

**Value Consensus and Measurement Inequivalence:
The Case of Sexual Emancipation Values**

Kirill Zhirkov¹ and Christian Welzel²

¹ University of Virginia

² Leuphana University Lueneburg

[Draft. Please don't cite or circulate]

Abstract

An ongoing controversy in comparative politics concerns the cross-cultural equivalence of measurement models for some social value scales that are widely used in cross-national research on political culture. In this article, we examine a particularly interesting example—the measurement equivalence of sexual emancipation values—using a multilevel Bayesian factor model. Our results confirm previous findings that measurement models underperform in countries with Muslim majorities and low levels of economic development. We then present new evidence indicating that the measurement model failure reflects extremely high levels of value consensus in these countries. This implies that measurement inequivalence will be found whenever societies with low values heterogeneity, manifested in near-zero response dispersions, are included in the analysis. We argue that, since the existence of tight vs. loose cultures is a real phenomenon, it should be a subject of systematic study in comparative political research instead of being treated as a measurement problem.

Keywords: cross-national research, measurement equivalence, value orientations

A growing body of comparative political research focuses on measurement equivalence. Its key assumption is that equivalence of measurement models is a necessary condition for any meaningful comparison of attitudinal measures across countries or groups (Davidov et al. 2014; also see Perez 2011). In particular, this concern has been expressed with regard to self-expression values and emancipative values that, according to revised modernization theory, link economic development with democratization (Inglehart and Welzel 2005; Welzel 2013). Recent studies indicate that the measures of self-expression and emancipative values are not equivalent across countries, casting doubt on findings that use these constructs (Aleman and Woods 2016; Sokolov 2018; but cf. Welzel and Inglehart 2016; Welzel et al. 2021).

Such studies, however, use multi-group confirmatory factor analysis (MG CFA), a method originally designed to simply document inequivalence without any built-in tools to explain it. To address this limitation of the method, researchers have come up with several analytic strategies. One of them is close examination of country-specific factor loadings and the application of substantive expertise to explain why differences are found (Meuleman and Billiet 2018). Another method is introduction of country-level covariates to account for outlying mean responses to specific survey questions (Davidov et al. 2012). Researchers can also use follow-up questions on how respondents understood specific survey items (Meitinger 2017). Most recently, multilevel Bayesian CFA modeling with random loadings has been proposed for this purpose (Meuleman and Schlueter 2018).

In this article, we use Bayesian CFA modeling to investigate the extent to which factor loadings for sexual emancipation values, one of the central variables in political culture research, reflect substantive economic and cultural characteristics of studied countries. In keeping with previous studies, we find that the measurement model for sexual emancipation values performs

worse in countries with low levels of economic development and in Muslim-majority societies. In the next step, however, we demonstrate that the reason why the measurement model performs relatively poorly is extremely high levels of value consensus—that is, low individual-level variation on moral values in these countries. Following this result, we argue that the existence of “tight” cultures in many societies is not a methodological artefact but rather an important social reality that needs to be studied by scholars of comparative politics and political culture.

Explanations of Inequivalence

Studies that have explored measurement inequivalence in value orientations suggest some potential sources. One is the society’s level of economic development. For instance, the estimated factor loadings for the postmaterialism index are closely associated with the countries’ per capita incomes (Van de Vijver and Poortinga 2002). Measurement reliability of emancipative values within a country is also positively related to its level of economic development (Welzel 2013). Overall, factor models for value orientations seem to work best in advanced industrial democracies (Aleman and Woods 2016).

Another variable that has been suspected to cause cross-country differences in survey response is religion. There are indeed pronounced cross-national differences in basic value orientations that are linked with the society’s dominant religious tradition, with self-expression values and emancipative values being highest in the Protestant countries of Northern Europe (Inglehart and Welzel 2005). This is not surprising, since religious heritage has a profound impact on such key social and political phenomena as partisan cleavages, welfare states, and educational systems (Grzymala-Busse 2012). Accordingly, the measurement models for emancipative values perform differently across groups of countries depending on the religious

tradition (Sokolov 2018). It makes the research on political culture vulnerable to the criticism of imposing the standards of modern Western Christianity upon the rest of the world.

Moreover, cross-national surveys face the challenge of linguistic differences: most of these surveys are originally designed in English and then translated into local languages. Even if the translations are equivalent, speakers of grammatically distant languages may process and answer the same questions in different ways (Perez and Tavits 2019).

This brief review suggests that measurement models for value orientations may perform better in (a) countries with higher levels of economic development, (b) Christian and specifically Protestant countries, and (c) countries with languages proximate to English. Overall, most cross-national surveys are designed by researchers from countries that are Western, educated, industrialized, rich, and democratic (WEIRD; Henrich, Heine, and Norenzayan 2010). If survey items are cognitively tailored to the mindsets of WEIRD people, it can create better measurement models in those countries—and raise questions about their comparative validity. We test this conjecture using a Bayesian CFA model with random factor loadings with sexual emancipation values as an example.

Data and Modeling

The sexual emancipation values scale plays a central role in modernization theory that links economic development, cultural change, and democratization. The scale consists of three questions that ask respondents about their attitudes toward homosexuality, abortion, and divorce. Answers range from 1 = *Never justifiable* to 10 = *Always justifiable*. We use data on these variables from the 6th wave of the World Values Survey (WVS) fielded in 2010–2014. This dataset covers 81,315 respondents in 57 countries. We also replicate our model for WVS Waves 3 and 5 (at least 50 countries each; see Table S1 in Supplementary Material). As the predictors of

measurement inequivalence, we use economic development measured as log GDP per capita in purchasing power parities according to the Penn World Table (Feenstra, Inklaar, and Timmer 2015), religious heritage (Norris and Inglehart 2011), and linguistic distance of a country's plurality language to English (Spolaore and Wacziarg 2016).

The estimated measurement model has both an individual-level component and a country-level component (see Figure 1). On the individual level, we estimate a standard CFA model with three observed variables—respondents' acceptance of homosexuality, abortion, and divorce—as indicators of a single latent construct, sexual emancipation values. However, in contrast to the standard CFA set-up, we estimate the country-level variances of factor loadings. We then use country-level covariates to predict differences in factor loadings across countries.

Formally, we analyze survey data on individuals indexed $i = 1, \dots, I$ nested in countries indexed $j = 1, \dots, J$. Denote the latent factor as η and the observed indicator variables as y_k indexed $k = 1, \dots, K$. Then, indicator k for individual i in country j is modeled as:

$$y_{kij} = \mu_{kj} + \lambda_{kj}\eta_{ij} + \varepsilon_{kij} + u_{kj},$$

where μ_{kj} is the country-specific intercept, λ_{kj} is the country-specific factor loading, η_{ij} is the unobserved factor score, ε_{kij} is the individual-level measurement error, and u_{kj} is the country-level measurement error. We use country-level covariates x_h indexed $h = 1, \dots, H$ to predict the variation of factor loadings across countries. Then, factor loading is expressed as:

$$\lambda_{kj} = \alpha_k + \sum_{h=1}^H \beta_{kh}x_{hj} + \delta_{kj},$$

where α_k is the intercept, β_{kh} is the covariate-specific coefficient, and δ_{kj} is the error term.

Combining these equations yields the integrated model:

$$y_{kij} = \mu_{kj} + \left(\alpha_k + \sum_{h=1}^H \beta_{kh} x_{hj} + \delta_{kj} \right) \eta_{ij} + \varepsilon_{kij} + u_{kj}.$$

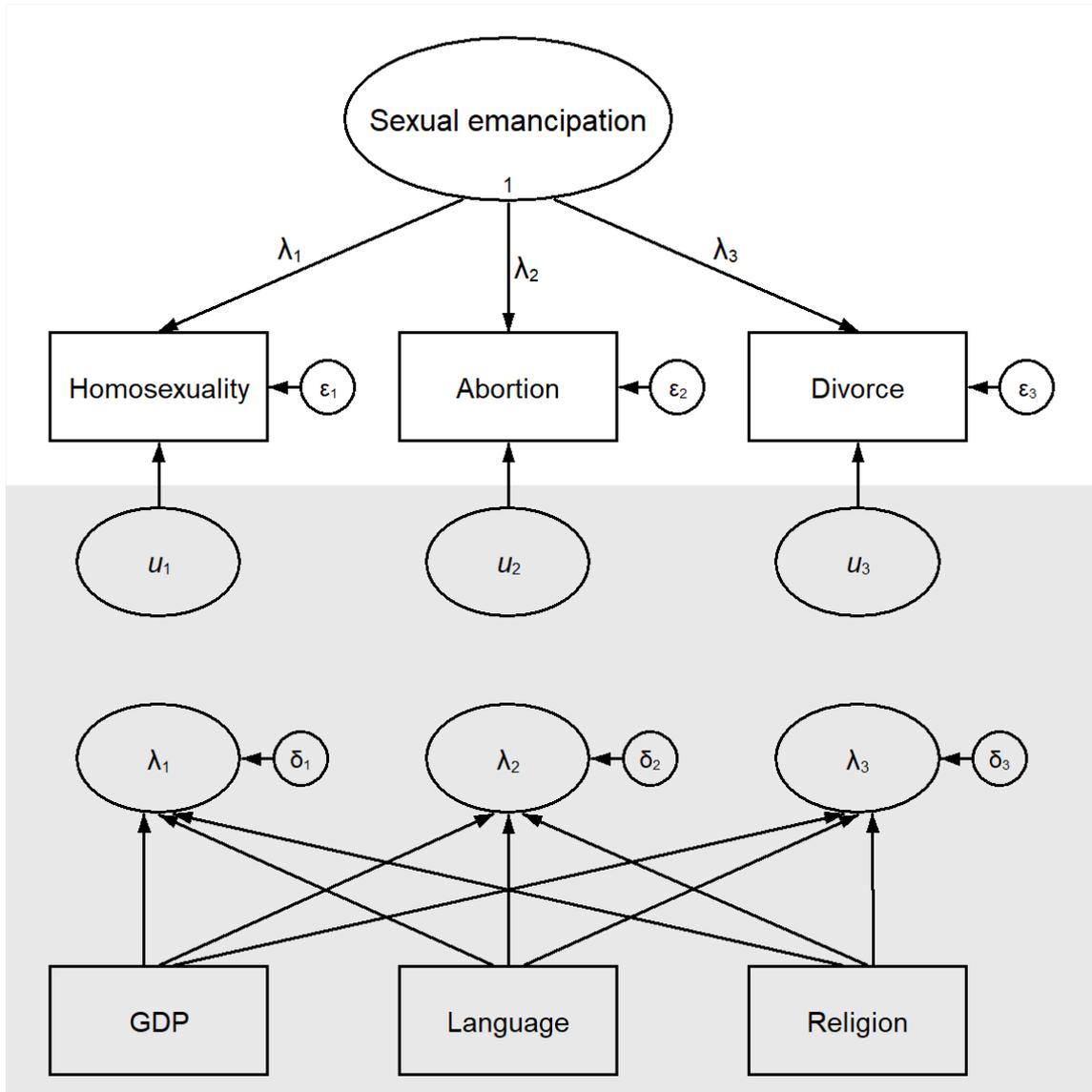


Figure 1. Bayesian CFA model with random loadings for sexual emancipation values. White background: individual level; grey background: country level

To estimate the parameters of interest, we employ Bayesian modeling in Mplus. For Bayesian estimation, Mplus uses a Markov chain Monte Carlo (MCMC) algorithm based on a Gibbs sampler. Convergence is monitored using the potential scale reduction criterion (Gelman

and Rubin 1992). Additionally, we run Kolmogorov–Smirnov tests to ensure equivalence of posterior parameter distributions produced by the MCMC chains. As uninformative (improper) priors, we use normal distribution with zero mean and infinite variance for factor loadings, structural effects, and intercepts and inverse-Gamma distribution with shape -1 and rate 1 for variances. The CFA models are estimated using two MCMC chains with 10,000 iterations each. Overall, all models indicate very good convergence by achieving stationary conditions within less than 2,000 iterations. However, we run the MCMC chains for a larger number of iterations in order to check for stability of the posterior distributions.

By default, Mplus treats the first half of iterations as burn-in and then uses the second half to calculate model parameters. As a result, reported posterior distributions are based on 5,000 iterations from each of the two chains (i.e., 10,000 iterations total). We report all point estimates for parameters as means of the respective posterior distributions.

Results

When running the model, we set the variance of the latent factor to the mean variance of the three indicators.¹ Consequently, the approximate range of estimated factor loadings is between zero (the lowest possible level of sexual emancipation values) and one (the highest possible level). We also rescale non-binary country-level covariates into a range from 0 (minimum observed value) to 1 (maximum observed value). As a result, estimated coefficients have intuitive interpretations. Coefficients for GDP per capita and distance from English represent estimated differences between countries with the lowest possible scores (least developed, plurality language most distant from English) and highest possible scores (most developed,

¹ When we re-estimate the model constraining one of the loadings instead of latent factor variance, we obtain substantively similar results with different scaling. See Figure S1 in Supplementary Material.

plurality language is English). Coefficients for religious zones represent estimated differences from Protestant countries (baseline category).

Figure 2 presents the results of the Bayesian CFA model. The estimates suggest that factor loadings for attitudes towards homosexuality and abortion are significantly lower in poorer and Muslim-majority countries. Factor loadings for attitudes toward homosexuality are also lower in Orthodox Christian societies. Closeness of a country’s plurality language to English does not show any significant or sizeable effects on the quality of the measurement model. The same findings apply in more or less the same fashion to WVS Waves 3 and 5 (see Figures S2 and S3 in Supplementary Material).

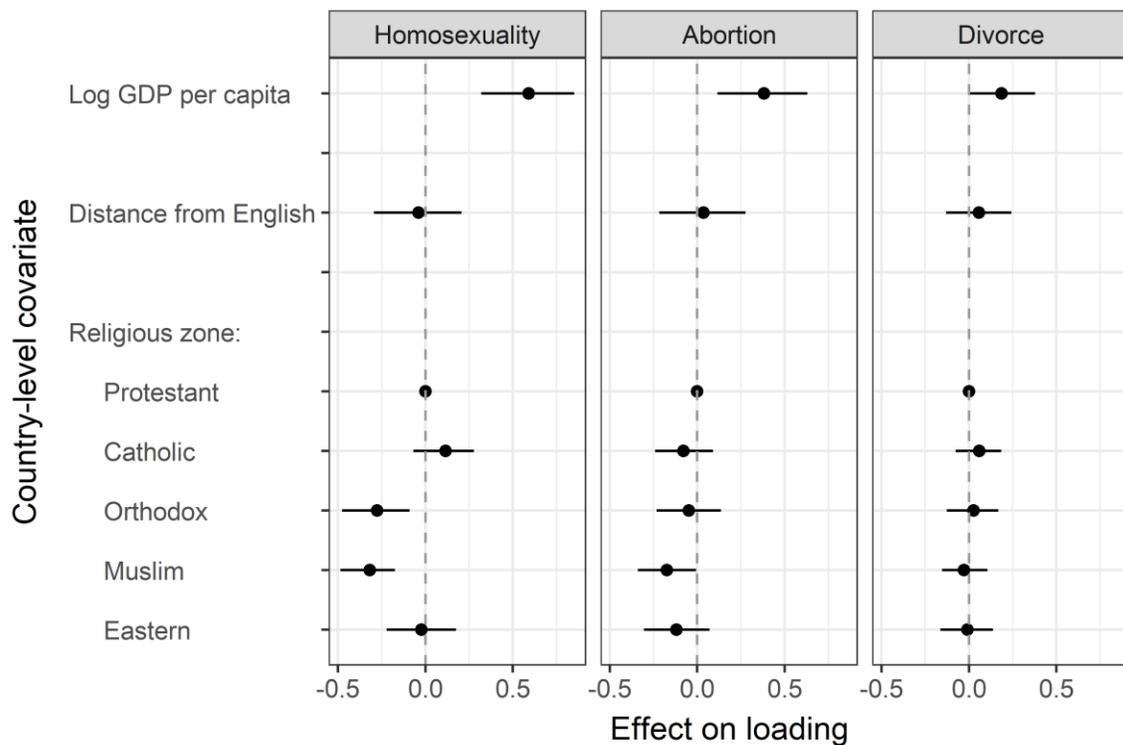


Figure 2. Effects of country-level covariates on factor loadings for the three indicators of sexual emancipation values, WVS Wave 6. Point estimates with 95% Bayesian credible intervals

The Issue of Value Consensus

Do these results indicate that attitudes toward homosexuality and abortion are less important as indicators of sexual emancipation values in poor and Muslim-majority societies? Or, in these countries, questions on attitudes toward homosexuality and abortion tap into a different value dimension than the question on divorce? Within the modern measurement equivalence framework, these would indeed be the conclusions. But they would be premature. There is an important alternative explanation for underperforming measurement models: high value consensus, or low cultural heterogeneity, within the most conservative societies (Blaydes and Grimmer 2020; Welzel et al. 2021). The presence of value consensus is crucial because of the inherent relationship between factor loadings and sample variances: extremely low variation in item responses automatically diminishes the factor loadings.

A factor-analytic model involves measuring latent variables (factors) using a larger number of observed variables (indicators). In matrix notation, it can be represented as follows:

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda} + \mathbf{\Theta},$$

where $\mathbf{\Sigma}$ is the symmetric covariance matrix, $\mathbf{\Lambda}$ is the matrix of factor loadings, $\mathbf{\Psi}$ is the symmetric matrix of correlations between factors, and $\mathbf{\Theta}$ is the covariance matrix of errors (uniquenesses). Suppose, without loss of generality, the model features three indicators and assumes the presence of a single unit-variance factor with no covariances between error terms.

Then, the equation can be represented in the following expanded form as:

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \times [\lambda_1 \quad \lambda_2 \quad \lambda_3] + \begin{bmatrix} \theta_1 & 0 & 0 \\ 0 & \theta_2 & 0 \\ 0 & 0 & \theta_3 \end{bmatrix},$$

or

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} \lambda_1^2 + \theta_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ \lambda_2\lambda_1 & \lambda_2^2 + \theta_2 & \lambda_2\lambda_3 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2 + \theta_3 \end{bmatrix},$$

where σ_{kk} is the variance of indicator k , $\sigma_{kk'}$ is the covariance of indicators k and k' , λ_k is the factor loading of indicator k , and θ_k is the unique variance of indicator k . The goal of fitting a factor model is to determine estimates for λ_k and θ_k parameters that reproduce the observed variances and covariances as closely as possible.

Consider the following equality:

$$\sigma_{kk} = \lambda_k^2 + \theta_k.$$

Suppose that the estimate for λ_k in country j is very small (close to zero). What are the potential causes? The standard explanation is that the factor model fits poorly for country j as the uniqueness of indicator k is large compared to the variance explained by the common factor (i.e., $\theta_k \gg \lambda_k^2$). There is, however, an alternative possibility: the variance of the indicator, σ_{kk} , can be extremely small in the first place. It is easy to see that, when the distribution of an indicator is extremely dense (i.e., $\sigma_{kk} \approx 0$), estimates of λ_k and θ_k have to be close to zero for purely mathematical reasons.

Therefore, if a large majority of respondents in a country agree that homosexuality, abortion, and divorce are “never justified,” these indicators have little variance—a substantively important finding that, nevertheless, produces factor models with near-zero loadings. We test this conjecture using a simple OLS regression that predicts within-country dispersions (standard deviations) on the three indicators of sexual emancipation values with GDP per capita, linguistic closeness to English, and religious zone. The results, presented in Figure 3, show strong support for the assumed relationship: poorer and Muslim-majority countries have significantly lower dispersions on all three indicators. The effects are strongest for attitudes toward homosexuality

and weakest for attitudes toward divorce. These results are, again, replicated for WVS Waves 3 and 5 (see Figures S4 and S5 in Supplementary Material).

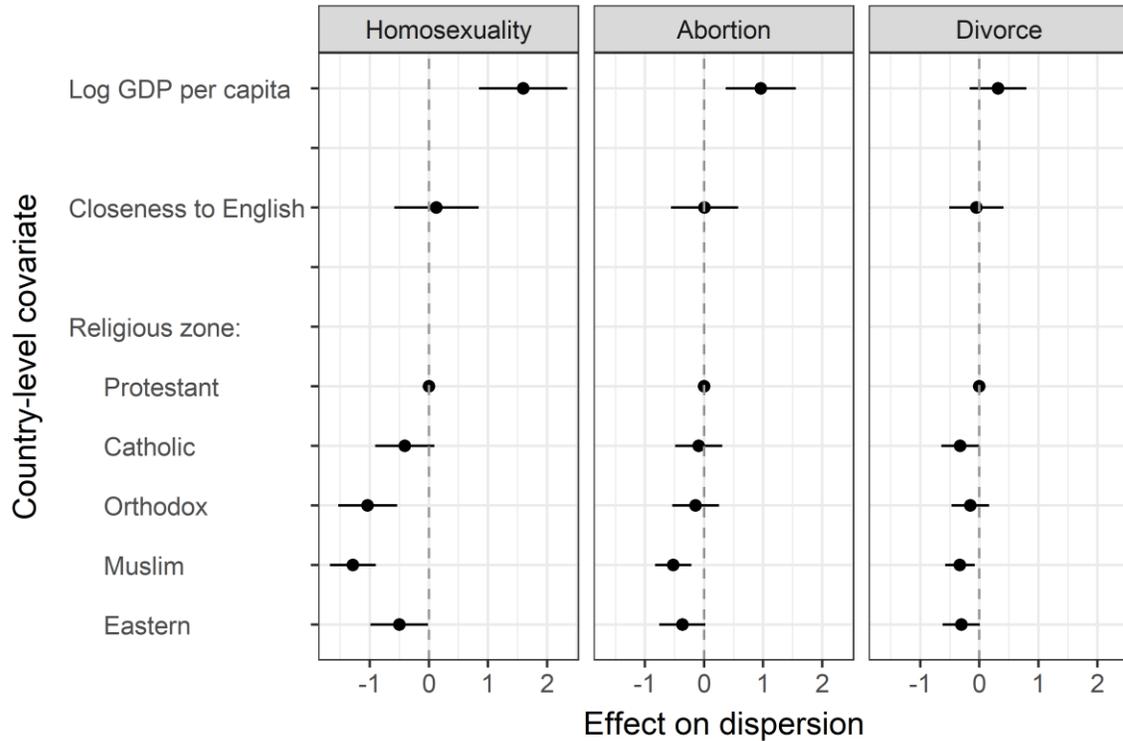


Figure 3. Effects of country-level covariates on within-country dispersions of the three indicators of sexual emancipation values, WVS Wave 6. Point estimates with 95% confidence intervals

In some countries, variation is almost nonexistent—that is, response distributions are close to being deterministic—and this observation is particularly true for attitudes toward homosexuality (see Figure 4 for distributions in countries with the lowest dispersions on the three indicators). Since the estimation of factor-analytic models is based on observed variances and covariances, fitting a factor model in countries with almost zero dispersions on the indicator items necessarily produces factor loadings that are close to zero. The takeaway, therefore, is that the measurement model for sexual emancipation values performs worse in poorer and Muslim-majority countries because the populations of these countries have extremely high levels of value

consensus: almost everyone rejects sexual emancipation. This is not an indication of measurement failure—on the contrary, it is an important substantive fact.

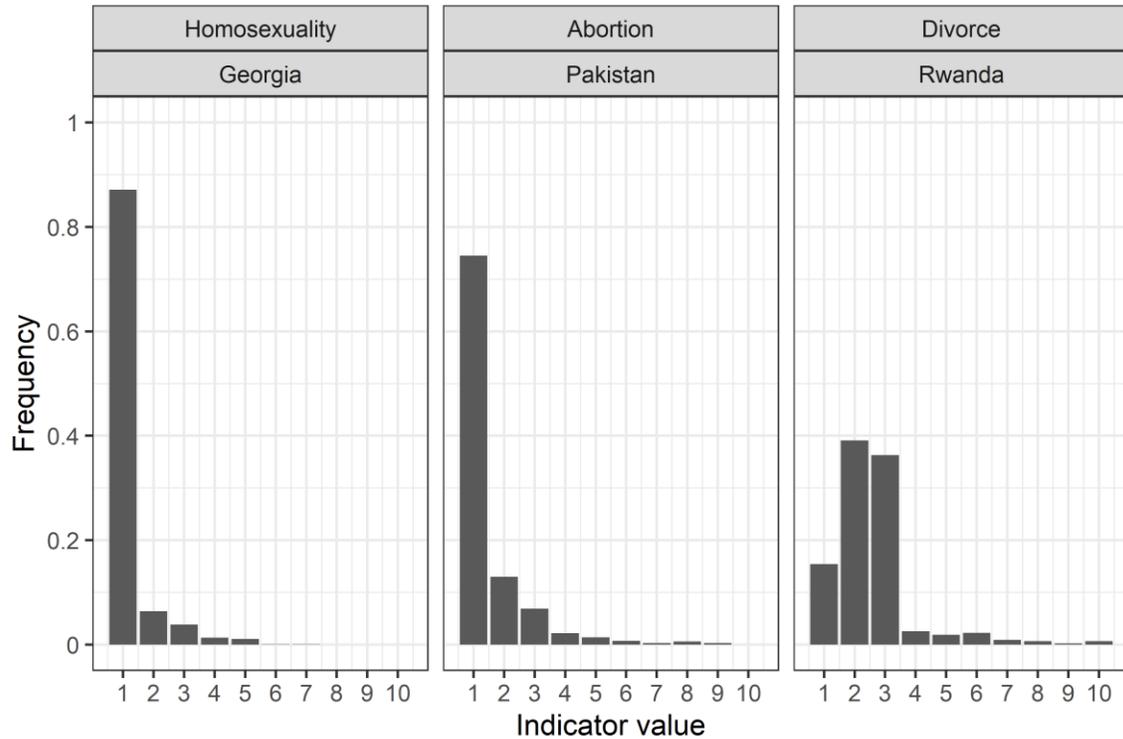


Figure 4. Histograms of reported attitudes; WVS wave 6

This pattern will be regularly produced in cross-cultural studies whenever there are widely disparate within-country dispersions on the social value scales. Extreme levels of value consensus (low dispersions) in certain groups of societies necessarily imply badly performing within-country measurement models. Therefore, measurement inequivalence may simply document large cross-national differences in value consensus—which can be exactly what is predicted by the theory being tested.

Implications

Our results have important implications for the distinction between measurement equivalence and substantive comparability that is generally overlooked in the existing literature. Contrary to a

popular perception, a failed equivalence test is a relatively weak result: it simply means that a scholar cannot establish whether the measures in question are comparable or not (Meredith 1993). Moreover, survey researchers have known for more than three decades that even individual-level survey responses within a single country are not perfectly equivalent (Brady 1985; for more recent evidence, see Perez and Hetherington 2014). However, it has not led to the abandonment of survey research, because researchers' goal is almost never to obtain and compare individual attitudinal scores. Instead, it is universally, although implicitly, accepted in the discipline that a mere possibility of non-comparability should not be an obstacle to exploring meaningful relationships between covariates of interest. Modern studies in political culture work within the same paradigm: they test theoretically motivated hypotheses on the aggregate level of analysis using multiple observations. Therefore, it is unclear why they should be held to a higher standard in proving comparability.

The phenomenon of value consensus highlights a major problem with this standard. Its adherents insist that country-level value scores cannot be compared if measurement equivalence is not established. But as we have shown, low items' dispersions necessarily imply low factor loadings. Therefore, the presence of countries with different levels of value consensus will always produce dissimilar factor loadings and, thus, a failed equivalence test.² The solution proposed within the measurement equivalence framework is to come up with items that vary similarly across countries. It means that, within this paradigm, value consensus is always treated as a methodological artefact produced by imperfect measurement. However, this contradicts massive evidence from across the social sciences that the existence of "tight" cultures—ones with strong norms and low tolerance of deviant behavior—is a profound and important reality

² Importantly, the opposite is not true: one can obtain failed equivalence tests for reasons other than low within-country dispersions.

(Gelfand et al. 2011). Mechanically eliminating any measure that does not pass conventional measurement equivalence tests would make the immensely significant concept of “tight” versus “loose” cultures untestable.

Conclusion

A growing number of studies in comparative politics examine the measurement properties of prominent value constructs, such as self-expression and emancipative orientations, concluding that they do not function equivalently across countries. But these studies simply document inequivalence without exploring its sources. In this paper, our goal has been to overcome this limitation. Using a Bayesian CFA model with random factor loadings, we have investigated differences in measurement models for sexual emancipation values and found that measurement models underperform in poor and in Muslim-majority countries. But the reason for this finding seems to be that people in poor and in Muslim-majority societies almost uniformly reject sexual freedoms (homosexuality in particular). Underperforming factor models, thus, can be an indicator of unequal levels of value consensus across countries.

These findings have important implications for the interpretation of inequivalence in political culture research. Within-country CFA models become increasingly meaningless as the analysis extends across the full range of cultural variation—which means that they cover societies with extreme levels of value consensus (i.e., near-zero dispersions). Such consensus is a profound reality in some societies, and it suppresses the factor loadings in measurement models for purely mathematical reasons. Therefore, equivalence testing is not always helpful in cross-national settings that include countries with extreme levels of value consensus—in such settings, failed equivalence tests become inevitable.

Overall, current practice of measurement equivalence testing in comparative political research assumes that individual-level response dispersions should be equally large in all analyzed countries—and that multi-item constructs should be designed to achieve such similarity. This assumption needs to be questioned. It denies existence of cross-national differences in value heterogeneity—despite the fact that shifts in moral consensus are an essential component of cultural change.

References

- Aleman, Jose, and Dwayne Woods. 2016. “Value Orientations from the World Values Survey: How Comparable Are They Cross-Nationally?” *Comparative Political Studies* 49 (8): 1039–67.
- Blaydes, Lisa, and Justin Grimmer. 2020. “Political Cultures: Measuring Values Heterogeneity.” *Political Science Research and Methods* 8 (3): 571–79.
- Brady, Henry E. 1985. “The Perils of Survey Research: Inter-Personally Incomparable Responses.” *Political Methodology* 11 (3/4): 269–91.
- Davidov, Eldad, Hermann Duellmer, Elmar Schlueter, Peter Schmidt, and Bart Meuleman. 2012. “Using a Multilevel Structural Equation Modeling Approach to Explain Cross-Cultural Measurement Noninvariance.” *Journal of Cross-Cultural Psychology* 43 (4): 558–75.
- Davidov, Eldad, Bart Meuleman, Jan Cieciuch, Peter Schmidt, and Jaak Billiet. 2014. “Measurement Equivalence in Cross-National Research.” *Annual Review of Sociology* 40: 55–75.
- Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. “The Next Generation of the Penn World Table.” *American Economic Review* 105 (10): 3150–82.

- Gelfand, Michele J., Jana L. Raver, Lisa Nishii, Lisa M. Leslie, Janetta Lun, Beng Chong Lim, Lili Duan, Assaf Almaliach, Soon Ang, Jakobina Arnadottir, Zeynep Aycan, Klaus Boehnke, Pawel Boski, Rosa Cabecinhas, Darius Chan, Jagdeep Chhokar, Alessia D'Amato, Montse Ferrer, Iris C. Fischlmayr, Ronald Fischer, Marta Fueoep, James Georgas, Emiko S. Kashima, Yoshishima Kashima, Kibum Kim, Alain Lempereur, Patricia Marquez, Rozhan Othman, Bert Overlaet, Penny Panagiotopoulou, Karl Peltzer, Lorena R. Perez-Florizno, Larisa Ponomarenko, Anu Realo, Vidar Schei, Manfred Schmitt, Peter B. Smith, Nazar Soomro, Erna Szabo, Naline Taveesin, Midori Toyama, Evert Van de Vliert, Naharika Vohra, Colleen Ward, and Susumu Yamaguchi. 2011. "Differences Between Tight and Loose Cultures: A 33-Nation Study." *Science* 332 (6033): 1100–04.
- Gelman, Andrew, and Donald B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7 (4): 457–72.
- Grzymala-Busse, Anna. 2012. "Why Comparative Politics Should Take Religion (More) Seriously." *Annual Review of Political Science* 15: 421–42.
- Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33 (2–3): 61–135.
- Inglehart, Ronald, and Christian Welzel. 2005. *Modernization, Cultural Change, and Democracy: The Human Development Sequence*. Cambridge: Cambridge University Press.
- Meitinger, Katharina. 2017. "Necessary but Insufficient: Why Measurement Invariance Tests Need Online Probing as a Complementary Tool." *Public Opinion Quarterly* 81 (2): 447–72.

Meredith, William. 1993. "Measurement Invariance, Factor Analysis and Factorial Invariance."

Psychometrika 58 (4): 525–43.

Meuleman, Bart, and Jaak Billiet. 2018. "Religious Involvement: Its Relation to Values and

Social Attitudes." In *Cross-Cultural Analysis: Methods and Applications*, edited by Eldad

Davidov, Peter Schmidt, Jaak Billiet, and Bart Meuleman, 2nd ed., 181–214. New York:

Routledge.

Meuleman, Bart, and Elmar Schlueter. 2018. "Explaining Cross-National Measurement

Inequivalence: A Bayesian Multilevel CFA With Random Loadings." In *Cross-Cultural*

Analysis: Methods and Applications, edited by Eldad Davidov, Peter Schmidt, Jaak

Billiet, and Bart Meuleman, 2nd ed., 363–90. New York: Routledge.

Norris, Pippa, and Ronald Inglehart. 2011. *Sacred and Secular: Religion and Politics*

Worldwide. 2nd edition. New York: Cambridge University Press.

Perez, Efren O. 2011. "The Origins and Implications of Language Effects in Multilingual

Surveys: A MIMIC Approach with Application to Latino Political Attitudes." *Political*

Analysis 19 (4): 434–54.

Perez, Efren O., and Marc J. Hetherington. 2014. "Authoritarianism in Black and White: Testing

the Cross-Racial Validity of the Child Rearing Scale." *Political Analysis* 22 (3): 398–412.

Perez, Efren O., and Margit Tavits. 2019. "Language Influences Public Attitudes toward Gender

Equality." *Journal of Politics* 81 (1): 81–93.

Sokolov, Boris. 2018. "The Index of Emancipative Values: Measurement Model

Misspecifications." *American Political Science Review* 112 (2): 395–408.

- Spolaore, Enrico, and Romain Wacziarg. 2016. "Ancestry, Language and Culture." In *Palgrave Handbook of Economics and Language*, edited by Victor Ginsburgh and Shlomo Weber, 174–211. Basingstoke: Palgrave Macmillan.
- Van de Vijver, Fons J. R., and Ype H. Poortinga. 2002. "Structural Equivalence in Multilevel Research." *Journal of Cross-Cultural Psychology* 33 (2): 141–56.
- Welzel, Christian. 2013. *Freedom Rising: Human Empowerment and the Quest for Emancipation*. New York: Cambridge University Press.
- Welzel, Christian, Lennart Brunkert, Stefan Kruse, and Ronald F. Inglehart. 2021. "Non-invariance? An Overstated Problem With Misconceived Causes." *Sociological Methods & Research*. Published ahead of print. <https://doi.org/10.1177/0049124121995521>
- Welzel, Christian, and Ronald F. Inglehart. 2016. "Misconceptions of Measurement Equivalence: Time for a Paradigm Shift." *Comparative Political Studies* 49 (8): 1068–94.

Supplementary Material

Table S1. Countries included in the WVS Waves 3, 5, and 6

WVS Wave 3

Albania, Argentina, Armenia, Australia, Azerbaijan, Belarus, Bosnia, Bulgaria, Chile, China, Colombia, Croatia, Czech Republic, Dominican Republic, El Salvador, Estonia, Finland, Georgia, Germany, Great Britain, Hungary, India, Japan, Latvia, Lithuania, Macedonia, Mexico, Moldova, Montenegro, New Zealand, Nigeria, Norway, Peru, Philippines, Poland, Romania, Russia, Serbia, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Ukraine, United States, Uruguay, Venezuela

Countries: 50, respondents: 70,946

WVS Wave 5

Argentina, Australia, Brazil, Bulgaria, Burkina Faso, Canada, Chile, China, Colombia, Cyprus, Ethiopia, Finland, France, Georgia, Germany, Ghana, Great Britain, Guatemala, Hong Kong, Hungary, India, Indonesia, Iran, Italy, Japan, Jordan, Malaysia, Mali, Mexico, Moldova, Netherlands, New Zealand, Norway, Poland, Romania, Russia, Rwanda, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, Trinidad and Tobago, Turkey, Ukraine, United States, Uruguay, Viet Nam, Zambia

Countries: 52, respondents: 73,300

WVS Wave 6

Algeria, Argentina, Armenia, Australia, Azerbaijan, Bahrain, Belarus, Brazil, Chile, China, Colombia, Cyprus, Ecuador, Estonia, Georgia, Germany, Ghana, Hong Kong, India, Iraq, Japan, Jordan, Kazakhstan, Kyrgyzstan, Lebanon, Malaysia, Mexico, Morocco, Netherlands, New Zealand, Nigeria, Pakistan, Palestine, Peru, Philippines, Poland, Qatar, Romania, Russia, Rwanda, Singapore, Slovenia, South Africa, South Korea, Spain, Sweden, Taiwan, Thailand, Trinidad and Tobago, Tunisia, Turkey, Ukraine, United States, Uruguay, Uzbekistan, Yemen, Zimbabwe

Countries: 57, respondents: 81,315

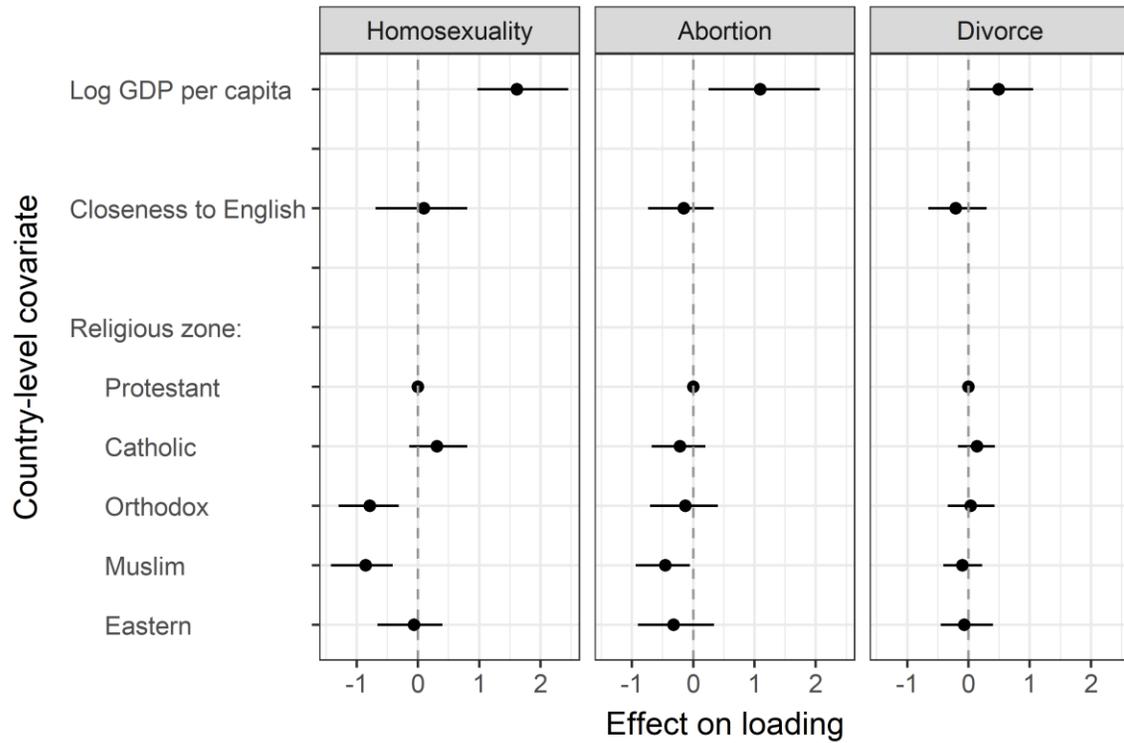


Figure S1. Effects of country-level covariates on factor loadings for the three indicators of sexual emancipation values, WVS Wave 6. Point estimates with 95% Bayesian credible intervals. Factor loading for “divorce” constrained to one

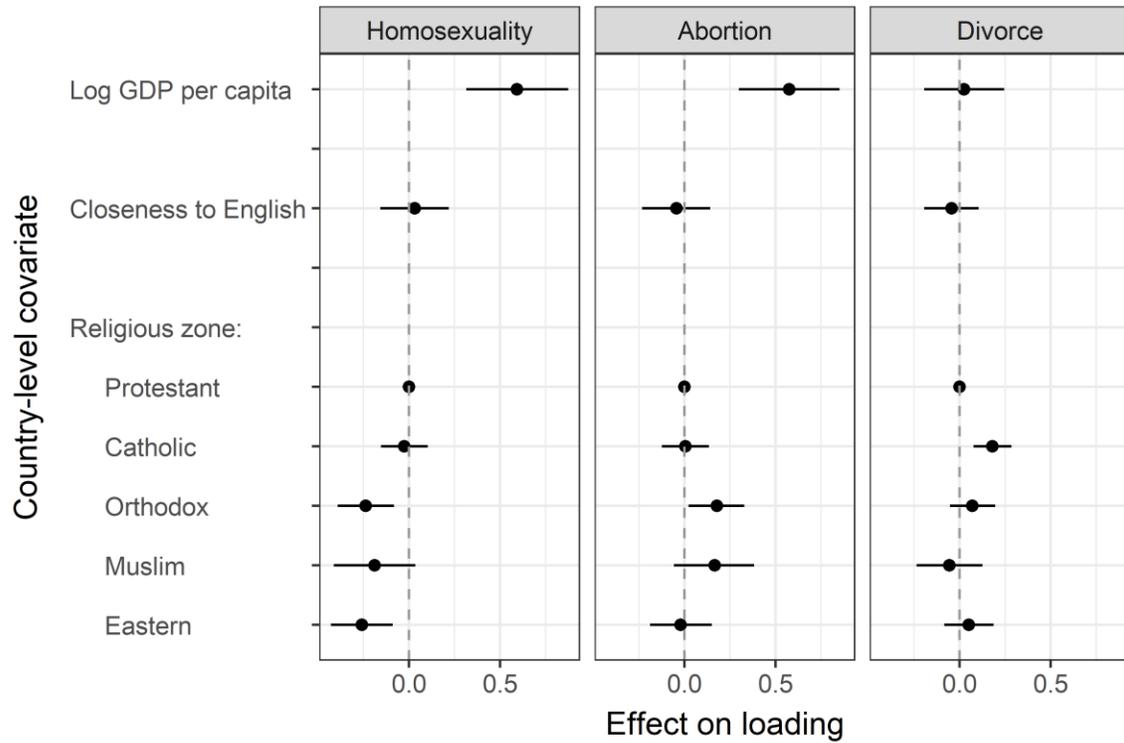


Figure S2. Effects of country-level covariates on factor loadings for the three indicators of sexual emancipation values, WVS Wave 3. Point estimates with 95% Bayesian credible intervals

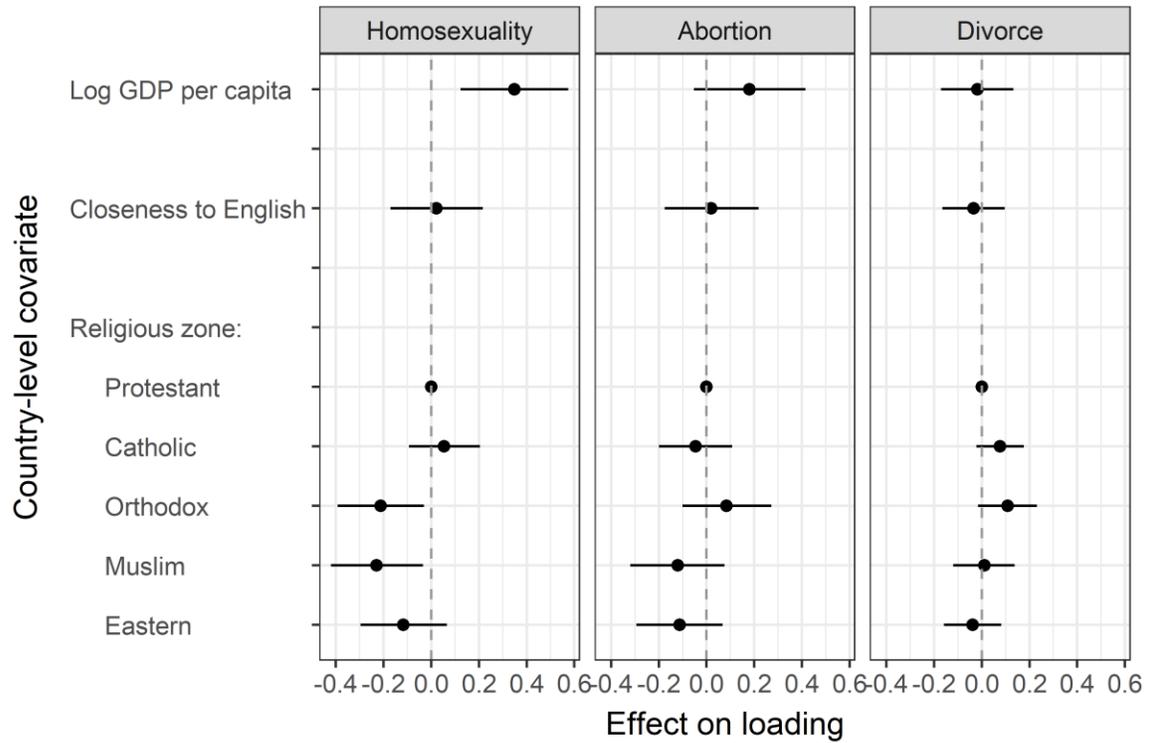


Figure S3. Effects of country-level covariates on factor loadings for the three indicators of sexual emancipation values, WVS Wave 5. Point estimates with 95% Bayesian credible intervals

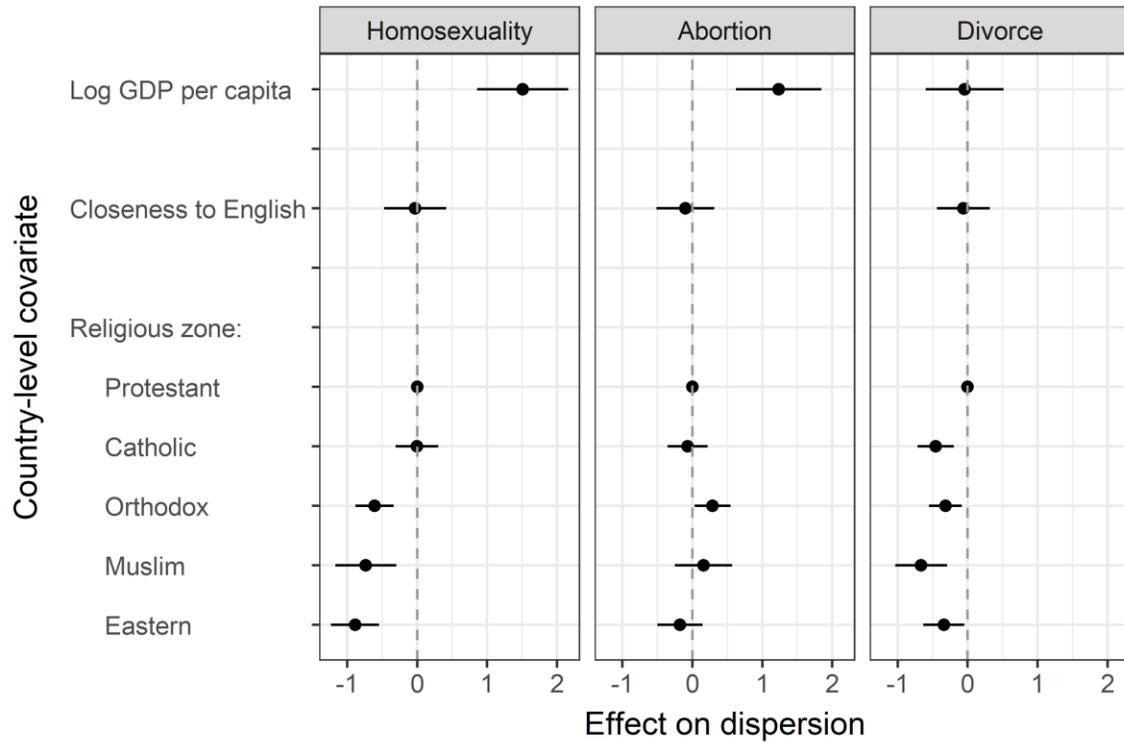


Figure S4. Effects of country-level covariates on within-country dispersions of the three indicators of sexual emancipation values, WVS Wave 3. Point estimates with 95% confidence intervals

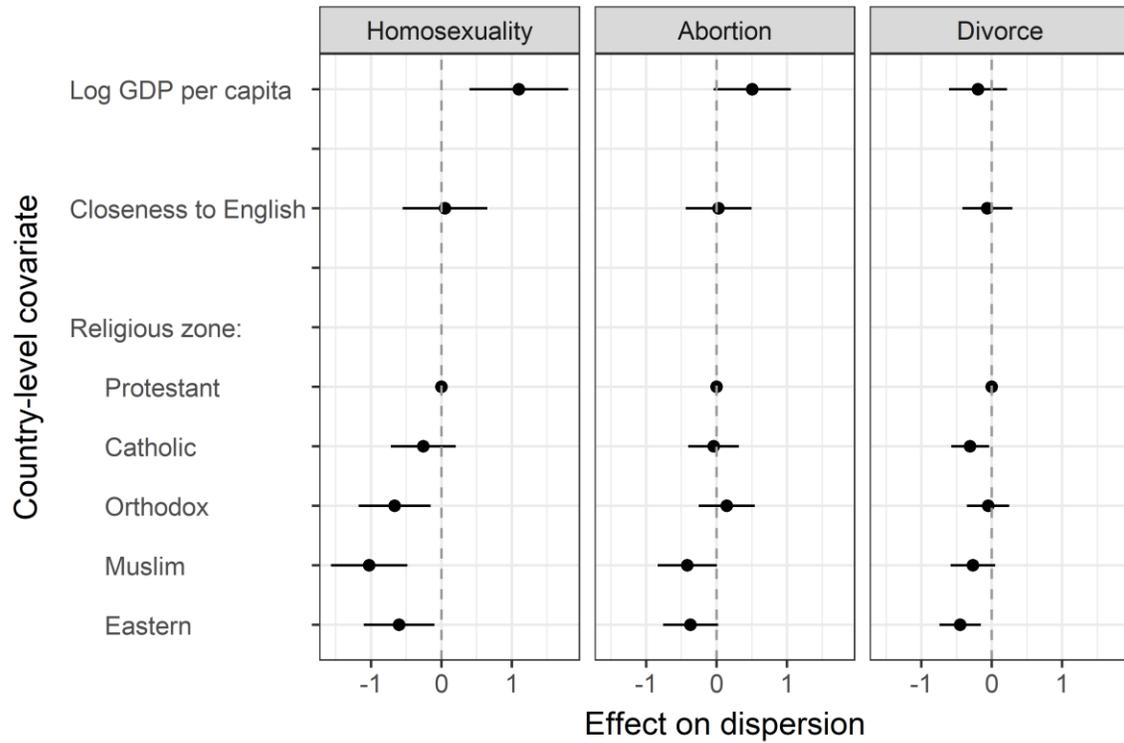


Figure S5. Effects of country-level covariates on within-country dispersions of the three indicators of sexual emancipation values, WVS Wave 5. Point estimates with 95% confidence intervals