Random Coefficient Models for Time-Series–Cross-Section Data: Monte Carlo Experiments of Finite Sample Properties∗

Nathaniel Beck† Jonathan N. Katz‡

Draft of April 1, 2006

∗We gratefully acknowledge the financial support of the National Science Foundation. We are thankful to Bill Greene and Rob Franzese for comments, to Larry Bartels for always reminding us that our judgment may outperform the data as well as to the anonymous reviewers for Political Analysis.

†Department of Politics; New York University; New York, NY 10003 USA; nathaniel.beck@nyu.edu

‡Division of the Humanities and Social Sciences; California Institute of Technology; Pasadena, CA 91125 USA; jkatz@caltech.edu
This article considers random coefficient models (RCMs) for time-series–cross-section data. These models allow for unit to unit variation in the model parameters. The heart of the paper compares the finite sample properties of the fully pooled estimator, the unit by unit (unpooled) estimator and the (maximum likelihood) RCM estimator. The maximum likelihood estimator RCM performs well, even where the data were generated so that the RCM would be problematic. In an appendix we show that the most common FGLS estimator of the RCM models in always inferior to the maximum likelihood estimator, and in smaller samples dramatically so.
1. RCMS AND TSCS DATA

Time-series–cross-sectional (TSCS) models have become very common in the study of comparative politics (defined expansively). In general, TSCS analysts seem willing to assume that all countries (units) are completely homogenous, such that a fully pooled model is appropriate. This assumption is often made without considering the alternatives. There are, of course, such alternatives. Many analysts allow for unit specific intercepts, that is, fixed effects. But there are relatively few attempts to go beyond this limited heterogeneity. Obviously we must assume enough homogeneity to allow for estimation; if every observation is unique, we can do no science. But it is not necessary to assume that the only alternative to complete uniqueness is complete pooling (or its close cousin, pooling other than for the unit specific intercepts). At first glance, the commitment to homogeneity is a bit odd, since a model which imposes less heterogeneity, the random coefficients model (RCM), has been well known for a quarter of a century. (We refer to the RCM as both a model and also a method for estimating the parameters of such a model.) Such models were considered in the light of comparative policy by Western (1998). Should TSCS analysts routinely entertain the RCM? Does it work well for the kinds of data and the kinds of questions typically seen in comparative politics? In this paper we continue the investigation of this question.

We begin by noting that the issue of whether to pool or not, or, more accurately, how much to pool, confronts every researcher. In an important, if uncited\(^1\) piece, Bartels (1996) argues that we are always in the position of deciding how much we should pool some observations with others, and we always have a choice ranging from complete pooling to assuming that the data have nothing to do with each other. He notes that, in general, political scientists seem to assume that either data completely pool or that some data is completely irrelevant, ignoring the in between position. The solution that Bartels proposes is that one should estimate a model allowing for varying degrees of pooling and then make a scientific decision after examining the locus of all such estimates. The procedure involves much judgment, since Bartels works in a purely cross-sectional context; in that context, the data alone can never determine the appropriate degree of pooling.

The situation is happier for the TSCS analyst, since we can assume complete homogeneity within units, and thus limit ourselves to models that allow for heterogeneity between units. It is this world we examine here. Western has described the RCM in a Bayesian context. The citation fate of his paper is no happier than that of Bartels.\(^2\) We thus must ask, if RCMs are so good, why do we not see TSCS analysts using them?\(^3\)

Since in this article we wish to focus on a comparison of the properties of the various estimators for the types of data typically seen in comparative politics, we omit many of the

---

\(^1\)The SSCI has 19 cites, with few, if any, being empirical applications.

\(^2\)Western has 27 cites. Few, if any, of these have an application of his proposed method.

\(^3\)It is possible that current interest in Bayesian methods will lead to a new interest in RCMs estimated via Bayesian methods. Much of this is stimulated by the work of Andrew Gelman, but it is still in its early stages. See Shor, Gelman, Park, Bafumi and Keele (2005) for an early conference paper out of this project. Gelman’s interest in this project is stimulated by his interest in the Bayesian multilevel (hierarchical) models (Park, Gelman and Bafumi, 2004). We return to the intimate tie between the RCM and the multilevel model below.
mathematical derivations, referring the interested reader to either the Hsiao’s (2003) revised text or our earlier paper (Beck and Katz, 2004). The next section of the paper lays out the notation of the RCM and the various estimators and discusses how we should evaluate the various estimators form the perspective of students of comparative politics. Section 3 provides Monte Carlo evidence on specific estimators and Section 4 concludes. The appendix consider the properties of one the most common estimators of RCM models, the Swamy-Hsiao feasible generalized least squares (FGLS) estimator.

2. ESTIMATING RANDOM COEFFICIENT MODELS

We assume standard TSCS data with a continuous dependent variable. The fully pooled model is:

\[ y_{i,t} = \mathbf{x}_{i,t} \beta + \epsilon_{i,t}; i = 1, \ldots, N; t = 1, \ldots, T \]  

(1)

where \( \mathbf{x} \) is a (row) vector of \( K \) independent variables and \( \beta \) is a \( K \)-vector of parameters. We assume that the errors are serially independent (perhaps after including a lagged dependent variable in the specification).\(^4\) We assume that \( \epsilon_{i,t} \stackrel{\text{iid}}{\sim} N(0, \sigma^2) \), so the error process is (for TSCS data) unusually simple. All these simplifying assumptions are made so that we can focus on the issue of parameter heterogeneity. Obviously in the real world the issues will be more complicated. The simple pooled model we have can be estimated by OLS.

The less pooled model has

\[ y_{i,t} = \mathbf{x}_{i,t} \beta_i + \epsilon_{i,t}; i = 1, \ldots, N; t = 1, \ldots, T \]  

(2)

We use the term “less pooled” because the notation allows for the \( \beta_i \) to follow a variety of patterns. We could, for instance, allow each unit’s \( \beta_i \) to be completely unrelated, the fully unpooled model, which would be estimated by unit by unit OLS.

We can test \( H_0: \text{Eq 1} \) against the alternative \( \text{Eq 2} \) by a standard \( F \)-test using the SSEs from the OLS estimations of both Eq 1 and 2. Typical procedure would then be to go with the pooled model if the \( F \)-ratio is small and with the fully unpooled (unit by unit OLS) model if it is large, with small and large being judged relative to 95% critical values of the \( F \)-distribution.

The “random coefficient model” (RCM) adjoins to Eq 2 the assumption that the \( \beta_i \) are all related, in particular they are draws from a multivariate normal distribution. Thus the RCM is Eq 2 with the additional assumption:

\[ \beta_i \sim N(\beta, \Gamma) \]  

(3)

where \( \Gamma \) is a matrix of variance and covariance terms to be estimated. \( \Gamma \) indicates the degree the heterogeneity of the unit parameters (\( \Gamma = 0 \) indicates perfect homogeneity). For some purposes it is simpler to write

\[ \beta_i = \beta + \nu_i \]  

(4)

\[ \nu_i \sim N(0, \Gamma) \]  

(5)

\(^4\)Our interest in this paper is not in dynamics, so the important issues about how to model dynamics need not detain us here.
\( y_{i,t} = x_{i,t}\beta + \{x_{i,t}\nu_i + \epsilon_{i,t}\} \). 

Thus the RCM can be seen as a linear model with a complicated error term (in braces).

While the assumption of parameters being drawn from a distribution is not a natural assumption for a classicist, the RCM can be estimated by classical (maximum likelihood) methods. The RCM framework also fits easily into a Bayesian setup; Western (1998), for example, uses Bayesian methods (and terminology) to estimate the RCM. While Bayesian numerical methods may be superior, with non-informative priors both Bayesians and classicists are going to get very similar asymptotic estimators and for the purposes of this paper it does not matter if one takes a classical or Bayesian position, so long as the Bayesian adopts non-informative priors. Thus the classicist may speak of “mixed estimation” while the Bayesian may speak of “hierarchical models,” but for our purposes these are more or less the same (and also the same as “multilevel” models (Steenbergen and Jones, 2002).

While the RCM and hierarchical and multilevel models are formally equivalent, they are often used in different research settings, and, most importantly for us, different sample sizes. Thus the RCM is typically seen in TSCS data, which typically has ten to fifty units observed annually for ten to fifty years. Applications of hierarchical or multilevel models often do not have the units observed over time framework, but rather units nested in other units. However, the number of first level units is identical, for our purposes, to the number of years in an RCM. Hierarchical models often have very few first level observations (say students in a classroom) or a huge number (survey respondents in a country). This article focuses on the typical sample sizes used in the typical comparative politics TSCS data. Thus all our Monte Carlo results are for \( N \)'s and \( T \)'s typical for such data, and for the remainder of this paper we refer only to the RCM.

Working with a classical interpretation, we use maximum likelihood methods. In particular, we use the methods of Pinheiro and Bates (2000), which are implemented in their “linear mixed estimation” R package. While the maximization is quite complicated, in theory this is just maximum likelihood, so we simply refer readers to the Pinheiro and Bates book for details.\(^5\)

So far we have discussed estimating the fundamental parameters of the RCM, the overall \( \beta \) and the variance of the distribution of the \( \beta_i \)'s, \( \Gamma \). One can clearly estimate the \( \beta_i \) using unit by unit OLS; such estimates will have usual optimal asymptotic properties, and, given the typically small number of observations per unit (\( T \)), rather high variability. Alternatively,\(^5\)

\(^5\)There are details galore. In particular, many analysts prefer “restricted” maximum likelihood (REAL) to more traditional maximum likelihood for these types of models. For simplicity we stick to the traditional maximum likelihood estimators (and note that our simulations indicate restricted and full maximum likelihood yield similar results for TSCS data). For those who prefer Stata to R, Stata also implements a maximum likelihood routine which performs very similar to the R package. There are many computational methods for actually maximizing the likelihood. (Greene, 2003, 512–7), for example, recommends simulated maximum likelihood. And of course one could use Markov chain Monte Carlo. Choosing between the methods of doing maximum likelihood estimation is not our interest here.
one could simply use the overall pooled estimate as the estimate for each of the $\beta_i$ which will have less variability but also will clearly not take into account the variability of the $\beta_i$. (So asymptotically using the pooled estimate to estimate the individual $\beta_i$ must be inferior to unit by unit OLS, but for finite (and small!) $T$ it could prove superior.

The RCM provides an alternative way to estimate the $\beta_i$. This consists of finding the “best linear unbiased predictor” (BLUP), which has $E(\beta_i) = \beta$ and lowest error loss in the class of such linear estimators. The various texts we refer to provide the formulae, but the BLUP can be seen as shrinking back the unit by unit estimates to the overall pooled estimate, with the degree of shrinkage a function of the uncertainty of the estimates. In our simulations we compare the RCM estimates of the $\beta_i$ to the unit by unit estimates.

Our Monte Carlo studies thus examine the finite sample properties of three different estimators: fully pooled, unit by unit and the RCM (estimated via maximum likelihood) for both the overall $\beta$ and the unit specific $\beta_i$. Those familiar with RCMs will note that we have omitted one other standard RCM estimator, which uses feasible generalized least squares (FGLS).

The FGLS estimator is due to Swamy (1971) though many know this work through Hsiao (2003). This estimator was derived decades ago, long before maximum likelihood estimation of the RCM was computationally practical. It is still implemented in standard statistical packages and is still discussed in current work as an alternative to full maximum likelihood estimation. Thus, for example, the currently leading text on panel (and TSCS) data, Hsiao (2003, 146–7), first presents the FGLS estimator and then states “[a]lternatively, we can use the Bayes model estimator suggested by [Smith].” He then provides one empirical example of an RCM, which was estimated via FGLS only.

In Beck and Katz (2004) we showed that the Swamy/Hsiao FGLS estimator can have quite poor finite sample properties. This is because making the estimated variance-covariance matrix positive definite required assuming that the sampling variance of the individual unit estimators was zero, an assumption that could be very bad in practice. While here we work only with the maximum likelihood estimator, and so need not worry about the FGLS estimator, the latter is still commonly used (as can be inferred from the cite in the previous paragraph). Since the properties of the FGLS estimator are orthogonal to our interests here, we present the major findings from our previous work in an appendix to this paper, with full details available in our previous paper. The appendix shows that the maximum likelihood estimator is better (and for smaller $T$ substantially better, than the Swamy-Hsiao FGLS estimator. Given that it is now quite feasible for any researcher to do full maximum likelihood estimation of the RCM, there is simply no reason to continue to consider the FGLS estimator.

Before turning to the Monte Carlos, we must first decide on how to assess the estimators. Since our focus on this paper is a comparison of the accuracy of the various estimators in estimating the fundamental parameters of the model, we report only root mean squared errors (RMSE) of the various estimates of either $\beta$ or $\beta_i$. Students of comparative politics will differ on whether they are which of those parameters are of more interest.
Evaluating RCMs for Students of Comparative Politics

Our basic query is, if RCMs are so good, why aren’t they used. One answer is that if we are only interested in estimating the mean $\beta$, not the $\beta_i$, then pooled OLS, provides a consistent, although not fully efficient, estimate of the hyperparameter, $\beta$. And of course the standard errors of the fully pooled OLS estimates will be wrong, since they fail to account for unit to unit parameter heterogeneity. But it well may be the case that the complexities of RCM estimation are not worth the bother if we only care about the mean $\beta$.

Consistency, of course, is an asymptotic property. That two estimators are each consistent does not guarantee that they perform equally well (or even nearly so) in finite samples. Thus we will perform experiments which compare the various estimators of $\beta$. Since our interests are in finite samples, we do this in the context of typical sample sizes commonly observed in the study of comparative politics. We return to this issue in the next subsection.

Students of comparative politics might also be interested in getting accurate estimates of the country level $\beta_i$. This might be of interest for exploratory work, where analysts simply wish to examine the variability of the country level $\beta_i$’s, or for graphical work which plots the $\beta_i$ against variables of interest. Of course there may be situations where researchers wish to use the individual estimates of the $\beta_i$ in their final conclusions (“the effect of inflation on approval is highest in Germany and France and lowest in Spain and Portugal”). If the RCM estimates of the $\beta_i$ are superior to the country by country OLS estimates, then analysts will find the RCM helpful. We thus compare the performance of the RCM, unit by unit OLS and fully pooled OLS (taking the estimate of the pooled $\beta$ as the estimator of each $\beta_i$) for estimating the $\beta_i$.

Finally we note that while not all students of comparative politics will make conclusions based on the country level $\beta_i$, all would like to account for variation in the $\beta_i$ by modeling

$$\beta_i = \alpha + z_i\kappa + \mu_i$$

where the $z_i$ are covariates that pertain to a country but do not vary over time (structural characteristics of the country). But substituting Eq. 8 into Eq. 2 shows that this is still an RCM (equivalent to Eq. 7), but one that contains non-random multiplicative terms interacting the $x$ and $z$ covariates. Thus while modeling the random coefficient as a function of structural covariates should be of great interest to students of comparative politics, this adds nothing new statistically to the RCM and hence we do not consider it further in this paper.

Choice of $N$ and $T$ and $k$ for the simulations

At this point we turn our inquiry to how the various estimators of RCMs perform for typical TSCS situations. Almost all published work in this area has involved asymptotics. Given that we are working with TSCS data, the justification of hierarchical methods with

---

6For examples of this type of exploratory work in the context of comparative analyses of surveys, see Kedar and Shively (2005). The analysis in that volume all have large first stage $T$’s, and so it is not difficult to use country by country OLS. For TSCS data the smaller $T$’s make the situation more complicated.
asymptotics in \( N \) is irrelevant to us.\(^7\) But the analysis of RCM methods looking at asymptotic properties as \( T \to \infty \) is equally useless; with an infinite amount of data per unit, the unit-by-unit OLS estimates are of course perfect and there is no need to pool (or to inquire whether the data might pool, since with infinite observations we know each coefficient perfectly and so unless they are identical, we will reject pooling). Thus we need to look at Monte Carlo evidence on the performance of various RCM estimators given the \( N_s \) and \( T_s \) we observe in actual TSCS data sets.

While there are a wide variety of TSCS studies of comparative politics, a typical number of countries (\( N \)) ranges from perhaps 15 to 50 (usually not less, sometimes more). The statistical properties of both the (unpooled) unit by unit estimator and the RCM are largely determined by \( T \), not \( N \). This is because the unit by unit estimator averages over \( T \) observations, and the RCM is a weighted average of the unpooled and fully pooled estimators. While it is possible that our results might not hold for very small \( N \)’s (say 3), our results are not sensitive to \( N \) for the typical \( N \)’s of comparative politics. We thus compare the estimators setting \( N = 20 \).

We will vary \( T \) so as to assess estimation issues for sample lengths seen in practice. We thus let \( T = 5, 10, \ldots, 50 \). We do not go beyond \( T = 50 \) since by that point the comparisons are clear cut and we are coming close to what would be the asymptotic results.

To focus attention on a comparison of the various estimators in the simplest case, we only examine a model with a single covariate (so \( k = 1 \)). We always draw from zero mean distributions, but estimate a single constant term (that is, in truth there are neither unit effects nor a constant term, but we estimate the constant term though no effects). Obviously as \( k \) increases the quality of both unit by unit OLS and the RCM will degrade (for a fixed \( T \)). A number of issues arise for the RCM for \( k > 1 \), particularly related to the variance covariance matrix, \( \Gamma \). In particular, \( \Gamma \) contains roughly \( \frac{K^2}{2} \) parameters. We postpone a comparison of estimators for the multivariate situation to a future paper.

**Design of the simulations**

The simulations consist of 1000 replications. Before the start of the replications we drew the \( N \times T \) regressors from \( x_{i,t} \ind \sim N(0, \sigma_x^2) \). By drawing the \( x_{i,t} \) only once we are able to simulate the case of fixed regressors, a standard assumption used in practice. We could, of course, allow for more complicated structure to the regressors, but this would just cloud the issue. Note that for all the experiments we set \( \sigma_x^2 = 0.01 \), this value was chosen so that given our \( \beta = 5 \) and variance on the error terms, \( \sigma_x^2 = 1 \), the average t-statistic on the unit by unit OLS estimates would be approximately 2 when \( T = 20 \). Then on each replication of the simulation we drew the \( N \) unit parameters from \( \beta_i \ind \sim N(\beta, \gamma^2) \). (Because we have only one random coefficient, \( \Gamma \) is scalar, and so we denote its variance as \( \gamma^2 \).) For all of the experiments presented here \( \beta \) was fixed at 5. We then generated the \( y_{i,t} = \beta_i x_{i,t} + \varepsilon_{i,t} \) where

---

\(^7\)All the estimators are consistent. If the RCM is the correct specification, pooled OLS is less efficient than is the RCM for estimating the overall \( \beta \), and the OLS standard errors will be misleading. In finite samples the RCM estimate of \( \beta \) is biased but unit by unit estimation is unbiased, though with larger squared error loss than the RCM estimator.
\( \varepsilon_{i,t} \overset{\text{ind}}{\sim} N(0, \sigma^2_{\varepsilon}) \) where \( \sigma^2_{\varepsilon} = 1 \). This is the simplest possible RCM model for TSCS data. Unlike our previous work, we do not focus on the accuracy of the standard errors. This is because if the data are generated by an RCM, then we know from theory that the RCM standard errors are more accurate than the OLS standard errors. Given the \( T \)'s we use, a comparison of the unit by unit standard errors is of little interest. In previous work there was a tradeoff of accuracy of standard errors and efficiency of estimation. As we shall see, this is not the case for the RCM, and hence we present no simulation results on the accuracy of standard errors.

3. MONTE CARLO RESULTS

Our first results compare the RCM (maximum likelihood) and OLS estimates of the overall \( \beta \) in Eq 3 when the data are generated by an RCM (with the standard deviation of the \( \beta_i, \gamma = 1.8 \)). Figure 1 shows the root mean squared (RMS) errors of the estimate of \( \beta \) as a function of \( T \) where \( T \) goes from 5 to 50 in increments of 5.

The results are clear. The RCM estimate of \( \beta \) always is more accurate than the corresponding OLS estimate, and the gain from using the RCM estimate is non-trivial. The advantage of the RCM is greatest for the smallest \( T \); at our minimal \( T = 5 \), the RCM is twice as accurate as is OLS. But even for our largest \( T = 50 \), the RCM is still 25% more accurate than OLS. Thus the efficiency gain of the RCM over OLS for estimating \( \beta \) is substantial, and makes it worthwhile to use the RCM to estimate the overall \( \beta \) if we think there is any variation in the unit \( \beta_i \).

The next experiment asks if we can be misled by the RCM if indeed there is no variation in the unit \( \beta_i \). If \( \gamma = 0 \) pooled OLS must dominate the RCM (since the OLS model is correct and has one fewer parameter to estimate). But even in this extreme situation, the advantage of OLS in terms of the RMS error of estimating \( \beta \) is trivial. For example, when \( T = 20 \), when \( \gamma = 0 \) the RMS error for the pooled OLS of \( \beta \) is .4858; for the RCM it is .4862. With even a tiny \( \gamma \) the RCM becomes superior to pooled OLS, and that superiority of course grows with \( \gamma \). But it is good to know that even when the pooled model is correct, the RCM performs at least 99% as well as pooled OLS. There is no fear that the RCM can mislead analysts into finding parameter variation when there is none.

Turning to the estimates of the unit \( \beta_i \), we have three different estimators: the unit by unit OLS estimates, the RCM estimates, and the pooled OLS estimates (so each \( \beta_i \) is assumed to equal the common estimate of \( \beta \). Keeping all parameters as in Figure 1, in Figure 2 we show the RMS errors of the three estimators of the \( \beta_i \).

The RCM again clearly dominates. Until \( T \) gets large (about 30), the pooled estimator does a better job of estimating the individual \( \beta_i \)'s than does unit by unit OLS; while eventually unit by unit OLS must give better estimates than the pooled estimator, there are many practical situations where investigators would prefer the pooled to the unit by unit estimates of the \( \beta_i \). Even for \( T > 30 \), where the unit by unit estimate is superior to the pooled estimate, the gains of the unit by unit estimator are not large (about 10%).

But we need not detain ourselves on this comparison, since the RCM estimate of the individual \( \beta_i \)'s is superior to either the fully pooled or unit by unit estimates. For all \( T \)'s the
Figure 1: Comparison of Root Mean Square Error for RCM and OLS estimators of $\beta$ as $T$ varies from 5 to 50. For all runs of the experiment $N = 20$, $\beta = 5$, $\gamma = 1.8$, $\sigma^2_x = 1$, and $\sigma^2_\epsilon = 0.01$.

gain from using the RCM is substantial, with that gain (over the better of the two OLS based estimates) being substantially over 50%. If one cares about the estimates of the individual $\beta_i$ these experiments show that one should use the RCM.

We have, so far, presented experiments where the RCM is likely to fare the best (though we have shown that even if there is no unit to unit variation the RCM does not mislead).
But what happens if there unit to unit parameter heterogeneity, but it is not normal. In Figure 3 we consider estimating the $\beta_i$ in a situation where 18 of the unit $\beta_i$’s are 5 and two are outliers, varying from 5 to 10. How well does the RCM do in picking up the $\beta_i$ here?

The answer is quite well. As we saw previously, when there is little or no unit to unit variation the RCM and pooled OLS perform similarly; when the first two coefficients are the
same, or almost the same, as the other eighteen, the RCM and pooled OLS perform about identically. But, as before, even when the parameters are all generated identically, the RCM performs as well as the pooled OLS estimator; it does not mislead even when there is no parameter variation.

As we induce two bigger and bigger outliers in the $\beta_i$ the RCM continues to perform well.
It is not the case that the highly non-normal way we drew the twenty \( \beta_i \) causes problems for the RCM. Thus, at least for this case, even when the \( \beta_i \) do not look at all like Eq 3 the RCM still yields good estimates of the \( \beta_i \). This is most reassuring, since the normality of the \( \beta_i \) is clearly an assumption driven more by mathematical convenience than by empirical reality.

The unit by unit estimates of the \( \beta_i \) are not affected at all by increased heterogeneity, but, unfortunately, given that \( T = 20 \), this performance is uniformly bad. As the outliers become more extreme unit by unit OLS starts to perform almost as well as pooled OLS (or, more accurately, pooled OLS performs worse as the outliers become more severe). But we need not concern ourselves with this comparison, since the RCM performs best in this situation, and its performance is not degraded as the outliers become more severe.

4. CONCLUSION

There is little doubt, as Western noted, that the RCM should appeal to scholars of comparative politics (broadly defined) who are not so naive as to assume that all countries (or units) are identical but who are sufficiently committed to statistical analysis that they cannot assume that all observations are \textit{sui generis}. The RCM appears to offer the analyst a flexible middle position. And it appears as though we can allow the data to tell us just how heterogeneous the units are.

As noted, Bruce Western’s analysis appears to be the only political science application of RCMs to TSCS data (that we know of), and that one appears to be largely for didactic purposes. So we must ask why RCMs have seen so little usage, and should they be used more? We stress that we only deal here with TSCS data and make no claims about panel or other types of data.

Our Monte Carlo experiments indicate that the RCM performs very well, both for the estimation of the overall \( \beta \) and the unit specific \( \beta_i \). This assumes that estimation is by some variant of maximum likelihood. As we saw in the Appendix, the performance of the FGLS RCM estimator is quite poor. We note that there are a number of different ways to actually maximize the likelihood; since we have investigated only one such way, we cannot compare the various alternatives. We would assume the estimates are similar, so those who like Markov chain Monte Carlo, or simulated likelihood, are probably quite safe in using those techniques to explore the likelihood.

Not only does the RCM perform well in terms of mean squared error, our results indicate that it does not mislead, in the sense that if there is little or no parameter heterogeneity, the RCM will not falsely find it. There are no places where simpler method (non-trivially) dominate the RCM in terms of mean square error. Thus, unlike other complicated routines, there seems little danger from starting with the RCM. It also appears that the RCM handles parameter dispersion that is highly non-normal, even though the model assumes normality. How robust the estimators are to all the possible violations of assumptions is well beyond the scope of any one paper, but none of our experiments (both reported and unreported) lead us to expect the RCM to be a brittle estimator.

If the RCM indicates little parameter heterogeneity (that is, a small and/or insignificant \( \gamma \)) then we might wish to go back to simpler, OLS based, methods, since these will be more
flexible. But the R implementation of the RCM is quite flexible; it can handle a variety of correlated error structures (including serially correlated errors with an ARMA structure), and it extends very nicely to non-linear models. Since we often find that the use of a lagged dependent variable simplifies the error structure, it is unclear how often the complicated errors structures will be needed, but for those who want them, they are there. But if, for example, spatial issues are of interest, and there is not much parameter heterogeneity, then it might be the case that starting with a simple linear model would allow researchers to better deal with spatial issues. Such tradeoffs will always exist for any method.

So why do applied TSCS analysts in political science seem not to use the RCM? One reason is that the real gains come from estimating the unit specific parameters, and many comparativists may be primarily interested in the overall parameters (including general determinants of the variability as in Eq 8). But it is hard to believe that there is not some interest in the unit specific parameters, even if only to check for outlying units.

The RCM is not a panacea. While it allows researchers to check for homogeneity across units, it is not the only way to do so. Thus we have recommended in other papers (Beck, 2001) that analysts assess homogeneity via cross-validation (leaving out a unit at a time); this still makes sense. The R routines come with a large variety of diagnostics to assess whether the assumptions of the RCM hold; these should be rigorously used. Clearly residuals need to be checked for time-series and spatial issues.

In this article we have looked only at the simplest case, one independent variable with a random coefficient and a single, fixed constant term. We do not know how the RCM will perform in more complicated situations. What happens if one has 10 independent variables? One thing that is almost certainly true is that “nuisance” control variables should probably be treated as fixed. And since the full $\Gamma$ covariance matrix has approximately $K^2/2$ parameters to estimate, it is likely to be problematic to estimate a full $\Gamma$ matrix for ten independent variables. But one can specify simpler structures, particularly by assuming that the random coefficients are all independent (so that only the diagonal elements of $\Gamma$ need to be estimated. While one may carp at that assumption, current practiced consists of making the much stronger assumption of homogeneous coefficients. Just because the RCM allows for a huge amount of flexibility does not mean that we need to use all that flexibility; general experience shows that imposing some structure on a flexible model often improves matters. But specific recommendations must await future work.

It may be that the poor performance of the FGLS estimator of the RCM led researchers to abandon the RCM. (In previous papers we were much more negative about the RCM because of our reliance on the FGLS approach.) Western’s Markov chain Monte Carlo approach was published only a decade ago; the Pinheiro and Bates R code has only been widely available for half a decade and the Stata code has only been out for a year. So the strongest conclusion of this paper is that TSCS analysts should think seriously about including the RCM as one of their principal empirical modeling tools. The costs are low and the gains are enormous.
5. APPENDIX: THE POOR FINITE SAMPLE PROPERTIES OF THE SWAMY/HSIAO FGLS ESTIMATOR

As noted in the text, a standard RCM estimator (e.g. Hsiao, 2003, 144–51) is a particular implementation of Swamy’s (1971) FGLS estimator. While we have no doubt that analysts should use maximum likelihood (or related) estimators for the RCM model, the FGLS method is still used, is implemented in standard statistical packages and is considered viable in standard texts (as shown by our discussion in Section 2 of Hsiao’s leading text). Since the poor finite sample properties of the FGLS estimator are clear, and we have laid out the details in our unpublished paper (Beck and Katz, 2004), we only lay out the brief argument here and present key Monte Carlo results. Since the derivation of the FGLS estimator is in Hsiao (2003, 145–6), we only present the final results here, referring the reader to Hsiao for the derivations.

To understand the issues, recall that the estimate of the overall mean from the pooled OLS is inefficient because it does not use all of the information in the structure of the model. An FGLS estimate of $\beta$ builds on the RCM as a linear model with a complicated error structure, as in Eq 7. As with all FGLS estimators, we start with the consistent but inefficient estimator of $\beta$, OLS. We then estimate the parameters of the non-spherical variance-covariance matrix of the errors, and then use this estimate in the standard GLS transformation. Hsiao shows that everything can be built up from the OLS estimators, $\hat{\beta}$ and $\hat{\beta}_i$(OLS estimates are designated by carets with tildes for FGLS).

Hsiao shows that the FGLS estimator can be seen as a weighted sum of unit by unit OLS estimators of the $\beta_i$,

$$\tilde{\beta} = \sum_{i=1}^{N} W_i \hat{\beta}_i$$

$$W_i = \left\{\sum_{i=1}^{N} [\Gamma + V_i]^{-1}\right\}^{-1} [\Gamma + V_i]^{-1}$$

$$V_i = \sigma_i^2 (X_i'X_i)^{-1}$$

where $X_i$ denotes the matrix of observations for unit $i$, $\Gamma$ is the variance matrix of random coefficients (Eq 3) and $V_i$ is the variance of $\hat{\beta}_i$.

Since this is feasible generalized least squares, we do not know $\Gamma$ or $\sigma_i^2$, but instead we use estimates of them from the initial OLS regressions. Swamy and Hsiao suggest the usual OLS estimate of $\sigma_i^2$. The question is how to estimate $\Gamma$? If we could directly observe the $\beta_i$ we could use the $N$ draws to construct an estimate of the covariance matrix in the usual fashion:

$$\tilde{\Gamma} = \frac{1}{N-1} \left( \sum_{i=1}^{N} (\beta_i - \frac{1}{N} \sum_{i=1}^{N} \beta_i)'(\beta_i - \frac{1}{N} \sum_{i=1}^{N} \beta_i) \right).$$

Since averaging is done over $N$, estimation will improve as $N$ grows; this is very important for panel analysts, but of cold comfort for TSCS analysts.
The problem is we do not observe \( \beta_i \); we have, instead, only a noisy estimates, \( \hat{\beta}_i \). So while we might consider just substituting \( \hat{\beta}_i \) for \( \beta_i \) in the definition of \( \tilde{\Gamma} \), this would lead us to over estimate the amount of variation in \( \beta_i \) since much of the variation in the \( \hat{\beta}_i \)'s is caused not by “real” parameter variability but purely by sampling error. We can correct for this sampling variability by subtracting it off. Swamy thus suggested that a plausible estimator of \( \Gamma \) to use in the FGLS estimation is

\[
\tilde{\Gamma} = \frac{1}{N-1} \left( \sum_{i=1}^{N} (\hat{\beta}_i - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i)' (\hat{\beta}_i - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i) \right) - \frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_i^2 (X_i'X_i)^{-1}. \tag{13}
\]

There is, however, a problem with this estimator: in finite samples \( \tilde{\Gamma} \) need not be positive definite, a necessary requirement for it to be a well defined covariance matrix. This is because we are subtracting off the estimated sampling variance which can be large for the typical values of \( T \) found in TSCS data.

The question is how to ensure that \( \tilde{\Gamma} \) is positive definite? Hsiao (2003, 146), building on Swamy, suggests that the second term of \( \tilde{\Gamma} \) be dropped. The rationale for this is asymptotic, that is, as \( NT \) gets large the sampling variance goes to zero. This fix is not correct in finite samples as it will tend to overestimate \( \Gamma \). The question for TSCS analysts is how badly does this omission of sampling variance affect the estimate of \( \Gamma \) and does this cause any problems in the estimates of \( \beta_i \) for typical values of \( N \) and \( T \) seen in actual research situations? Since these are problems in finite samples we will have to assess the claims via Monte Carlo simulations.

**Monte Carlo evidence**

We replicate our standard experiment for assessing the estimate of the \( \beta_i \), but this time simply comparing the Swamy-Hsiao FGLS estimates to the maximum likelihood estimates (replicating the conditions of Figure 2). Results are in Figure 4.

Clearly the FGLS estimator performs considerably worse than the maximum likelihood estimator. As \( T \) gets larger this inferiority declines, but even for our largest \( T = 50 \), the FGLS RMS errors are twice those of the maximum likelihood errors.

Why does the FGLS estimator perform so poorly. As noted above, it is because in finite samples the variance term is underestimated (by the assumption that the sampling variance of the unit by unit estimates of \( \beta_i \) is zero). In Figure 5 we compare the RMS errors of the estimate of \( \gamma \) for the FGLS and maximum likelihood procedures. For very small \( T \leq 10 \) the FGLS estimate of \( \gamma \) is horrible; even past that, the RMS error for the FGLS estimator is never less than fives times as large as for the maximum likelihood estimator. The Hsiao assumption that the sampling variance is zero, used to ensure positive definiteness of a variance-covariance matrix, is the culprit here.

**Conclusion**

When the FGLS estimators for the RCM were developed there were no computationally feasible alternatives to FGLS. The world has changed, and maximum likelihood is now quite
feasible. The maximum likelihood estimator strongly dominates the FGLS estimator (and always will be better in all possible ways). Given that the maximum likelihood estimator is now implemented in commonly used packages, there is simply no reason for anyone to again consider the Swamy-Hsiao FGLS estimator. It may once have been the best available compromise, but its compromise is both costly and completely unnecessary today.
Figure 5: Comparison of Root Mean Square Error for RCM and FGLS estimators of $\gamma$ as $T$ varies from 5 to 50. For all runs of the experiment $N = 20$, $\beta = 5$, $\gamma = 1.8$, $\sigma^2_\varepsilon = 1$, and $\sigma^2_\varepsilon = 0.01$. 
REFERENCES


