantitative analysis shows that, in addition to data choices and publication attributes, these empirical estimates vary with the choices of theoretical frameworks and empirical strategies. Given these results, what is the inference about RMB misalignment we can draw from this large set of RMB misalignment estimates? To address this question, we consider equation (5), which is the JMA version of the best approximate of the true model.

One strategy is to consider the distribution of Y_i assuming X_{ij}'s are at their average values. Recall that the JMA coefficient estimates are weighted averages of the corresponding estimates from individual model specifications in the model space. We follow Buckland, *et al.* (1997) to calculate the mean and variance of these weighted estimates (Appendix A.7).

Figure 2. The Probability of the RMB Misalignment: Five Exchange Rate Policy Regimes
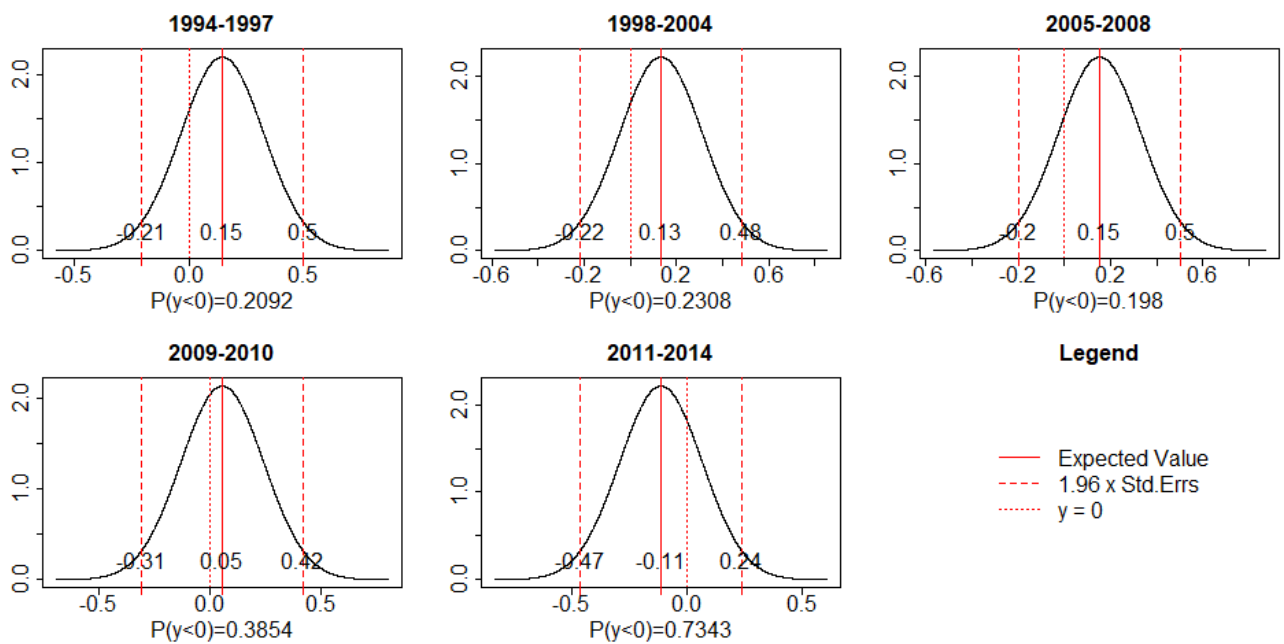
Notes: The Figure plots the probability density of the RMB misalignment estimator implied by the JMA model combination (equation 5) that is the best approximate of the true model for each of the five exchange rate policy regime periods. The solid line indicates the expected level of misalignment, the two dash lines present the symmetric 95% confident interval, and the reference point of zero is given by the doted red line. The numbers above the red ticks are the 2.5%, 50% and 97.5% quantiles. The equation P(y<0) under the x-axis shows the probability of the RMB is overvalued.

Figure 2 presents the density plots of the resulting RMB misalignment for each of the five periods. The mean RMB misalignment estimate is positive and indicative of undervaluation in the first four historical periods, and is negative and indicative of overvaluation in the last one. The probability of an overvaluation estimate ranges from 19.8% for 2005-2008 to 73.43% for 2011-2014. Indeed, the 95% confidence intervals of these plots cover both positive and negative values and, thus, offer no statistically significant evidence to substantiate either the claim that the RMB is undervalued or is overvalued.

An astute reader may question if the use of mean values of the study characteristics offers the best representation. To shed further light on the issue, we consider another four alternative choices of combining study characteristic types. They are the combinations of data characteristics and theoretical and estimation specifications that are a) the most frequently used, b) the least frequently used, c) those used in the most cited English study and d) those used in the most cited Chinese study. For each of these four composite cases, the relevant publication attributes assume their respective sample average values. In other words, we extract information about RMB misalignment conditional on selected choices of study characteristics of data, theoretical and estimation specifications, and the averages of publication attributes.

In forming composites of study characteristics, we note that some study characteristic types cannot be combined together at the same time. For instance, the adoption of a BEER specification does not go with a non-cointegration setup. Focusing only on possible choices, the combination of the most frequently used data characteristics and theoretical and estimation specifications includes these study characteristic types: Annual, Non-PPP-based, Dollar-based RER, BEER, Time series, Cointegration and Reduced-form, and the combination of the least frequently used ones are Non-Annual, PPP-based, Non-Dollar-based ER, Other Approaches, Non-Time-series, Cointegration and Structural.[26]

The density plots of the RMB misalignment under the specifications of the most and least frequently used study characteristic types are presented in Figures 3a and 3b. Two observations are in order. First, the mean misalignment estimates in Figure 3a are smaller than the corresponding ones in Figure 3b. On the average, a study equipped with the least employed – compared with the most employed - data characteristic and specifications of theoretical and estimation setups tends to yield a stronger evidence of RMB undervaluation.

---

[26]   The Non-Time series type includes the panel setup and, thus, can be combined with cointegration.

Figure 3a. The Probability of the RMB Misalignment Implied by the Most Frequently Used Study Characteristics: Five Exchange Rate Policy Regimes
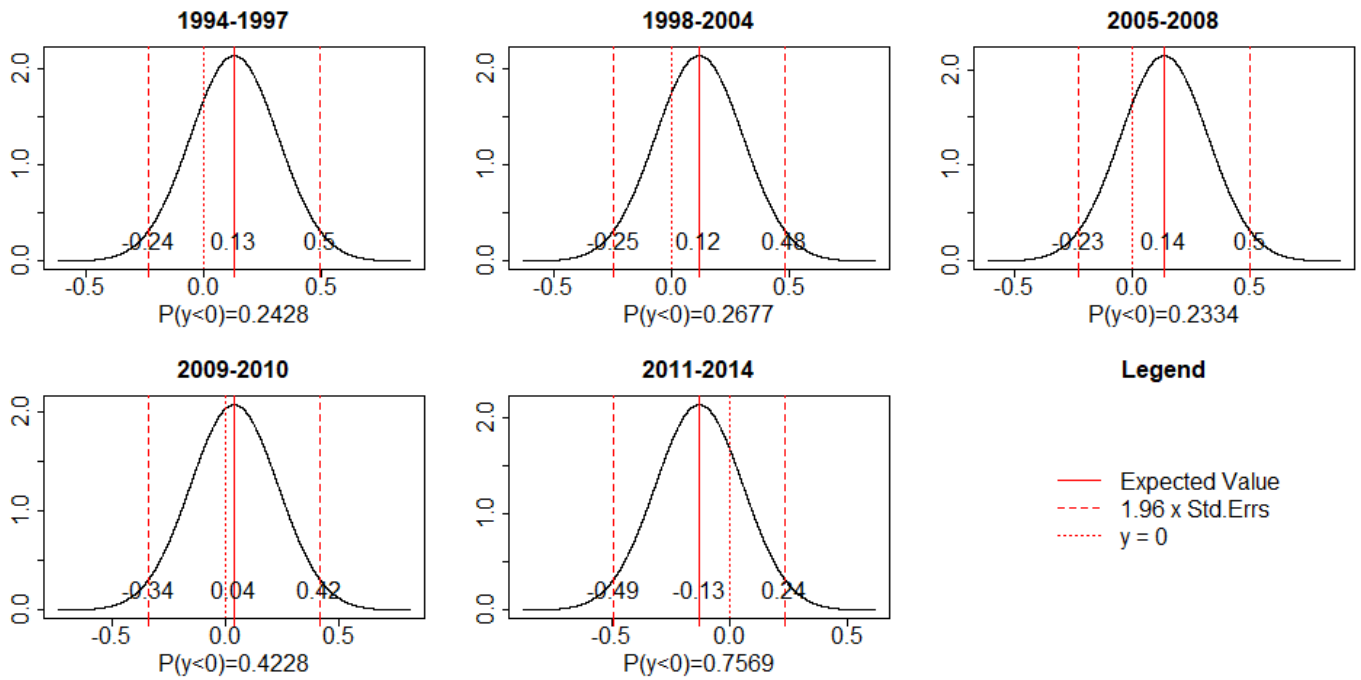
Notes: See notes to Figure 2 and the text.

Figure 3b. The Probability of the RMB Misalignment Implied by the Least Frequently Used Study Characteristics: Five Exchange Rate Policy Regimes
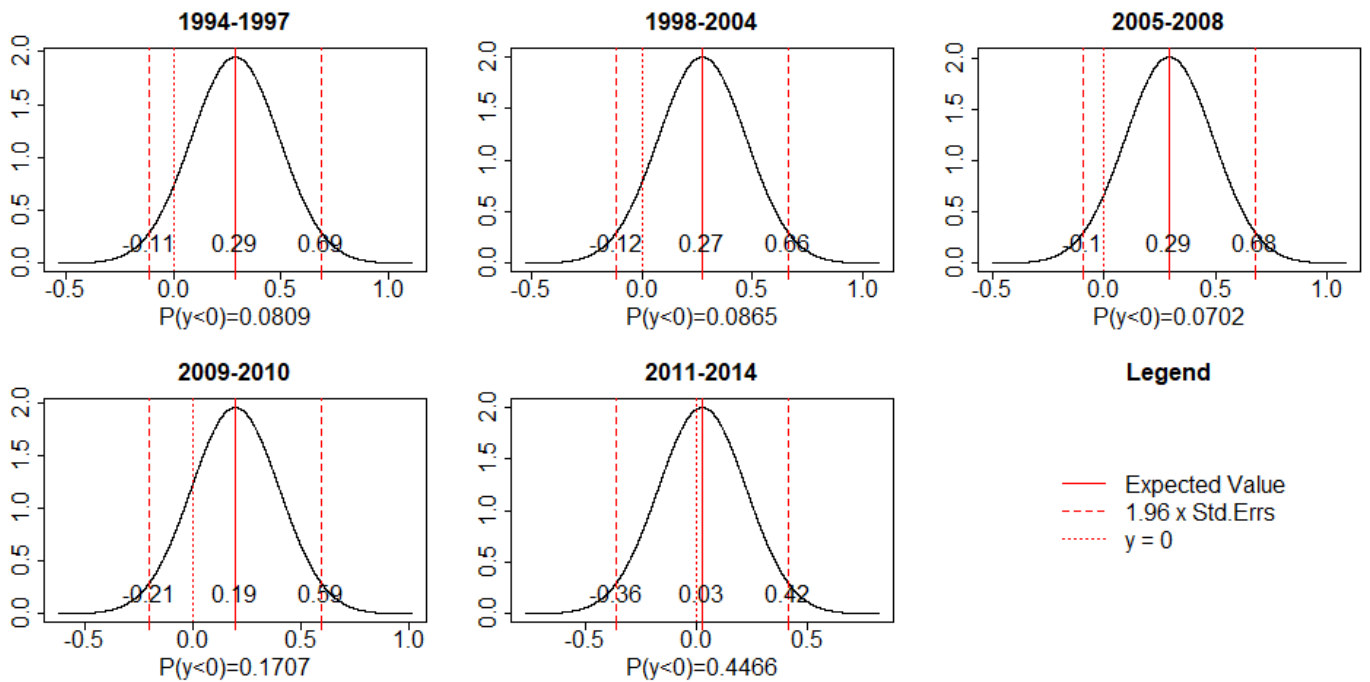
Notes: See notes to Figure 2 and the text.

Figure 4a.   The Probability of the RMB Misalignment Implied by the Study Characteristics adopted by the Most Cited English Study: Five Exchange Rate Policy Regimes
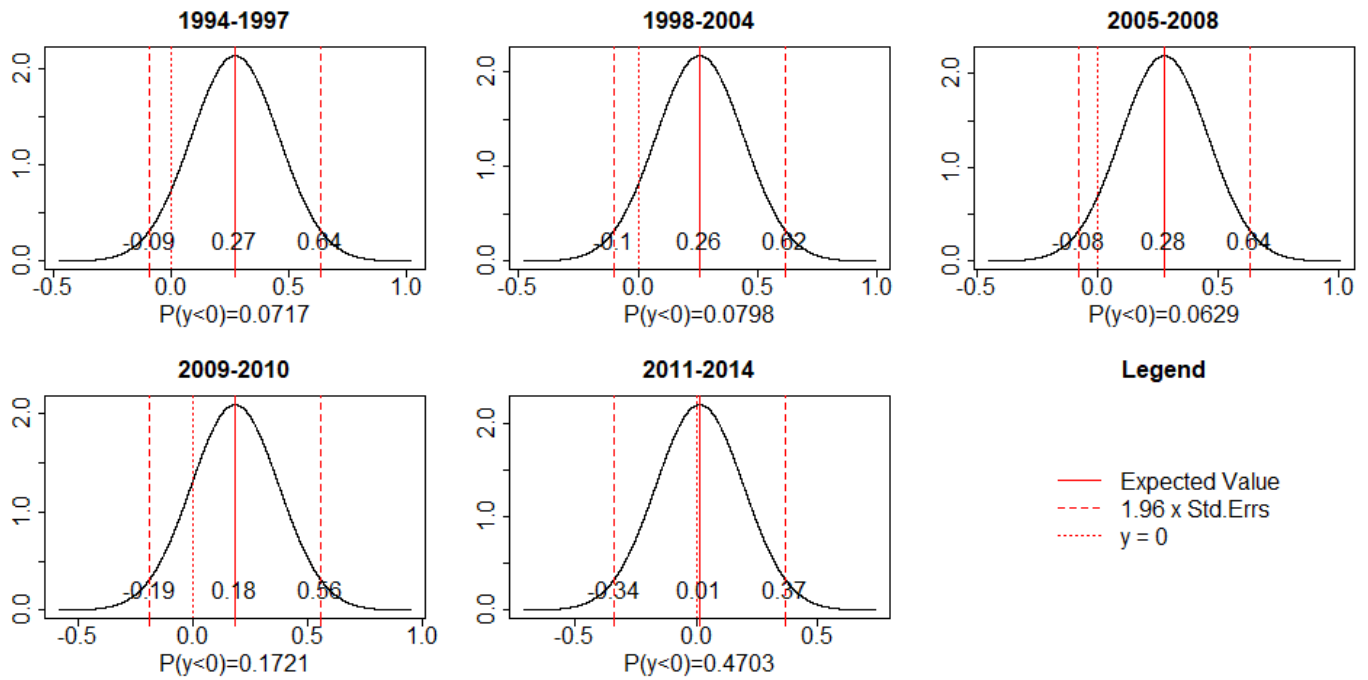
Notes: See notes to Figure 2 and the text.

Figure 4b.   The Probability of the RMB Misalignment Implied by the Study Characteristics adopted by the Most Cited Chinese Study: Five Exchange Rate Policy Regimes

Notes: See notes to Figure 2 and the text.

Second, the null hypothesis of zero misalignment is not rejected based on the 95% confidence interval criterion. Relatively speaking, the composite of least used study characteristic types yields a smaller probability mass of a negative (overvaluation) misalignment estimate; specifically, there is less than a 10% chance to observe an overvalued RMB misalignment estimate for the periods of 1994-1997, 1998-2004, and 2005-2008. Of course, one have to assert if it represents a reasonable statistical evidence of RMB undervaluation.
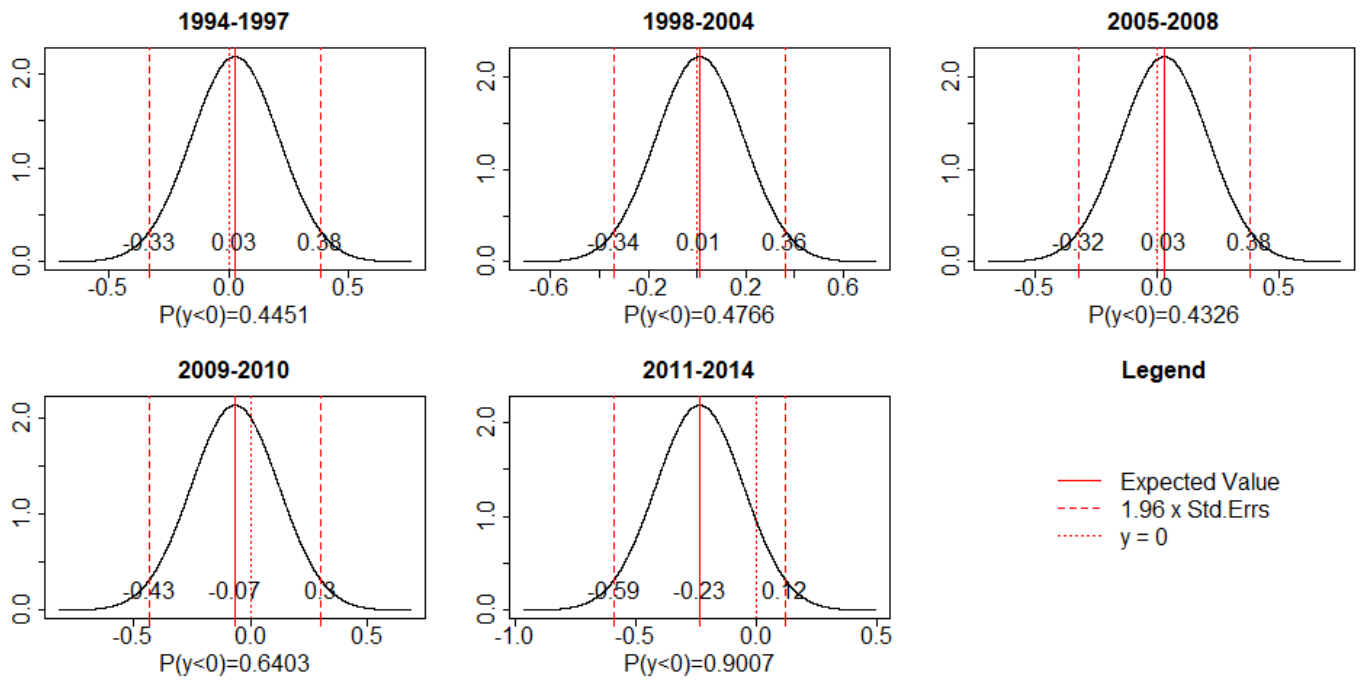
Figure 4a and 4b present density plots of the RMB misalignment when a study shares characteristics on data, theoretical setup and estimation procedure with the most cited, respectively, English and Chinese study. We consider the English and Chinese studies separately because they possibly cater for different readerships. The citation count can be interpreted as a barometer of profession's assessment of the importance and/or quality of the study. In our sample of studies, Cheung, *et al.* (2007) is the study in English, and Shi and Yu (2005) is the one in Chinese that has the highest number of (total and per-year) citations. Further, both are published in academic journals, and met the expected academic vigor implied by the respective peer-review processes.[27] Specifically, Cheung, *et al.* (2007) employ annual, PPP-based real exchange rate against US dollar, and work with a reduced form Penn effect setup using panel data. Shi and Yu (2005), on the other hand, use quarterly, non-PPP based effective real exchange rate data, and conduct the analysis using the BEER approach with time-series, cointegration and reduce-form setup.

The means of the density plots in Figure 4a are noticeably larger than the corresponding ones in Figure 4b – the differences are larger than 20%. The study characteristics adopted by the well-cited Chinese study tend to yield less undervalued estimates – a result that echoes the finding that affiliation with a China's institution is associated with less undervalued estimates.

Again, the 95% confidence interval criterion does not reject the no-misalignment null hypothesis. The specification according to the well cited Chinese study, compared with the one of the well-cited English study, offers a stronger evidence of non-rejection; especially for the first three historical periods.

5.   Robustness Check

---

[27]   As of March 11, 2019, Cheung, *et al.* (2007) has a citation count of 293, followed by 277 of Frankel (2006). Among the Chinese studies, Shi and Yu (2005) has a citation count of 196, followed by 92 of Zhang (2000). The journals in which Cheung, *et al.* (2007) and Shi and Yu (2005) have the highest impact factor/ranking in the respective collections of English and Chinese journals.

In this section, we present results from a few alternative model averaging procedures, which are described in Appendix A.4. The first alternative is the frequentist method based on the heteroskedasticity-robust Mallows model averaging criterion (Hansen, 2007; Liu and Okui, 2013). The corresponding vector of model averaging weights, $\hat{w}_{rM}$, is obtained by minimizing the Mallows model averaging criterion $C_{Mallows}(w)$,

$$\hat{w}_{rM} = \underset{w \in \{w \leq 1, w \geq 0, w'1 = 1\}}{\arg\min} C_{Mallows}(w), \tag{8}$$

where

$$C_{Mallows}(w) = w'\hat{\varepsilon}'^{OLS}\hat{\varepsilon}^{OLS}w + 2tr(\Omega)'Pw, \tag{9}$$

$\hat{\varepsilon}^{OLS} = (\hat{\varepsilon}_1^{OLS}, ..., \hat{\varepsilon}_{2^{12}}^{OLS})$ contains least squares residual estimates from models in the model space, Ω is the heteroskedastic diagonal variance-covariance matrix, $P = [tr(Q_1(Q_1'Q_1)^{-1}Q_1')', ..., tr(Q_{2^{12}}(Q_{2^{12}}'Q_{2^{12}})^{-1}Q_{2^{12}}')']$, and Q_τ is the matrix of explanatory variables of model τ ($\tau = 1, ..., 2^{12}$). The heteroskedasticity-robust Mallows model averaging estimator of β_j ($j = 1, ..., J$) is

$$\hat{\beta}_{j,rM} = \sum_{\tau=1}^{2^{12}} \hat{w}_{\tau,rM}\hat{\beta}_{j,\tau}I(\beta_{j,\tau}), \tag{10}$$

where $\hat{w}_{\tau,rM}$ is the τ-th element of $\hat{w}_{rM}$.

The results of the heteroskedasticity-robust Mallows model averaging are presented in Table 5 under the column "R MMA." It is apparent that the inclusion weights and weighted coefficient estimates are almost the same as those obtained from the JMA method. In fact, both methods give the same set of study characteristics that have an inclusion weight larger than 0.5 (column "BEST"). They yield the same model specification that has the highest model weight, and includes all the study characteristics with an inclusion weight larger than 0.5. That is, both methods yield very similar evidence of effects of study characteristics.

The robust Mallows estimation results suggest that the estimates in the two periods 2009 to 2010 and 2011 to 2014 show a relatively weak evidence of undervaluation. However, we verified that, similar to the case of Jackknife approach, the 2009-2010 result is spurious in the sense that it becomes insignificant once clustering effects are controlled for. Similar significant results for 2009-2010 estimates reported for other cases in Table 5 are also found to be spurious and sensitive to clustering effects.

Table 5. Model Averaging Results Derived from the Mallows Criterion, AIC Method, BIC Method, and the Bayesian Model Averaging Approach

	R MMA			AIC			BIC			BMA		
	IW	MA	BEST	IW	MA	BEST	IW	MA	BEST	PIP	Post Est	BEST
Weights			0.4318			0.2230			0.5989			0.6004
Intercept		-0.0113	-0.0308		-0.0287	-0.0308		-0.0452	**-0.0535**		-0.0451	-0.0534
		(0.0503)	(0.0327)		(0.0330)	(0.0286)		(0.0300)	(0.0268)			
Data Characteristics												
Non-Annual	0.7492	-0.0725	**-0.0938**	0.9974	-0.0854	**-0.0938**	0.9961	-0.0904	**-0.0906**	0.9960	-0.0903	**-0.0905**
		(0.0227)	(0.0173)		(0.0220)	(0.0196)		(0.0200)	(0.0196)		(0.0208)	(0.0194)
PPP-based	0.8980	0.1101	**0.1168**	1.0000	0.1164	**0.1168**	1.0000	0.1139	**0.1123**	1.0000	0.1137	**0.1122**
		(0.0264)	(0.0248)		(0.0222)	(0.0214)		(0.0217)	(0.0214)		(0.0216)	(0.0213)
Non-Dollar-based RER	0.6827	-0.0322	**-0.0357**	0.8682	-0.0331	**-0.0357**	0.3213	-0.0116		0.3196	-0.0116	
		(0.0170)	(0.0165)		(0.0146)	(0.0157)		(0.0094)			(0.0191)	
Theoretical and Estimation Specifications												
FEER	0.1835	0.0252		0.3739	0.0140		0.0486	0.0017		0.0485	0.0017	
		(0.0209)			(0.0173)			(0.0024)			(0.0119)	
Penn Effect	0.8028	0.1660	**0.1970**	0.9999	0.1884	**0.1970**	0.9995	0.2105	**0.2196**	0.9995	0.2103	**0.2194**
		(0.0471)	(0.0391)		(0.0357)	(0.0283)		(0.0303)	(0.0266)		(0.0309)	(0.0264)
Other Approaches	*0.0749*	-0.0056		0.3041	-0.0041		0.0419	-0.0009		0.0419	-0.0009	
		(0.0054)			(0.0115)			(0.0016)			(0.0080)	
Non-Time-series	0.0957	0.0117		0.3337	0.0060		0.0430	0.0007		0.0430	0.0007	
		(0.0108)			(0.0091)			(0.0012)			(0.0058)	
Cointegration	0.8976	0.1437	**0.1806**	1.0000	0.1691	**0.1806**	0.9994	0.1753	**0.1756**	0.9994	0.1751	**0.1754**
		(0.0433)	(0.0254)		(0.0285)	(0.0224)		(0.0237)	(0.0224)		(0.0243)	(0.0223)
Structural	0.8167	0.1466	**0.1872**	0.9996	0.1711	**0.1872**	0.9993	0.1864	**0.1892**	0.9993	0.1862	**0.1890**
		(0.0362)	(0.0203)		(0.0302)	(0.0183)		(0.0200)	(0.0183)		(0.0215)	(0.0182)
The Publication Attributes												
Non-Journal	0.8893	0.0687	**0.0810**	0.9994	0.0775	**0.0810**	0.9971	0.0797	**0.0800**	0.9970	0.0796	**0.0799**
		(0.0178)	(0.0172)		(0.0189)	(0.0183)		(0.0184)	(0.0184)		(0.0189)	(0.0183)
Mainland	0.9718	-0.1164	**-0.1251**	1.0000	-0.1231	**-0.1251**	1.0000	-0.1192	**-0.1166**	1.0000	-0.119	**-0.1165**
		(0.0173)	(0.0158)		(0.0161)	(0.0156)		(0.0159)	(0.0152)		(0.0158)	(0.0151)
Non-Academics	0.8421	0.0629	**0.0728**	0.9999	0.0724	**0.0728**	0.9998	0.0802	**0.0842**	0.9998	0.0802	**0.0841**
		(0.0172)	(0.0154)		(0.0155)	(0.0144)		(0.0149)	(0.0136)		(0.0149)	(0.0135)
Exchange Rate Policy Regime												
1998-2004		-0.0147	-0.0160		-0.0158	-0.0160		-0.0182	-0.0193		-0.0182	-0.0193
		(0.0177)	(0.0167)		(0.0171)	(0.0170)		(0.0171)	(0.0170)		(0.0170)	(0.0169)
2005-2008		0.0054	-0.0013		0.0005	-0.0013		-0.0016	-0.0022		-0.0016	-0.0022
		(0.0238)	(0.0206)		(0.0202)	(0.0200)		(0.0201)	(0.0200)		(0.0200)	(0.0199)
2009-2010		-0.0925	**-0.0974**		-0.0966	**-0.0974**		-0.0978	**-0.0982**		-0.0977	**-0.0981**
		(0.0306)	(0.0287)		(0.0293)	(0.0292)		(0.0293)	(0.0293)		(0.0291)	(0.0291)
2011-2014		-0.2601	**-0.2656**		-0.2654	**-0.2656**		-0.2655	**-0.2655**		-0.2652	**-0.2652**
		(0.0326)	(0.0306)		(0.0264)	(0.0264)		(0.0264)	(0.0264)		(0.0263)	(0.0263)
Adj.R2			0.3150			0.3150			0.3119			

Notes: The columns labelled "R MMA," "AIC," and "BIC" present the frequentist moving averaging results derived from, respectively, the heteroskedasticity-robust Mallows model averaging criterion, the AIC and BIC. The column labelled "BMA" presents the results derived from the Bayesian model averaging approach. The posterior inclusion probabilities, the posterior means and their posterior standard deviations in parentheses, and the results pertaining to the model with the largest posterior inclusion probability are presented under, respectively, the headings of "PIP," "Post Est," and "BEST". Coefficient estimates that are significant at the level of 10% or better are in bold. See the Notes to Table 3.

We report the results of model averaging results based on the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) under the columns labelled "AIC" and "BIC." The information criterion approach uses estimates of information measure of models in the model space to derive their relative model weights (Buckland, *et al.*, 1997; Burnham and Anderson, 2002).

The evidence provided by these two alternative model averaging methods are quite comparable to those of the Jackknife and heteroskedasticity-robust Mallows method, despite some minor differences. For instance, for study characteristic types such as "Non-dollar based RER," "FEER," "Other Approaches," and "NonTime-Series," the AIC approach gives a larger while the BIC approach a smaller inclusion weight. The model with the highest weight identified by the AIC approach (column "BEST") is the same as those by the Jackknife and heteroskedasticity-robust Mallows method. The "best" model identified by the BIC approach (Column "BEST"), on the other hand, includes one less study characteristic – "Non-Dollar-based RER." In general, these frequentist model averaging methods offer qualitatively similar evidence on study characteristic effects.

In addition to frequentist model averaging, we consider a BMA alternative. To implement BMA, we adopted two commonly used conservative priors: the uniform prior probability on the model space is used to generate posterior model probabilities and the unit information prior g-UIP for parameter estimation (Zeugner and Feldkircher, 2015). The BMA results are presented under the column "BMA" in Table 5.

The posterior inclusion probability of a study characteristic type reported under the heading of "PIP" is given by the sum of the posterior probabilities of models that include the type, and is used to determine whether the variable should be included in the regression. The usual rule of thumb is to label a variable non-ignorable if its PIP is larger than 0.5 (footnote 22). The BMA identifies eight non-ignorable study characteristic types. The model specification that includes non-ignorable variables is known as the median probability model. In the current exercise, the median probability model and the highest probability model, which is the model specification that has the highest posterior model probability (Barbieri and Berger, 2004), are the same. The highest probability model presented under the column labelled "BEST" garners a posterior model probability of 60%, and is quite similar to the "best" model identified by the BIC based frequentist model averaging.

For completeness, we applied these alternative model averaging techniques to the specification with the benchmark study characteristics as regressors. The results are very similar to those obtained from the JMA approach. These results are presented in Appendix A.8 for brevity.

Overall, the empirical effects of the study characteristics on RMB misalignment estimates are quite robust to these alternative model averaging methods.

6.   Concluding Remarks

In this study, we conduct a meta-regression analysis of RMB misalignment estimates using the JMA method to account for model selection and estimation uncertainties, and allow for heteroskedasticity and non-nested models. We consider a sample of 925 RMB misalignment estimates in percentages that span from 1994 to 2014 obtained from 69 studies.

The JMA meta-analysis shows that these misalignment estimates are systematically affected by the data used, the adopted theoretical setup and empirical strategy, and publication attributes. Specifically, studies that use PPP-based data, a Penn effect setup, a cointegration specification, or is a non-academic journal publication are likely to give strong evidence of RMB undervaluation. Studies that adopt a BEER setup, use a non-cointegration technique, have an author affiliated with a mainland China institution, or generate estimates for the period 2011 to 2014 tend to yield weak evidence of RMB undervaluation. Less conclusive evidence is found for the effects of the use of non-dollar based RER data, time series setup, reduced-form regression, and author's academic affiliation. These empirical findings are quite robust to the choice of benchmark study characteristic types, and to alternative model averaging methods including the heteroskedasticity-robust Mallows approach, the AIC and BIC approaches, and the BMA.

When we extend the analysis to infer the evidence of RMB misalignment, it is hard to draw a definte conclusion. Specifically, using the JMA results, we evaluate the probabilistic property of the RMB misalignment estimate implied by a few hypothetical composites of study characteristics. The composites are defined by a) the average values of study characteristics, b) the most frequently used study characteristics, c) the least frequently used study characteristics, d) the study characteristics of the most cited English study, and e) the study characteristics of the most cited Chinese study. For all these hypothetical cases, the evidence of a misaligned RMB, in general, is weak. While the magnitude of RMB misalignment can be swayed by study characteristics, it is not easy to get a definite statistical evidence against the notion of RMB is not misaligned.

The inability to conclude statistically that the RMB is misaligned seems contradictory to the "common" wisdom of a "severely" undervalued RMB. Aside from the possibility that the RMB is fairly valued, what are the reasons that we do not obtain a sharp inference from the meta-analysis? One possible reason is the inherent difficulty of pinning down the correct level of misalignment. In the case of the RMB, Cheung, *et al*. (2007), Dunaway, *et al*. (2009), and Schnatz (2011), for example, point out that the lack of a consensual equilibrium exchange rate model, substantial data revisions, and sensitivity to alternative model assumptions can lead to diverse misalignment estimates. The very wide spread of estimates is likely due to alternative and different views on the equilibrium exchange rate.

The meta-analysis approach allows us to pool and aggregate information from these studies. However, if these misalignment estimates do not represent random (measurement) variations around the true value within and across individual studies, the pooling may not improve the precision of estimation. When empirical studies are based on different models with different implications for the equilibrium exchange rate, the pooling will not help in sharpening inferences about the degree of misalignment. Our results on the heterogeneity of RMB misalignment estimates, and the inability to draw a sharp inference are indicative of limited gain of information in aggregating these data. It is likely that these estimates are generated from different models and different approaches that imply different equilibrium RMB values, and hence different levels of misalignment. Overall, it is prudent to acknowledge the possibility that these empirical studies and their misalignment estimates are not necessarily very informative for assessing a precise level of RMB misalignment, and for recommending related policy remedies.

APPENDIX

A.1.   The Sample of Studies

ID	Study	Publication type	Language
1	Agya and Jun (2015)	Journal	English
2	Aflouk, Jeong, Mazier and Saadaoui (2010)	Journal	English
3	Almas, Grewal, Hvide and Ugurlu (2017)	Journal	English
4	Benassy-Quere and Lahreche-Revil (2008)	Journal	English
5	Benassy-Quere, Bereau and Mignon (2009)	Journal	English
6	Benassy-Quere, Lahreche-Revil and Mignon (2011)	Journal	English
7	Chang (2007)	Journal	English
8	Chang (2008)	Journal	English
9	Chang and Qin (2004)	Journal	English
10	Chen (2009)	Journal	English
11	Cheung, Chinn and Fujii (2007)	Journal	English
12	Cheung, Chinn and Fujii (2009)	Journal	English
13	Cheung, Chinn and Fujii (2010)	Journal	English
14	Cheung, Chinn and Fujii (2017)	Journal	English
15	Chou and Shih (1998)	Journal	English
16	Christoph and Hossfeld (2014)	Journal	English
17	Coudert and Couharde (2007)	Journal	English
18	Cui (2013)	Journal	English
19	Frankel (2006)	Journal	English
20	Funke and Rahn (2005)	Journal	English
21	Gan, Ward, Su and Cohen (2013)	Journal	English
22	Garroway, Hacibedel, Reisen and Turkisch (2012)	Journal	English
23	Giannellis and Koukouritakis (2018)	Journal	English
24	Hall, Kenjegaliev, Swamy and Tavlas (2013)	Journal	English
25	Hu and Chen (2010)	Journal	English
26	Lipman (2011)	Journal	English
27	Lü (2007)	Journal	English
28	Nouira, Plane and Sekkat (2011)	Journal	English
29	Peng, Lee and Gan (2008)	Journal	English
30	Schroder (2013)	Journal	English
31	Yang and Bajeux-Besnainou (2006)	Journal	English
32	Yi (2010)	Journal	English
33	You and Sarantis (2011)	Journal	English
34	You and Sarantis (2012a)	Journal	English
35	You and Sarantis (2012)	Journal	English
36	Zhang and Chen (2014)	Journal	English
37	Chen, Deng and Kemme (2008)	Working paper	English
38	Cline (2007)	Working paper	English
39	Cline (2008)	Working paper	English
40	Garton and Chang (2005)	Working paper	English
41	Jeong and Mazier (2003)	Working paper	English
42	Jeong, Bao and Mazier (2007)	Working paper	English
43	Li (2009)	Working paper	English
44	MacDonald and Dias (2007)	Working paper	English
45	Sinnakkannu and Vnair (2010)	Working paper	English
46	Zhang (2010)	Working paper	English
47	Zhang (2012b)	Working paper	English
48	Li (2015)	Master Thesis	English
49	Benassy-Quere, Duran-Vigneron, Lahreche-Revil and Mignon (2004)	Book Chapter	English
50	Cheung, Chinn and Fujii (2012)	Book Chapter	English

51	Cline (2013a)	IIE	English
52	Cline (2013b)	IIE	English
53	Cline (2014a)	IIE	English
54	Cline (2014b)	IIE	English
55	Cline and Williamson (2008)	IIE	English
56	Cline and Williamson (2009)	IIE	English
57	Cline and Williamson (2010a)	IIE	English
58	Cline and Williamson (2010b)	IIE	English
59	Cline and Williamson (2011)	IIE	English
60	Cline and Williamson (2012a)	IIE	English
61	Cline and Williamson (2012b)	IIE	English
62	Subramanian (2010)	IIE	English
63	Shi and Yu (2005)	Journal	Chinese
64	Sun and Sun (2013)	Journal	Chinese
65	Wang (2015)	Journal	Chinese
66	Wang and Cai (2007)	Journal	Chinese
67	Wang and Lin (2013)	Journal	Chinese
68	Wang and Yao (2008)	Journal	Chinese
69	Zhang (2000)	Journal	Chinese

Notes: IIE refers to Peterson Institute for International Economics.

A.2. Study Characteristic Types and Regime Dummy Variables

Study Characteristic Types	Description
a) Data Characteristics	
Annual	=1 if annual data are used.
Non-Annual	=1 if data are not annual.
PPP-based[1]	=1 if PPP-based data derived from ICP surveys are mainly used.
Non-PPP-based	=1 if market based data from, say, IFS, World Bank, or BIS are mainly used.
Dollar-based RER	=1 if bilateral real RMB-US dollar exchange rate is used.
Non-Dollar-based ER	=1 if bilateral real RMB-US dollar exchange rate is not used, including RMB real effective exchange rate, bilateral real RMB against Japanese yen or euro exchange rate or RMB nominal (effective) exchange rate.
b) Theoretical and Estimation Specifications	
BEER[2]	=1 if a model from the family of behavioral equilibrium exchange rate models or the productivity approach is used.
FEER[3]	=1 if the fundamental equilibrium exchange rate model, IMF macroeconomic balance approach or the nature rate of exchange approach is used.
Penn Effect	=1 if the Penn effect approach is used.
Other Approaches	=1 if other frameworks, such as the absolute or relative PPP framework or shadow price of foreign exchange approach is used
Time series	=1 if time series technique is used.
Non-Time-series	=1 if panel technique or cross-sectional technique is used.
Cointegration	=1 if cointegration framework is used.
Non-cointegration	=1 if non-cointegration framework is used.
Reduced-form	=1 if reduced-form setup is used.
Structural	=1 if structural setup is used.
c) Publication Attributes	
Journal	=1 if the study is published in a peer-reviewed journal.
Non-Journal	=1 if the study is not published in a peer-reviewed journal, e.g. book chapters and working papers.
Mainland	=1 if any one of the authors of the study is affiliated with a mainland China institution.
Non-mainland	=1 if no author of the study is affiliated with a mainland China institution.
Academics	=1 if all authors of the study are affiliated with academic affiliations (e.g. university).
Non-Academics	=1 if any one of authors are not affiliated with academic affiliations. The authors may be affiliated with government (e.g. central bank), think tanks (e.g. IIE), international organizations (e.g. IMF, Asian Development Bank) or industries.
Regime Dummy Variables	
1994-1997	=1 when the RMB misalignment estimate falls within the period of 1994 to 1997.
1998-2004	=1 when the RMB misalignment estimate falls within the period of 1998 to 2004.
2005-2008	=1 when the RMB misalignment estimate falls within the period of 2005 to 2008.

2009-2010		=1 when the RMB misalignment estimate falls within the period of 2009 to 2010.	
2011-2014		=1 when the RMB misalignment estimate falls within the period of 2011 to 2014.	

Notes: The variables marked with bold-font are benchmark characteristic types. ① Here we do not distinguish the different rounds of ICP data because of data complexity. However, note that ICP revision has a pronounced implication for estimating currency misalignment (Cheung and Fujii, 2014); ② "BEER" refers to the family of behavioral equilibrium exchange rate models, which consists of the standard behavioral equilibrium exchange rate model (BEER), the permanent equilibrium exchange rate model (PEER), the equilibrium real exchange rate model (ERER), the Goldman Sachs dynamic equilibrium exchange rate (GSDEER). Besides, the productivity approach solely considers the productivity differentials as exchange rate determinant, which can be incorporated into BEER specification with other variables, such as interest differentials (Cheung, *et al.*, 2010). Therefore, the productivity approach is also included in BEERs; ③ "FEER" refers to fundamental equilibrium exchange rate model (FEER), IMF macroeconomic balance approach (MB) and the nature rate of exchange approach (NATREX), which are quite similar theoretically.

A.3. Descriptive Statistics of RMB Misalignment Estimates: Different Exchange Rate Policy Regimes

	Obs	Mean	SE	Rank-Sum
Sample Periods				
1994-1997	166	9.7%	17.1%	0.36
1998-2004	402	10.1%	22.4%	0.00
2005-2008	222	19.0%	20.5%	0.00
2009-2010	55	12.2%	15.0%	0.41
2011-2014	80	-0.7%	22.2%	0.00

Notes: See Notes to Table 2.

A.4. Jackknife model averaging and Other Selected Model Averaging Procedures

The general form of model averaging is introduced here. The notations are slightly different from those in text.

A.4.1 Model Setup

Consider the linear regression model:

$$y = X\theta + \varepsilon$$

where $y = (y_1,...,y_T)'$ is a $T \times 1$ vector of the dependent variable and ε is a $T \times 1$ vector of random residual terms. The $T \times K$ matrix $X = (X_1, X_2,..., X_K)$ contains the K potential exogenous explanatory variables, and $X_j, j = 1, 2,..., K$ is a $T \times 1$ vector of the j-th explanatory variable. The coefficients of these K variables are in the $K \times 1$ θ vector.

Assume $E(\varepsilon|X) = 0$ and $Var(\varepsilon|X) = \Omega$, where Ω is a $T \times T$ positive definite symmetric matrix that allows for heteroskedastic and/or autocorrelated residuals . Under these assumptions, $X\hat{\theta}$ is an unbiased estimator of y if $\hat{\theta}$ is unbiased.

The K potential explanatory variables offer 2^K potential model specifications. Assume all these model specifications are under consideration. The size of the model space is 2^K. Let X_τ contains variables of the τ-th model, and the τ-th model is represented by $y = X_\tau \theta_\tau + \varepsilon_\tau$. Note that the OLS estimator; $\hat{\theta}_\tau^{OLS} = (X_\tau' X_\tau)^{-1} X_\tau' y$ is unbiased under heteroskedasticity and autocorrelation. Let $\hat{\varepsilon}_\tau = y - X_\tau (X_\tau' X_\tau)^{-1} X_\tau' y$ be the corresponding OLS residual vector.

A.4.2 Jackknife model averaging

Jackknife model averaging (JMA) is based on the Cross-Validation Criterion applied to Jackknife residuals. In this exercise, we consider the so-called delete-1 Jackknife residual. The delete-1 Jackknife residual of a specific observation, say, the *i*-th observation is the error obtained

from predicting the i-th observation with the coefficient vector estimated from the sample with the i-th observation deleted. Schematically, 1) estimate the model with the sample without the i-th observation, 2) predict the i-th observation based on the estimates from 1) and the corresponding i-th right-hand-side variable(s), 3) the Jackknife residual is the difference of the actual and predicted values.

A convenient way of calculating Jackknife residual vectors is by adjusting OLS residual vectors. For the τ-th model in the model space, the Jackknife residual is $\hat{\varepsilon}_\tau^{Jack} = Diag_\tau \hat{\varepsilon}_\tau$, where $Diag_\tau$ is a $T \times T$ diagonal matrix with i-th (i=1,2,...,T) diagonal element equaling to $\left(1 - x'_{i,\tau}\left(X'_\tau X_\tau\right)^{-1} x_{i,\tau}\right)^{-1}$, and $x_{i,\tau}$ is the i-th explanatory observation in model M_τ. After we estimating all the models, we have a $T \times 2^K$ Jackknife residual matrix which is $\hat{\boldsymbol{\varepsilon}}^{Jack} = \left(\hat{\varepsilon}_1^{Jack}, \hat{\varepsilon}_2^{Jack}, ..., \hat{\varepsilon}_{2^K}^{Jack}\right)$.

Given these notations, JMA chooses a $2^K \times 1$ weight vector $\hat{w}$ that minimizes the least-squares cross-validation criterion, that is,

$$\hat{w} = \underset{w \in \{w \le 1, w \ge 0, w'1 = 1\}}{\arg\min} w' \hat{\boldsymbol{\varepsilon}}'^{Jack} \hat{\boldsymbol{\varepsilon}}^{Jack} w .$$

Given $\hat{w}$, the weighted estimator of the j-th (i=1,2,...,K) coefficient θ_j is given by

$$\hat{\theta}_j = \sum_{\tau=1}^{2^K} \hat{w}_\tau \hat{\theta}_{j,\tau}^{OLS} I\left(\hat{\theta}_{j,\tau}\right) .$$ The weighted standard error of $\hat{\theta}_j$ is

$$SE\left(\hat{\theta}_j\right) = \sum_{\tau=1}^{2^K} \hat{w}_\tau \left[Var\left(\hat{\theta}_{j,\tau}^{OLS}\right) + \left(\hat{\theta}_{j,\tau}^{OLS} - \hat{\theta}_j\right)^2\right]^{1/2} I\left(\hat{\theta}_{j,\tau}\right),$$ and the inclusion weight of $\hat{\theta}_j$ is $IW\left(\hat{\theta}_j\right) = \sum_{\tau=1}^{2^K} \hat{w}_\tau I\left(\hat{\theta}_{j,\tau}\right)$,

where $I\left(\hat{\theta}_{j,\tau}\right)$ is an indicator function depending on whether θ_j is included in the τ-th model.

The cross-validation is based on the prediction residual error sum of squares (Syed, 2011). JMA chooses the cross-validation criterion as the target function. The model weights calculated by JMA minimize the prediction residual error sum of squares of the combined model. Hansen and Racine (2012) show that the combined model achieves the lowest possible expected squared error even in the presence of heteroskedasticity. The combined model is considered the optimal model in the model space under consideration.

When assigning a model weight to each model specification in the model space, the covariance of residuals across model specifications is considered. The model weight measures the contribution of a model specification to the optimal model combination. The model with the largest model weight contributes the most in forming the optimal model combination. The combined model is optimal in sense of having the lowest possible expected squared error of the model. Therefore, the more a model specification contributes to the optimal model combination, the better the model is in sense of having a lower possible expected squared error.

A.4.3 *Model weights based on other kinds of Frequentist Model averaging*
(1)   Heteroskedasticity-robust model averaging based on the Mallows criterion

The model averaging procedure based on the Mallows criterion is robust to heteroskedasticity represented by the presence of non-homogenous diagonal elements in Ω, the variance-covariance matrix of residuals.

The heteroskedasticity-robust Mallows model averaging (MMA) uses OLS residual estimates. After estimating all the models in model space, we have a $T \times 2^K$ residual matrix $\hat{\boldsymbol{\varepsilon}}^{OLS} = \left(\hat{\varepsilon}_1, \hat{\varepsilon}_2, ..., \hat{\varepsilon}_{2^K}\right)$. Model weights are obtained from minimizing the heteroskedasticity-robust Mallows criterion:

$$\hat{w}^{Mallows} = \underset{w \in \{w \le 1, w \ge 0, w'1=1\}}{\arg\min} \quad w' \hat{\varepsilon}'^{OLS} \hat{\varepsilon}^{OLS} w + 2tr(\Omega)' Pw,$$

where $P = \left(tr\left(X_1\left(X_1'X_1\right)^{-1}X_1'\right)' \quad tr\left(X_2\left(X_2'X_2\right)^{-1}X_2'\right)' \quad \cdots \quad tr\left(X_{2^K}\left(X_{2^K}'X_{2^K}\right)^{-1}X_{2^K}'\right)' \right)$ is a $T \times 2^K$ matrix.

In early studies (Gorman and Toman, 1966), the Mallows criterion was used to compare the sum of squared bias errors in different models. Here, the heteroskedasticity-robust MMA selects a set of model weights to minimize the sum of squared bias errors after combining the models. Similar to JMA, the heteroskedasticity-robust MMA considers the covariances of residuals across model specifications.

(2)   Model averaging based on Information Criteria
     Model averaging based on the information criterion approach assumes the residuals follow an identical and independent normal distribution. AIC and BIC are two information criteria that are widely used in model selection. Following the same model setup described above, we calculate the AIC or BIC for the τ-th model and denote it as IC_τ. The weight for the τ-th model is:

$$w_\tau^{IC} = \frac{e^{-IC_\tau/2}}{\sum_{j=1}^{2^K} e^{-IC_j/2}}.$$

If a model specification has a high "quality information content" in terms of likelihood and parsimony as measured by an information criterion, it has a relatively high model weight. The model specification with the highest model weight is the best model specification in the model space in terms of the information criterion.

A.4.4   Bayesian Model averaging
     Bayesian model averaging is based on Bayes' Theorem. In BMA, there are two parts: the estimation of regression coefficients and the estimation of model probability. Under the setup introduced in A.4.1, BMA assumes the residuals are IID normal.
     When estimating regression coefficients, a widely used assumption is the Zellner's *g* prior which specifies how certain the researcher is about the coefficients are zero. A large g means the researcher is very uncertain that the coefficients equal 0, and a small g means the researcher is very certain that the coefficients equal 0. A popular choice of g is $g = T$ which is called the unit information prior. Under the Zellner's g prior, the posterior Bayesian (vector) estimator of coefficients in the τ-th model is:

$$\hat{\theta}_\tau^{bayes} = \frac{g}{1+g}\hat{\theta}_\tau^{OLS} = \frac{g}{1+g}\left(X_\tau'X_\tau\right)^{-1}X_\tau'y.$$

The posterior model probability of the τ-th model is:

$$p(M_\tau | y, X_\tau) = \frac{p(y|M_\tau, X_\tau)p(M_\tau)}{\sum_{i=1}^{2^K} p(y|M_i, X_i)p(M_i)},$$

where $p(y|M_\tau, X)$ is the marginal likelihood[28], and $p(M_\tau)$ is the prior model probability. There are different ways to choose the prior model probability. One approach is to choose prior variable probability and then calculate prior model probability. Suppose the inclusion of a representative

---

[28] The marginal likelihood is also called integrated likelihood. The marginalized variables are parameters in the model, which means $p(y|M_\tau, X) = \int p(y|M_\tau, X, \beta) p(\beta|M_\tau, X) d\beta$.

variable X_j follows Bernoulli distribution and the prior inclusion probability is p_j, the prior model probability is:

$$p(M_\tau) = \prod_{j \in \boldsymbol{\tau}} p_j \prod_{l \notin \boldsymbol{\tau}} (1 - p_l),$$

where the bold $\boldsymbol{\tau}$ is a set of index indicating which variable is included in the τ-th model. A commonly and conservatively used choice of priors, the uniform model prior, is $p_j = 0.5$, $j = 1,2,...,K$.

The expectation of coefficient θ_j is:

$$E(\theta_j \mid \boldsymbol{y}, \boldsymbol{X}) = \sum_{\tau=1}^{2^K} p(M_\tau \mid \boldsymbol{y}, \boldsymbol{X}_\tau) \hat{\theta}_{j,\tau}^{bayes}.$$

The expected standard deviation of coefficient θ_j is:

$$SD(\theta_j) = \sqrt{\sum_{\tau=1}^{2^K} \left\{ \left[Var\left(\hat{\theta}_{j,\tau}^{bayes}\right) + \left(\hat{\theta}_{j,\tau}^{bayes}\right)^2 \right] p(M_\tau \mid \boldsymbol{y}, \boldsymbol{X}_\tau) \right\} - E(\theta_j \mid \boldsymbol{y}, \boldsymbol{X})^2}.$$

And, the posterior inclusion probability of coefficient is:

$$PIP(\theta_j \mid \boldsymbol{y}, \boldsymbol{X}) = \sum_{\tau=1}^{2^K} p(M_\tau \mid \boldsymbol{y}, \boldsymbol{X}_\tau) \boldsymbol{I}\left(\hat{\theta}_{j,\tau}^{bayes}\right).$$

A.5. The Model Weights from the Jackknife Model Averaging Estimation

A.5.1 Non-benchmark Types as Regressors
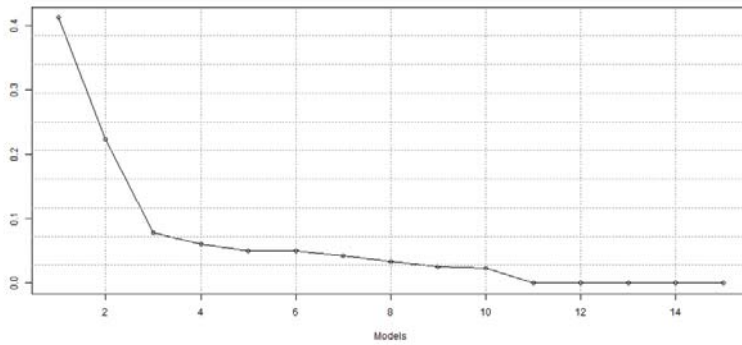
A.5.1 Benchmark Types as Regressors
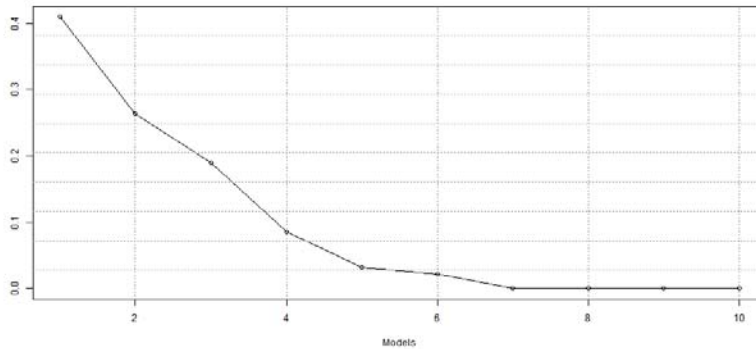
A.6. Comparison of Results

	Cheung and He (forthcoming)		Bineau (2010)		Korhonen and Ritola (2011)	
	Significance	Sign	Significance	Sign	Significance	Sign
Annual	insignificant	+	insignificant	+		
Non-Annual	insignificant	+/-	insignificant	-	insignificant	
PPP-based	significant	+				
Non-PPP-based	insignificant	-				
Dollar-based RER	insignificant	+	significant	-	significant	+
Non-Dollar-based ER	insignificant/significant	+/-	significant	+	insignificant	
BEER	significant	-	insignificant	+	insignificant	
FEER	insignificant	+	significant	+		
Penn Effect	insignificant	+	--	-		
Other Approaches	insignificant	+/-	insignificant/significant	+/-		
Time series	significant	-	--	-	significant	-
Non-Time-series	insignificant/significant	+	significant	+	insignificant	
Cointegration	insignificant	+				
Non-cointegration	insignificant	-				
Reduced-form	significant	-				
Structural	significant	+				
Journal	insignificant	-	significant	-	significant	+
Non-Journal	insignificant	+	significant	+	insignificant	
Mainland	insignificant	-			insignificant	
Non-mainland	insignificant	+				
Academics	insignificant	-				
Non-Academics	insignificant/significant	+/-			significant	-
1994-1997	insignificant	+				
1998-2004	insignificant	+				
2005-2008	insignificant	+				
2009-2010	significant	-				
2011-2014	significant	-				

Notes: In the columns labelled "Significance", "insignificant/significant" means at least two elements are in the corresponding type, and the significances are opposite. In the columns labelled "Sign", "+" means the positive sign, "-" means negative sign; "+/-" means at least two elements are in the corresponding type, and the signs are opposite.

A.7. The Mean and the Variance of Weighted Estimators

Following Buckland, *et al.* (1997), the weighted mean of y_i is given by

$$\hat{y}_i = \sum_\tau \hat{w}_\tau \hat{y}_{i,\tau} = \sum_\tau \hat{w}_\tau \mathbf{x}'_{i,\tau} \boldsymbol{\beta}_\tau ,$$

where $\hat{w}_\tau$ is the model weight of the τ-th model, $\hat{y}_{i,\tau}$ is the predict of y_i by using the τ-th model, $\mathbf{x}_{i,\tau}$ the data used for predicting y_i by the τ-th model, $\boldsymbol{\beta}_\tau$ is the estimates of the coefficients.
The variance of y_i is given by

$$Var\left(y_i | \mathbf{x}_i\right) = Var\left(\hat{y}_i + \hat{\varepsilon}_i | \mathbf{x}_i\right)$$
$$= Var\left(\hat{y}_i | \mathbf{x}_i\right) + Var\left(\hat{\varepsilon}_i | \mathbf{x}_i\right)$$

where

$$Var\left(\hat{y}_i | \mathbf{x}_i\right) = Var\left(\sum_\tau \hat{w}_\tau \hat{y}_{i,\tau} | \mathbf{x}_i\right)$$
$$= \sum_\tau \hat{w}_\tau^2 Var\left(\hat{y}_{i,\tau} | \mathbf{x}_i\right) + \sum_\tau \sum_{\kappa \neq \tau} \hat{w}_\tau \hat{w}_\kappa \, cov\left(\hat{y}_{i,\tau}, \hat{y}_{i,\kappa} | \mathbf{x}_i\right).$$

Assuming perfect correlation, which means $cov\left(\hat{y}_{i,\tau}, \hat{y}_{i,\kappa} | \mathbf{x}_i\right) \approx Var\left(\hat{y}_{i,\tau} | \mathbf{x}_i\right) \approx Var\left(\hat{y}_{i,\kappa} | \mathbf{x}_i\right)$, the above equation becomes:

$$Var\left(\hat{y}_i | \mathbf{x}_i\right) = \left[\sum_\tau \hat{w}_\tau \sqrt{Var\left(\hat{y}_{i,\tau} | \mathbf{x}_i\right)}\right]^2 .$$

Besides,

$$E\left[\left(\hat{y}_{i,\tau} - E\left(y_i | \mathbf{x}_i\right)\right)^2\right]$$
$$= E\left[\left(\hat{y}_{i,\tau} - E\left(y_{i,\tau} | \mathbf{x}_{i,\tau}\right)\right)^2 + \left(E\left(y_{i,\tau} | \mathbf{x}_{i,\tau}\right) - E\left(y_i | \mathbf{x}_i\right)\right)^2\right]$$
$$= Var\left(\hat{y}_{i,\tau} | \mathbf{x}_{i,\tau}\right) + \left(y_{i,\tau} - y_i\right)^2 ,$$

and,

$$Var\left(\hat{y}_{i,\tau} | \mathbf{x}_{i,\tau}\right)$$
$$= Var\left(\mathbf{x}'_{i,\tau} \boldsymbol{\beta}_\tau\right)$$
$$= \mathbf{x}'_{i,\tau} Var\left(\boldsymbol{\beta}_\tau\right) \mathbf{x}_{i,\tau} .$$

A.8.   Model Averaging Results Derived from the Mallows Criterion, AIC Method, BIC Method, and the Bayesian Model Averaging Approach: Benchmark study characteristic types as regressions.

	R MMA			AIC			BIC			BMA		
	IW	MA	BEST	IW	MA	BEST	IW	MA	BEST	PIP	Post Est.	BEST
Weights			0.4226			0.6071			0.5121			0.5110
Intercept		0.3462	**0.4112**		0.3782	**0.3744**		0.3834	**0.4112**		0.3830	0.4109
		(0.0755)	(0.0399)		(0.0477)	(0.0492)		(0.0535)	(0.0405)			
Data Characteristics												
Annual	0.4765	0.0418		0.8258	0.0408	**0.0515**	0.2394	0.0119		0.2388	0.0119	
		(0.0238)			(0.0196)	(0.0218)		(0.0104)			(0.0244)	
Non-PPP-based	0.8188	-0.1018	**-0.0903**	0.9999	-0.1048	**-0.1110**	0.9958	-0.0935	**-0.0903**	0.9957	-0.0934	**-0.0902**
		(0.0349)	(0.0188)		(0.0258)	(0.0255)		(0.0242)	(0.0191)		(0.0264)	(0.0178)
Dollar-based RER	0.9414	0.065	**0.0658**	0.9999	0.0690	**0.0696**	0.9989	0.0688	**0.0658**	0.9988	0.0687	**0.0657**
		(0.0161)	(0.0163)		(0.0150)	(0.0166)		(0.0154)	(0.0165)		(0.0156)	(0.0148)
Theoretical and Estimation Specifications												
BEER	0.6248	-0.1047	**-0.1724**	0.9747	-0.1248	**-0.1122**	0.8890	-0.1409	**-0.1724**	0.8879	-0.1406	**-0.1722**
		(0.0428)	(0.0274)		(0.0423)	(0.0368)		(0.0356)	(0.0279)		(0.0608)	(0.0245)
Time-series	0.7423	-0.0904	**-0.1208**	1.0000	-0.1023	**-0.0967**	1.0000	-0.1157	**-0.1208**	1.0000	-0.1156	**-0.1206**
		(0.0279)	(0.0195)		(0.0202)	(0.0238)		(0.0193)	(0.0197)		(0.0196)	(0.0168)
Non-Cointegration	1.0000	-0.151	**-0.1859**	1.0000	-0.1764	**-0.1749**	0.9999	-0.1703	**-0.1859**	0.9999	-0.17	**-0.1857**
		(0.0508)	(0.0326)		(0.0329)	(0.0338)		(0.0376)	(0.0331)		(0.0408)	(0.0267)
Reduced-form	0.3752	-0.0493		0.7547	-0.0438	**-0.0575**	0.2673	-0.0210		0.2678	-0.0211	
		(0.0316)			(0.0240)	(0.0250)		(0.0171)			(0.0402)	
Publication Attributes												
Journal	0.7812	-0.0555	**-0.0794**	0.9897	-0.0641	**-0.0610**	0.9578	-0.0713	**-0.0794**	0.9572	-0.0712	**-0.0793**
		(0.0180)	(0.0148)		(0.0195)	(0.0156)		(0.0185)	(0.0150)		(0.0240)	(0.0173)
Non-mainland	0.9624	0.1006	**0.1085**	1.0000	0.1057	**0.1049**	1.0000	0.1057	**0.1085**	1.0000	0.1056	**0.1083**
		(0.0160)	(0.0154)		(0.0160)	(0.0162)		(0.0164)	(0.0156)		(0.0163)	(0.0156)
Academics	0.9201	-0.0388	**-0.0447**	0.9698	-0.0437	**-0.0449**	0.7449	-0.0333	**-0.0447**	0.7424	-0.0332	**-0.0447**
		(0.0132)	(0.0132)		(0.0147)	(0.0140)		(0.0139)	(0.0133)		(0.0233)	(0.0140)
Time Dummies												
1998-2004		-0.0058	-0.0113		-0.0062	-0.0048		-0.0093	-0.0113		-0.0093	-0.0113
		(0.0181)	(0.0169)		(0.0172)	(0.0171)		(0.0172)	(0.0171)		(0.0171)	(0.0168)
2005-2008		0.0128	0.0036		0.0092	0.0104		0.0074	0.0036		0.0075	0.0036
		(0.0229)	(0.0206)		(0.0204)	(0.0208)		(0.0208)	(0.0208)		(0.0207)	(0.0200)
2009-2010		-0.0838	**-0.0974**		-0.0883	**-0.0857**		-0.0932	**-0.0974**		-0.0931	**-0.0973**
		(0.0318)	(0.0284)		(0.0298)	(0.0288)		(0.0302)	(0.0288)		(0.0300)	(0.0291)
2011-2014		-0.2467	**-0.2566**		-0.2526	**-0.2515**		-0.2547	**-0.2566**		-0.2545	**-0.2563**
		(0.0324)	(0.0305)		(0.0266)	(0.0308)		(0.0268)	(0.0309)		(0.0266)	(0.0264)
Adj.R2			0.2965			0.3012			0.2965			

Notes: See the notes to Table 5.

References

Amini, Shahram M., and Christopher F. Parmeter. "Comparison of model averaging techniques: Assessing growth determinants." *Journal of Applied Econometrics* 27.5 (2012): 870-876.

Balassa, Bela. "The purchasing-power parity doctrine: a reappraisal." *Journal of Political Economy* 72.6 (1964): 584-596.

Barbieri, Maria Maddalena, and James O. Berger. "Optimal predictive model selection." *The Annals of Statistics* 32.3 (2004): 870-897.

Bates, John M., and Clive WJ Granger. "The combination of forecasts." *Journal of the Operational Research Society* 20.4 (1969): 451-468.

Bénassy-Quéré, Agnès, Sophie Béreau, and Valérie Mignon. "Robust estimations of equilibrium exchange rates within the G20: A panel BEER approach." *Scottish Journal of Political Economy* 56.5 (2009): 608-633.

Bénassy-Quéré, Agnès, Amina Lahrèche-Révil, and Valérie Mignon. "Is Asia responsible for exchange rate misalignments within the G20?" *Pacific Economic Review* 13.1 (2008): 46-61.

Benos, Nikos, and Stefania Zotou. "Education and economic growth: A meta-regression analysis." *World Development* 64 (2014): 669-689.

Bergin, Paul R., Reuven Glick, and Alan M. Taylor. "Productivity, tradability, and the long-run price puzzle." *Journal of Monetary Economics* 53.8 (2006): 2041-2066.

Bergsten, C. Fred. "The Chinese exchange rate and the US economy." *Testimony before the Senate Committee on Banking, Housing and Urban Affairs, January* 31 (2007).

Bineau, Yannick. "Renminbi's misalignment: A meta-analysis." *Economic Systems* 34.3 (2010): 259-269.

Buckland, Steven T., Kenneth P. Burnham, and Nicole H. Augustin. "Model selection: an integral part of inference." *Biometrics* (1997): 603-618.

Burnham, Kenneth P., and David R. Anderson. "A practical information-theoretic approach." *Model Selection and Multimodel Inference, 2nd ed. Springer, New York* (2002).

Capon, Noel, John U. Farley, and Scott Hoenig. "Determinants of financial performance: a meta-analysis." *Management Science* 36.10 (1990): 1143-1159.

Campos, Nauro F., Jarko Fidrmuc, and Iikka Korhonen. "Business cycle synchronisation and currency unions: A review of the econometric evidence using meta-analysis." *International Review of Financial Analysis* 61 (2019): 274-283.

Cenedese, Gino, and Thomas Stolper. "Currency fair value models." *Handbook of Exchange Rates* (2012): 313-342.

Chang, Gene Hsin. "Estimation of the Undervaluation of the Chinese Currency by a Non-linear Model." Asia-Pacific Journal of Accounting & Economics 15.1 (2008): 29-40.

Chen-Yuan, Tung, and Sam Baker. "RMB revaluation will serve China's self-interest." *China Economic Review* 15.3 (2004): 331-335.

Cheung, Yin-Wong. "Exchange Rate Misalignment - The Case of the Chinese Renminbi (Chapter 27)." *Handbook of Exchange Rates*, edited by Jessica James, Ian W. Marsh and Lucio Sarno, John Wiley & Sons, Inc. (2012) 751-765.

Cheung, Yin-Wong, Menzie D. Chinn, and Eiji Fujii. "The overvaluation of renminbi undervaluation." *Journal of International Money and Finance* 26.5 (2007): 762-785.

Cheung, Yin-Wong, Menzie D. Chinn, and Eiji Fujii. "Measuring misalignment: Latest estimates for the Chinese yuan." *The US-Sino Currency Dispute: New Insights from Economics, Politics and Law, VoxEU, April* (2010).

Cheung, Yin-Wong, Menzie D. Chinn, Antonio Garcia Pascual, and Yi Zhang. "Exchange rate prediction redux: new models, new data, new currencies." Journal of International Money and Finance 95 (2019): 332-362.

Cheung, Yin-Wong, Menzie Chinn, and Xin Nong. "Estimating currency misalignment using the Penn effect: It is not as simple as it looks." *International Finance* 20.3 (2017): 222-242.

Cheung, Yin‑Wong, and Eiji Fujii. "Exchange rate misalignment estimates — sources of differences." International Journal of Finance & Economics 19.2 (2014): 91-121.

Cheung, Yin-Wong, and Shi He. "Truths and myths about RMB misalignment: A meta-analysis." *Comparative Economic Studies* (forthcoming), https://doi.org/10.1057/s41294-019-00093-0.

Chinn, Menzie D. "Before the fall: were East Asian currencies overvalued?" *Emerging Markets Review* 1.2 (2000): 101-126.

Cline, William R., and John Williamson. Estimates of Fundamental Equilibrium Exchange Rates, May 2010. PB10-15. *Peterson Institute for International Economics*, (2010).

Coggin, T. Daniel, and John E. Hunter. "A meta-analysis of mutual fund performance." *Review of Quantitative Finance and Accounting* 3.2 (1993): 189.

Ćorić, Bruno, and Geoff Pugh. "The effects of exchange rate variability on international trade: a meta-regression analysis." *Applied Economics* 42.20 (2010): 2631-2644.

Corsetti, Giancarlo, Paolo Pesenti, and Nouriel Roubini. "What caused the Asian currency and financial crisis?" *Japan and the World Economy* 11.3 (1999): 305-373.

Coudert, Virginie, and Cecile Couharde. "Real equilibrium exchange rate in China is the renminbi undervalued?" *Journal of Asian Economics* 18.4 (2007): 568-594.

Doucouliagos, Hristos, and Mehmet Ali Ulubaşoğlu. "Democracy and economic growth: a meta-analysis." *American Journal of Political Science* 52.1 (2008): 61-83.

Dunaway, Steven, Lamin Leigh, and Xiangming Li. "How robust are estimates of equilibrium real exchange rates: The case of China." *Pacific Economic Review* 14.3 (2009): 361-375.

Égert, Balázs, and László Halpern. "Equilibrium exchange rates in Central and Eastern Europe: A meta-regression analysis." *Journal of Banking & Finance* 30.5 (2006): 1359-1374.

Eicher, Theo S., Lindy Helfman, and Alex Lenkoski. "Robust FDI determinants: Bayesian model averaging in the presence of selection bias." *Journal of Macroeconomics* 34.3 (2012): 637-651.

Engel, Charles. "Exchange Rate Policies: A Federal Reserve Bank of Dallas Staff Paper." *DIANE Publishing* (2010).

Engel, Charles, Dohyeon Lee, Chang Liu, Chenxin Liu, and Steve Pak Yeung Wu. "The uncovered interest parity puzzle, exchange rate forecasting, and Taylor rules." *Journal of International Money and Finance* 95 (2019): 317-331.

Fidrmuc, Jarko, and Ronja Lind. "Macroeconomic impact of Basel III: Evidence from a meta-analysis." *Journal of Banking & Finance* (2018).

Fragoso, Tiago M., Wesley Bertoli, and Francisco Louzada. "Bayesian model averaging: A systematic review and conceptual classification." *International Statistical Review* 86, 1 (2018): 1-28.

Frankel, Jeffrey. "On the yuan: The choice between adjustment under a fixed exchange rate and adjustment under a flexible rate." *CESifo Economic Studies* 52, 2 (2006): 246-275.

Gan, Christopher, Bert Ward, Su Ting Ting, and David A. Cohen. "An empirical analysis of China's equilibrium exchange rate: A co-integration approach." *Journal of Asian Economics* 29 (2013): 33-44.

Gorman, John W., and R. J. Toman. "Selection of variables for fitting equations to data." *Technometrics* 8, 1 (1966): 27-51.

Government Accountability Office. "International Trade: Treasury Assessments Have Not Found Currency Manipulation, but Concerns about Exchange Rates Continue", *Report to Congressional Committees* (2005) GAO-05-351 (Washington, D.C.: Government Accountability Office, April).

Granger, Clive WJ, and Ramu Ramanathan. "Improved methods of combining forecasts." *Journal of Forecasting* 3, 2 (1984): 197-204.

Groot, Wim, and Henriette Maassen Van Den Brink. "Overeducation in the labor market: a meta-analysis." *Economics of Education Review* 19.2 (2000): 149-158.

Hall, Stephen G., Amangeldi Kenjegaliev, P. A. V. B. Swamy, and George S. Tavlas. "Measuring currency pressures: The cases of the Japanese yen, the Chinese yuan, and the UK pound." *Journal of the Japanese and International Economies* 29 (2013): 1-20.

Hansen, Bruce E. "Least squares model averaging." *Econometrica* 75, 4 (2007): 1175-1189.

Hansen, Bruce E., and Jeffrey S. Racine. "Jackknife model averaging." *Journal of Econometrics* 167.1 (2012): 38-46.

Havranek, Tomas, Zuzana Irsova, Karel Janda, and David Zilberman. "Selective reporting and the social cost of carbon." *Energy Economics* 51 (2015): 394-406.

Havranek, Tomas, Roman Horvath, Zuzana Irsova, and Marek Rusnak. "Cross-country heterogeneity in intertemporal substitution." *Journal of International Economics* 96, 1 (2015): 100-118.

Havranek, Tomas, Zuzana Irsova, and Olesia Zeynalova. "Tuition Fees and University Enrolment: A Meta-Regression Analysis." *Oxford Bulletin of Economics and Statistics* 80, 6 (2018): 1145-1184.

Havranek, Tomas, Marek Rusnak, and Anna Sokolova. "Habit formation in consumption: A meta-analysis." *European Economic Review* 95 (2017): 142-167.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky. "Bayesian model averaging: a tutorial." *Statistical Science* (1999): 382-401.

Hu, Chuntian, and Zhijun Chen. "Renminbi already overappreciated: Evidence from FEERs (1994-2008)." *China Economist* (2010), 26, 64-78

International Monetary Fund Communications Department. "IMF Staff Completes the 2015 Article IV Consultation Mission to China." *Press Release No. 15/237* (2015). Available at: http://www.imf.org/external/np/sec/pr/2015/ pr15237.htm.

Iršová, Zuzana, and Tomáš Havránek. "Determinants of horizontal spillovers from FDI: Evidence from a large meta-analysis." *World Development* 42 (2013): 1-15.c

Kass, Robert E., and Adrian E. Raftery. "Bayes factors." *Journal of the American Statistical Association* 90.430 (1995): 773-795.

Korhonen, Iikka, and Maria Ritola. "Renminbi misaligned—results from meta-regressions." *Asia and China in the Global Economy* (2011). 97-121.

Kravis, Irving B., and Robert E. Lipsey. "Toward an explanation of national price level (Princeton Studies in International Finance No. 52)." *Princeton University, Princeton, NJ* (1983).

Kravis, Irving B., and Lipsey, Robert E.. The Assessment of National Price Levels," in Sven W. Arndt and J. David Richardson, eds., *Real Financial Linkages among Open Economies*, (1987) Cambridge, MA: MIT Press, 97-134.

Kravis, Irving B., Alan Heston, Robert Summers, Alicia Civitello, and S. P. Dhar. "International comparisons of real product and purchasing power-United Nations International Comparison Project: phase II." (1978).

Li, Ker-Chau. "Asymptotic optimality for Cp, CL, cross-validation and generalized cross-validation: Discrete index set." *The Annals of Statistics* 15, 3 (1987): 958-975.

Liu, Qingfeng, and Ryo Okui. "Heteroscedasticity-robust Cp model averaging." *The Econometrics Journal* 16, 3 (2013): 463-472.

Magnus, Jan R., and Giuseppe De Luca. "Weighted-average least squares (WALS): a survey." *Journal of Economic Surveys* 30, 1 (2016): 117-148.

Magnus, Jan R., Owen Powell, and Patricia Prüfer. "A comparison of two model averaging techniques with an application to growth empirics." *Journal of Econometrics* 154, 2 (2010): 139-153.

Magnus, Jan R., Alan TK Wan, and Xinyu Zhang. "Weighted average least squares estimation with nonspherical disturbances and an application to the Hong Kong housing market." *Computational Statistics & Data Analysis* 55.3 (2011): 1331-1341.

Meese, Richard A., and Kenneth Rogoff. "Empirical exchange rate models of the seventies: Do they fit out of sample?" *Journal of International Economics* 14.1-2 (1983): 3-24.

Meyer, Klaus E., and Evis Sinani. "When and where does foreign direct investment generate positive spillovers? A meta-analysis." *Journal of International Business Studies* 40.7 (2009): 1075-1094.

Moral-Benito, Enrique. "Model averaging in economics: An overview." *Journal of Economic Surveys* 29.1 (2015): 46-75.

Morrison, Wayne M., and Marc Labonte. "China's currency policy: An analysis of the economic issues." *Current Politics and Economics of Northern and Western Asia* 22.4 (2013): 507.

Rossi, Barbara. "Exchange rate predictability." *Journal of Economic Literature* 51.4 (2013): 1063-1119.

Sala-i-Martin, Xavier, Gernot Doppelhofer, and Ronald I. Miller. "Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach." *American Economic Review* (2004): 813-835.

Samuelson, Paul A. "Theoretical notes on trade problems." *The Review of Economics and Statistics* (1964): 145-154.

Samuelson, Paul A. "Facets of Balassa-Samuelson thirty years later." *Review of International Economics* 2.3 (1994): 201-226.

Schnatz, Bernd. "Global imbalances and the pretence of knowing fundamental equilibrium exchange rates." *Pacific Economic Review* 16.5 (2011): 604-615.

Shi, Jianhuai, and Yu, Haifeng. "Renminbi equilibrium exchange rate and China's exchange rate misalignment: 1991–2004." *Economic Research Journal* 40.4 (2005): 34-45.

Stanley, Tom D. "Wheat from chaff: Meta-analysis as quantitative literature review." *Journal of Economic Perspectives* 15.3 (2001): 131-150.

Steel, Mark FJ. "Model Averaging and Its Use in Economics." *Journal of Economic Literature* (forthcoming).

Syed, Ali R. "A review of cross validation and adaptive model selection." *Georgia State University* (2011).

United States Department of the Treasury. "Report to the Congress – Report to Congress on International Economic and Exchange Rate Policies, May 1992". *U.S. Department of the Treasury, Office of International Affairs* (1992).

United States Department of the Treasury. "Interim Report to the Congress on International Economic and Exchange Rate Policies, July 1994". *U.S. Department of the Treasury, Office of International Affairs* (1994).

United States Department of the Treasury. "Report to the Congress – Report to Congress on international economic and exchange rate policies, December 2006". *U.S. Department of the Treasury, Office of International Affairs* (2006).

United States Department of the Treasury. "Report to the Congress – Macroeconomic and foreign exchange policies of major trading partners of the United States, October 2018". *U.S. Department of the Treasury, Office of International Affairs* (2018).

Valickova, Petra, Tomas Havranek, and Roman Horvath. "Financial development and economic growth: A meta-analysis." *Journal of Economic Surveys* 29.3 (2015): 506-526.

Wan, Alan TK, and Xinyu Zhang. "On the use of model averaging in tourism research." *Annals of Tourism Research* 36.3 (2009): 525-532.

Wan, Alan TK, Xinyu Zhang, and Guohua Zou. "Least squares model averaging by Mallows criterion." *Journal of Econometrics* 156.2 (2010): 277-283.

Wang, Yajie, Xiaofeng Hui, and Abdol S. Soofi. "Estimating renminbi (RMB) equilibrium exchange rate." *Journal of Policy Modeling* 29.3 (2007): 417-429.

Wooster, Rossitza B., and David S. Diebel. "Productivity spillovers from foreign direct investment in developing countries: A meta-regression analysis." *Review of Development Economics* 14.3 (2010): 640-655.

World Bank. "Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program — ICP." *Washington, DC: World Bank* (2013).

World Bank. "Purchasing power parities and real expenditures of World economics: A comprehensive report of the 2011 International Comparison Program (English)." *International Comparison Program (ICP). Washington, DC: World Bank Group* (2015).

Wright, Jonathan H. "Bayesian model averaging and exchange rate forecasts." *Journal of Econometrics* 146.2 (2008): 329-341.

You, Kefei, and Nicholas Sarantis. "Structural breaks and the equilibrium Chinese yuan/us dollar real exchange rate: a FEER approach." *Review of International Economics* 19.5 (2011): 791-808.

Zeugner, Stefan, and Martin Feldkircher. "Bayesian model averaging employing fixed and flexible priors: The BMS package for R." *Journal of Statistical Software* 68.4 (2015): 1-37.

Zhang, Xinyu, Alan TK Wan, and Guohua Zou. "Model averaging by jackknife criterion in models with dependent data." *Journal of Econometrics* 174.2 (2013): 82-94.

Zhang, Zhibai, and Langnan Chen. "A new assessment of the Chinese RMB exchange rate." *China Economic Review* 30 (2014): 113-122.